Kneule, Andreas

Working Paper

Betriebswirtschaftliche Einsatzmöglichkeiten von Cognitive Computing

Wismarer Diskussionspapiere, No. 03/2018

Provided in Cooperation with:
Hochschule Wismar, Wismar Business School

Suggested Citation: Kneule, Andreas (2018) : Betriebswirtschaftliche Einsatzmöglichkeiten von Cognitive Computing, Wismarer Diskussionspapiere, No. 03/2018, ISBN 978-3-942100-60-1, Hochschule Wismar, Fakultät für Wirtschaftswissenschaften, Wismar

This Version is available at:
https://hdl.handle.net/10419/180619

Fakultät für Wirtschaftswissenschaften
Wismar Business School

Andreas Kneule

Betriebswirtschaftliche Einsatzmöglichkeiten von Cognitive Computing

Heft 03/2018

W D P

Wismarer Diskussionspapiere / Wismar Discussion Papers

Die Fakultät für Wirtschaftswissenschaften der Hochschule Wismar, University of Applied Sciences – Technology, Business and Design bietet die Präsenzstudiengänge Betriebswirtschaft, Wirtschaftsinformatik und Wirtschaftsrecht sowie die Fernstudiengänge Betriebswirtschaft, Business Consulting, Business Systems, Facility Management, Quality Management, Sales and Marketing und Wirtschaftsinformatik an. Gegenstand der Ausbildung sind die verschiedenen Aspekte des Wirtschaftens in der Unternehmung, der modernen Verwaltungstätigkeit, der Verbindung von angewandter Informatik und Wirtschaftswissenschaften sowie des Rechts im Bereich der Wirtschaft.
Nähere Informationen zu Studienangebot, Forschung und Ansprechpartnern finden Sie auf unserer Homepage im World Wide Web (WWW): https://www.fww.hs-wismar.de/.

Herausgeber: Prof. Dr. Hans-Eggert Reimers
Fakultät für Wirtschaftswissenschaften
Hochschule Wismar
University of Applied Sciences – Technology, Business and Design
Philipp-Müller-Straße
Postfach 12 10
D – 23966 Wismar
Telefon: ++49/(0)3841/753 7601
Fax: ++49/(0)3841/753 7131
E-Mail: hans-eggert.reimers@hs-wismar.de

Vertrieb: Fakultät für Wirtschaftswissenschaften
Hochschule Wismar
Postfach 12 10
23952 Wismar
Telefon: ++49/(0)3841/753-7468
Fax: ++49/(0) 3841/753-7131
E-Mail: Silvia.Kaetelhoen@hs-wismar.de
Homepage: https://www.fww.hs-wismar.de/

ISSN 1612-0884
ISBN 978-3-942100-60-1
JEL- Klassifikation M14, O32

Printed in Germany

Inhaltsverzeichnis

1 Einleitung 4

2 Cognitive Computing 6
2.1 Kognitionswissenschaft und Kognition 6
2.2 Künstliche Intelligenz 7
2.3 Big Data und Business Intelligence 10
2.4 Cognitive Computing 11
2.5 Zusammenfassung 18

3 Cloud Computing 19

4 Geschäftsprozesse und deren Entwicklung 24
4.1 Grundlagen 24
4.2 Ein angepasstes Vorgehensmodell zur Prozessentwicklung mit Cognitive Computing 32

5 Das kognitive System IBM Watson 35
5.1 DeepQA: Watson als Forschungsprojekt 36
5.2 Watson heute: als kommerzielles Produkt 38
5.3 Technische Rahmenbedingungen der Watson Developer Cloud 39
5.4 APIs und Applikationen der Watson Developer Cloud 41

6 Betriebswirtschaftliche Anwendung von Cognitive Computing 52
6.1 Vorgehensmodell zum Trainieren kognitiver Anwendungen 52
6.2 Routine-Prozesse (Typ III) 55
6.3 Know-how- und entscheidungsintensive Prozesse (Typ I und II) 65

7 Schlussfolgerung und Ausblick 67

8 Literaturverzeichnis 69

1 Einleitung

Künstliche Intelligenz (KI) ist ein Trendthema. Fortschritte bei verfügbarer Rechenleistung, Speicherkapazität und Trainingsdaten haben zu einer Renaissance altbekannter Verfahren geführt (vgl. Bager 2016). Von diesem Trend profitiert auch das Thema Cognitive Computing.

So kündigt der Bundesverband Informationswirtschaft, Telekommunikation und neue Medien e.V. (BITKOM) seit 2015 regelmäßig Durchbrüche und hohe Wachstumsraten für Cognitive Computing an (vgl. Weber 2015b; Shahd 2016; Weber 2017a) und unterstreicht dessen Bedeutung insbesondere für Unternehmen, deren Geschäftsmodell auf dem Sammeln, Verdichten und Interpretieren großer Datenmengen beruht (beispielsweise Banken oder Versicherungen, vgl. Weber 2015a, S. 17). Auf der anderen Seite mahnen kritische Stimmen zur Vorsicht: Das Beratungs- und Marktforschungsunternehmen Gartner Inc. stuft Cognitive Computing in ihrem „Hype Cycle 2017“ als „überzogene Erwartung“ ein und sieht den produktiven Breiteneinsatz erst in fünf bis zehn Jahren (vgl. Panetta 2017).

In einer weiteren Pressemitteilung zu KI konstatiert Gartner Inc. „Hype“ und „AI Washing“: Aufgrund der erreichten Fortschritte und des Interesses an KI-Technologien würden Softwareanbieter beinahe jedes Produkt als KI-Lösung vermarkten beziehungsweise um derartige Funktionen und Bausteine erweitern – ohne darzulegen, welche konkreten betriebswirtschaftlichen Anwendungsmöglichkeiten sich daraus ergeben (vgl. Moore 2017).

Cognitive Computing ist weder in der akademischen noch in der unternehmerischen Welt eindeutig definiert. Das Ziel der Arbeit ist daher die Beantwortung folgender Fragen:

1. Welche Einsatz- und Anwendungsmöglichkeiten bietet Cognitive Computing einem in Deutschland ansässigen Unternehmen?
2. Wie kann Cognitive Computing betriebswirtschaftlich sinnvoll eingesetzt werden?
3. Welche Voraussetzungen müssen für den Einsatz erfüllt sein?

Als Einstieg in das Thema wägt Kapitel 2 verschiedene Definitionen von „Cognitive Computing“ und kognitiven Computersystemen ab, insbesondere deren Eigenschaften und Architektur. Zudem stellt es benötigte Grundlagen aus Kognitionswissenschaft, KI, Business Intelligence (BI) und Big Data dar. Kognitive Computersysteme werden meist als Software-as-a-Service (SaaS) angeboten. Kapitel 3 erörtert Grundlagen des Themas Cloud Computing und die hierfür relevanten, datenschutzrechtlichen Regelungen auf Grundlage der ab dem 25.05.2018 gültigen europäischen Datenschutz-Grundverordnung (DSGVO).

Die betriebswirtschaftliche Anwendung von technischen Neuerungen wie Cognitive Computing erfolgt in Form der Verbesserung von bestehenden (Geschäfts-)Prozessen entweder als revolutionäre Prozesserneuerung oder evolutionäre Prozessverbesserung. Kapitel 4 skizziert die dazu erforderlichen, theoretischen Grundlagen, stellt ein allgemeines Vorgehensmodell zum operativen Prozessmanagement vor und modifiziert es anschließend für die Verwendung kognitiver Computersysteme als Prozessinnovation. Diese Erweiterungen stellen die Qualität der Prozesse sowie die systematische Prüfung der Anforderungen, der Voraussetzungen und der Risiken des kognitiven Systems sicher.

Kapitel 5 konzentriert sich auf das wahrscheinlich bekannteste kognitive Computersystem: IBM (International Business Machines Corporation) Watson, welches 2011 an dem Fernsehquiz „Jeopardy!“ teilnahm und zwei menschliche Spitzenspieler besiegte (vgl. Kramer 2011). Abschnitt 5.1 stellt kurz die Funktionsweise und die Eigenschaften dieser ersten Inkarnation Watsons dar, welche es zu einem kognitiven System im Sinne der festgelegten Definition machen.

Seit 2014 vertreibt IBM die Watson Developer Cloud kommerziell als SaaS. Dabei handelt es sich um eine Sammlung von Application Programming Interfaces (APIs) und Applikationen, beispielsweise zur Erstellung eines Chatbots oder zur Klassifikation von Texten oder Bildern. Abschnitt 5.2 ff. gibt einen Überblick über die Funktionalität sämtlicher Module. Da für dieses Produkt keine Interna veröffentlicht werden, basiert die Darstellung auf den online zugänglichen Dokumentationen und Redbooks von IBM.

Es existieren noch zahlreiche weitere Produkte unter dem Markennamen Watson, denn „bei IBM heißt jetzt alles Watson. Aus dem Jeopardy gewinnenden Computer wurde eine Marke. [...] alles wird ’cognitive’“ (Weber 2017b). Diese Produkte, wie beispielsweise Watson Health zur Unterstützung von medizinischer Krebsbehandlung sowie kognitive Produkte anderer Hersteller werden in dieser Arbeit nicht erörtert.

Kapitel 6 führt die Vorarbeiten zusammen. Da das Training für die Ergebnisse eines kognitiven Systems entscheidend ist, systematisiert Abschnitt 6.1 diesen Vorgang durch Anwendung von Cross Industry Standard Process for Data Mining (CRISP-DM), einem Vorgehensmodell des Data Mining. Anschließend bewertet es die Möglichkeiten, IBM Watson Developer Cloud zur Verbesserung von Routine-Prozessen einzusetzen: durch Ausgliederung von Prozessen an den Kunden, Effizienzsteigerung und Automatisierung von Prozessen oder die Unterstützung bestehender BI-Systeme. Kognitive Computersysteme sollen sich insbesondere zur Unterstützung anspruchsvoller, menschlicher Wissensarbeit eignen, wie sie in Know-how- und Entscheidungsintensiven Prozessen vorkommt (vgl. Hull und Nezhad 2016, S. 4); diese These wird Abschnitt 6.3 einer Überprüfung unterzogen. Damit sind die betriebswirtschaftlichen Einsatzmöglichkeiten von Cognitive Computing dargestellt und das Ziel der Arbeit erreicht.

2 Cognitive Computing

In diesem Kapitel werden zunächst für das Verständnis von Cognitive Computing grundlegende Begriffe und Definitionen der Kognitionswissenschaft, der KI und der BI erläutert.

Die Arbeit nähert sich dem Begriff Cognitive Computing, indem sie ihn zunächst als wissenschaftliche Disziplin auffasst und dessen Verbindungen zu anderen Disziplinen untersucht. Anschließend wird das Ergebnis von Cognitive Computing, kognitive Computersysteme, aus zwei Blickwinkeln betrachtet: „von außen", auf ihre wahrnehmbaren Eigenschaften (Features) hin, und „von innen", also Architektur, Aufbau und Arbeitsweise eines solchen Systems.

2.1 Kognitionswissenschaft und Kognition

Kognition stammt von dem lateinischen Wort „cognoscere" beziehungsweise dem griechischen Wort „gignoskein" (jeweils: erkennen, wahrnehmen, wissen, kennenlernen). Der Begriff fasst primär alle Funktionen zusammen, welche es Menschen[1] ermöglichen, sich intelligent zu verhalten und Probleme angemessen und effizient zu lösen: insbesondere Wahrnehmung, Erkennung, Kategorisierung, Aufmerksamkeit, Erinnerung, Lernen, Schlussfolgerung, Entscheidungsfindung und Kommunikation (vgl. Stephan und Walter 2013, S. 1–4).

Die Kognitionswissenschaft versucht zu ergründen, was komplexe Systeme zu kognitiven Leistungen befähigt (vgl. Stephan und Walter 2013, S. 4; Hurwitz, Kaufman und Bowles 2015, S. 11; und Gudivada 2016, S. 5). Sie betrachtet Kognition als eine Art der Informationsverarbeitung; kognitive Prozesse sind damit letztendlich Berechnungsvorgänge (computational) beziehungsweise Informations- oder Symbolverarbeitung (repräsentational) und bestehen aus spezialisierten Subsystemen, welche miteinander interagieren. Die Kognitionswissenschaft ist ein interdisziplinäres, integratives Forschungsgebiet, insbesondere aus den Disziplinen Anthropologie, Linguistik, Neurowissenschaft, Philosophie, Psychologie und der Informatik, besonders deren Teilgebiet Künstliche Intelligenz.

Ein wichtiger Ansatz in der Kognitionswissenschaft sind Theorien der dualen Entscheidungssysteme (dual process theories) zur Erklärung kognitiver Funktionen wie Entscheidungsfindung und Schlussfolgerung. Sie nimmt die Existenz zweier verschiedener, sich ergänzender Modi zur Informationsverarbeitung an, im Allgemeinen als System 1 und System 2 bezeichnet. System 1 arbeitet schnell und mit geringem Aufwand (Intuition oder „Bauchgefühl"), unterliegt

[1] In dieser Arbeit wird regelmäßig zwischen weiblicher und männlicher Form gewechselt, wertungsfrei als Pars pro Toto.

jedoch keiner bewussten Kontrolle und basiert auf in der Vergangenheit unbewusst verarbeiteten Strukturen (beispielsweise Vorurteilen), ist also fehleranfällig. System 2 ist hingegen langsam, schwerfällig und energieaufwändig, unterliegt aber der willentlichen Steuerung und ermöglicht somit die rationale Entscheidungsfindung. System 1 ist vorgeschaltet und liefert fortlaufend Eindrücke und Abschätzungen, welche nachträglich durch das rationale System überprüft und eventuell korrigiert werden (vgl. Evans 2008, S. 256, 265–268; Magrabi und Bach 2013, S. 277-280; und Kahneman 2015, S. 31–44).

2.2 Künstliche Intelligenz

Cognitive Computing hat ausgeprägte Verbindungen zur KI-Forschung. Bevor diese im Abschnitt 2.4 näher beleuchtet werden, sind zunächst die Künstliche Intelligenz (KI) an sich sowie Begriffe und Techniken dieser Disziplin zu beschreiben.

2.2.1 Grundbegriffe

Künstliche Intelligenz ist ein „Teil der Informatik, [...] der sich mit der Konzeption, Formalisierung, Charakterisierung, Implementation und Evaluierung von Algorithmen befasst, welche Probleme lösen, die bislang nur mit menschlicher Intelligenz lösbar waren“ (Schmidt 2013, S. 44). *Kognitive KI* als Teilgebiet der KI verlangt zusätzlich, dass die verwendeten Systeme und Algorithmen nach ähnlichen Prinzipien funktionieren wie menschliche Informationsverarbeitungsprozesse (vgl. ebenda, S. 44–45), also letztendlich die Erstellung kognitiver Modelle, welche ausgewählte kognitive Leistungen näherungsweise repräsentieren.

Die KI kennt zwei grundlegende Ansätze: In der klassischen beziehungsweise *Symbol verarbeitenden KI* wird das Wissen explizit durch Symbole dargestellt (Wissensrepräsentation). Dies ermöglicht logikbasierte Schlussfolgerungen, das Generieren neuen Wissens und Problemlösung. Entscheidungen beziehungsweise Ergebnisse des Systems sind für menschliche Anwender nachvollziehbar (vgl. Lämmel und Cleve 2008, S. 15–17). Für den Themenkomplex Cognitive Computing sind insbesondere Taxonomien und Ontologien relevant.

Taxonomien ordnen Objekte aus einer Domäne (Ausschnitt aus der realen Welt) in baumartig hierarchisch gegliederte Klassen an. Unterklassen erben alle Eigenschaften ihrer Superklassen. Die Objekte selbst werden nicht näher definiert oder beschrieben (vgl. Pfuhl 2012). Ein Beispiel aus dem Gesundheitswesen ist die ICD-10 (Internationale statistische Klassifikation der Krankheiten

und verwandter Gesundheitsprobleme)[2].

Ontologien sind Wissensmodelle, auf welche sich eine Gruppe von Akteuren geeinigt hat und die eine Domäne konzeptualisieren: Sie beschreiben formal, für einen festgelegten Zweck, die in der Domäne vorkommenden Entitäten und deren Beziehungen zueinander (vgl. Studer 2016; Unland 2012). Ontologien werden häufig im standardisierten Datenmodell Resource Description Framework (RDF) modelliert. Es kennt die Objekttypen Ressource (Entitäten), Property (Eigenschaften von Entitäten) und Statements (Aussagen). Aussagen bestehen wiederum aus Subjekt, Prädikat und Objekt (beispielsweise „Max Mustermann kocht Suppe“) und werden daher Tripel (englisch triple) beziehungsweise 3-Tupel genannt (vgl. Carstensen u. a. 2010, S. 166). Mit SPARQL (SPARQL Protocol and RDF Query Language) steht eine standardisierte Abfragesprache für RDF bereit.

Der zweite Ansatz, die *subsymbolische* beziehungsweise *konnektionistische KI*, beruht hingegen auf sehr vielen einfachen Verarbeitungselementen, welche untereinander verbundenen sind und Informationen austauschen. In den letzten Jahren haben insbesondere Künstliche neuronale Netzwerke (KNNs) an Bedeutung gewonnen, auf welche in Abschnitt 2.2.2 detailliert eingegangen wird.

Beim *maschinellen Lernen* beziehungsweise *Maschinenlernen* (machine learning) handelt es sich um den Versuch, in vorhandenen Daten (Beispielen / Erfahrungen) Muster, Abhängigkeiten oder Anomalien zu erkennen und daraus allgemeingültiges, auf neue Daten anwendbares Wissen zu generieren (vgl. Cleve und Lämmel 2016, S. 14), beispielsweise das Erkennen von Gesichtern in Bildern. Es sind drei Arten des Lernens zu unterscheiden (vgl. Russell und Norvig 2012, S. 811–812):

- Beim *überwachten Lernen* bekommt das KI-System eine Trainings- und eine Lernmenge gegeben und hat die Aufgabe, eine Funktion (Hypothese) zu finden, welche die Trainingsmenge möglichst gut auf die Lernmenge abbildet. Das Ziel wird also von dem Anlernenden vorgegeben. Typischer Anwendungsfall ist die Klassifikation.
- Das *verstärkende Lernen* kann als Sonderfall des überwachten Lernens betrachtet werden. Die Anlernerin gibt das Ziel nicht direkt vor, sondern steuert das System indirekt über Verstärker: Bei gewünschten Ausgaben wird es „belohnt“, bei unerwünschten hingegen „bestraft“ (oder zumindest nicht belohnt). Das System verarbeitet diese Information und passt sich an, um möglichst hohe Belohnungen zu erzielen.
- Beim *unüberwachten Lernen* soll das System hingegen neue, bislang unbekannte Zusammenhänge entdecken, beispielsweise eine Gesamtmenge in Untergruppen aus zueinander möglichst ähnlichen Objekten zerlegen

[2] Online einsehbar unter https://www.dimdi.de/static/de/klassi/icd-10-gm/kodesuche/onlinefassungen/htmlgm2018/index.htm

(Cluster-Analyse, vgl. Cleve und Lämmel 2016, S. 57). Das Ergebnis ist zu Beginn des Lernens unbekannt.

Bedeutsam ist der Unterschied zwischen Daten, Informationen und Wissen (vgl. ebenda, S. 37–38). *Daten* (Plural von Datum im Sinne einer Informationseinheit) sind zunächst eine Ansammlung von Zeichen, zum Beispiel „-3,2“. Zwar gilt für die Zeichen eine Syntax (hier: arabische Ziffern basierend auf dem Dezimalsystem), sie verfügen aber über keinen Kontext. Informationen sind hingegen Daten mit Bezug und damit Bedeutung, beispielsweise „Preiselastizität der Nachfrage Produkt X = -3,2“. Wissen ist eine Information, welche man zu nutzen weiß. Im Beispiel führt eine geringe Preisreduzierung zu einem überproportionalen Anstieg der Nachfrage, da eine sehr elastische Preiselastizität der Nachfrage vorliegt (vgl. Varian 2007, S. 320–325).

Bei Daten sind drei verschiedene Typen zu unterscheiden. *Strukturierte Daten* sind auf definierte Art angeordnet und der Datentyp steht fest, beispielsweise in einer relationalen Datenbank. Die Daten sind somit relativ einfach auszulesen. *Unstrukturierte Daten* weisen hingegen keine formelle Struktur auf (Beispiel: Fließtext), die Analyse ist daher komplex und aufwändig. Dazwischen liegen semi-strukturierte Daten, bei denen nur Teile strukturiert sind; so legen die Multipurpose Internet Mail Extensions (MIME) das Datenformat von E-Mails fest und damit, wo Empfänger, Absender und Text der Mail stehen; der Text selber ist jedoch unstrukturierter Fließtext (vgl. Cleve und Lämmel 2016, S. 37–38).

Natural Language Processing (NLP, maschinelle Sprachverarbeitung) ist ein Gebiet der Linguistik und hat die Verarbeitung natürlicher Sprache auf Computersystemen zum Inhalt. Die Sprachanalyse erfolgt dabei sowohl durch manuell erzeugte, sprachabhängige Regelwerke als auch durch statistische Verfahren und maschinell angelernte Modelle (vgl. Schröder 2013, S. 71–74). Typische Anwendungen des NLP sind die Informationsextraktion (information extraction), beispielsweise zur Erkennung von Eigennamen und deren Relationen zueinander (Named Entity Recognition – NER); die Zuordnung von Wörtern zu Wortarten (Part-of-Speech Tagging) und grammatikalische Analysen (vgl. Carstensen u. a. 2010, S. 594–605).

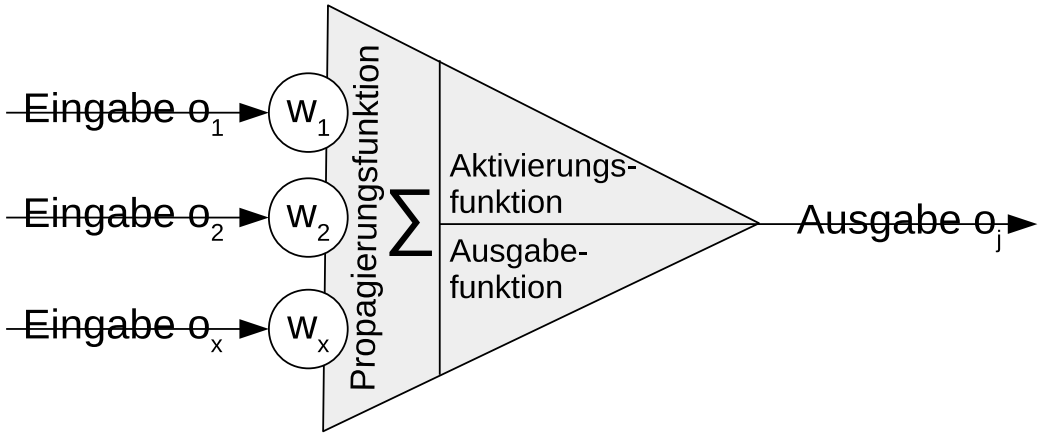
2.2.2 Künstliche neuronale Netzwerke

Künstliche neuronale Netzwerke (KNN) sind der „*Versuch, die in der Gehirnforschung gewonnenen Erkenntnisse über das Zusammenspiel aus Nervenzellen (Neuronen) und deren Verbindungen (Synapsen) zu modellieren*“ (Trinkwalder 2016, S. 131).

Als Verarbeitungselement dient im KNN ein künstliches Neuron. Es nimmt eine oder mehrere, gewichtete Eingaben entgegen und ermittelt daraus eine Ausgabe, siehe Abbildung 1 (vgl. Lämmel und Cleve 2008, S. 195–201; und

Russell und Norvig 2012, S. 845–848).

Abbildung 1: Aufbau eines künstlichen Neurons



Quelle: eigene Darstellung nach Lämmel und Cleve 2008, S. 197

Die Funktionen eines Neurons sind mathematisch einfach zu berechnen. Der Kern liegt im Zusammenschluss vieler einfacher Neuronen über gerichtete und gewichtete Verbindungen oder eine Verarbeitungsfunktion zu einem KNN.

Ein KNN wird typischerweise mit Beispielen trainiert (*Trainingsmenge*). Hierzu wird die Ausgabe des Netzes immer wieder berechnet und nach jedem Durchlauf werden die Verbindungsgewichte modifiziert. Dies geschieht bis das Netz eine zufriedenstellende Qualität erreicht, es die Beispiele der Testmenge hinreichend gut verarbeitet.

Ein KNN kann also Aufgaben lösen, ohne dass der Lösungsweg explizit durch einen Algorithmus beschrieben wird. Das durch Trainieren erlernte Wissen ist implizit im gesamten KNN abgelegt – in der Struktur des Netzes (beispielsweise Anzahl der Neuronen, Anzahl der verdeckten Schichten usw.) und der Gewichtung w_i der verschiedenen Eingänge der Neuronen (vgl. Lämmel und Cleve 2008, S. 15–18, 201–205; oder Haun 2014, S. 54, 79 und 83). Daher kann ein KNN, wie jedes subsymbolische Verfahren, die von ihm gelieferten Ergebnisse nicht begründen.

2.3 Big Data und Business Intelligence

Im Zusammenhang mit Cognitive Computing spielt Big Data eine Rolle, ein ebenfalls nicht allgemeingültig definierter Begriff. Im Rahmen dieser Arbeit werden darunter Massendaten verstanden, welche vier Eigenschaften aufweisen und dadurch im Rahmen „klassischer relationaler Datenbanktechnik“ nicht verarbeitbar sind: hohes Volumen (Tera- bis Zettabyte), hohe Anforderungen an Transport- und Verarbeitungsgeschwindigkeit, hohe Qualität bzw. Genauigkeit

der Daten und die Verwendung verschiedener Datentypen (strukturiert, unstrukturiert, semi-strukturiert). Diese Eigenschaften werden als „die 4 Vs“ (volume, velocity, veracity, variety) zusammengefasst (vgl. Fasel und Meier 2016, S. 5–6; und Hurwitz, Kaufman und Bowles 2015, S. 56–57).

Business Intelligence (BI) fasst Techniken und Architekturen zur Gewinnung, Verwaltung und Verarbeitung des Unternehmenswissens zusammen (vgl. Cleve und Lämmel 2016, S. 3). Auch dieser Begriff ist nicht einheitlich definiert, gängig ist aber die Unterteilung in drei Sichtweisen oder Typen (vgl. Kemper, Mehanna und Baars 2010, S. 1–5). Die BI im engeren Sinne beschränkt sich auf Applikationen, welche unmittelbar zur Entscheidungsunterstützung dienen; insbesondere Online Analytical Processing (OLAP) und Management Information Systems (MIS). Das analyseorientierte BI-Verständnis fügt dem Anwendungen hinzu, bei denen eine Benutzerin interaktiv Analysen vorbereitet oder durchführt; beispielsweise Data Mining (Techniken, Methoden und Algorithmen zur Datenanalyse und Mustererkennung) oder Text Mining (Techniken, Methoden und Algorithmen zur Strukturierung von Texten und Extraktion von Informationen daraus, vgl. Heyer, Quasthoff und Wittig 2012, S. 3–4). Das weite BI-Verständnis erweitert dieses um alle Anwendungen, welche direkt oder indirekt für die Entscheidungsunterstützung eingesetzt werden; beispielsweise ETL-Prozesse (Extrahieren, Transformieren, Laden) und Data-Warehouses (DWHs).

Analytische Informationssysteme (advanced analytics beziehungsweise business analytics) sind Bestandteil des analyseorientierten BI-Verständnisses und umfassen drei Analysetechniken: deskriptive Analysen (descriptive analytics) werten Vergangenheitsdaten aus, prädiktive Analysen (predictive analytics) versuchen aus Vergangenheits- und aktuellen Daten Prognosen für die Zukunft abzuleiten, und präskriptive Analysen (prescriptive analytics) verknüpfen prädiktive Analysen mit Regelwerken und Modellen, um Entscheidungsvorschläge zu unterbreiten oder sogar automatisiert Entscheidungen zu fällen (vgl. Hummeltenberg 2014).

2.4 Cognitive Computing

Cognitive Computing ist weder in der akademischen noch in der unternehmerischen Welt einheitlich definiert. Viele Hersteller bieten ihre jeweils eigene Interpretation an. Gleichwohl kristallisieren sich zwei grobe Gemeinsamkeiten heraus: Cognitive Computing beschreibt Computersysteme, welche lernen, selbstständig Schlussfolgerungen ziehen und mit Menschen oder anderen Maschinen interagieren; und vereint verschiedene Schlüsseltechnologien, insbesondere NLP, Maschinenlernen und KNN (vgl. Zaino 2014; und Hull und Nezhad 2016, S. 7).

2.4.1 Cognitive Computing als Wissenschaftsdisziplin

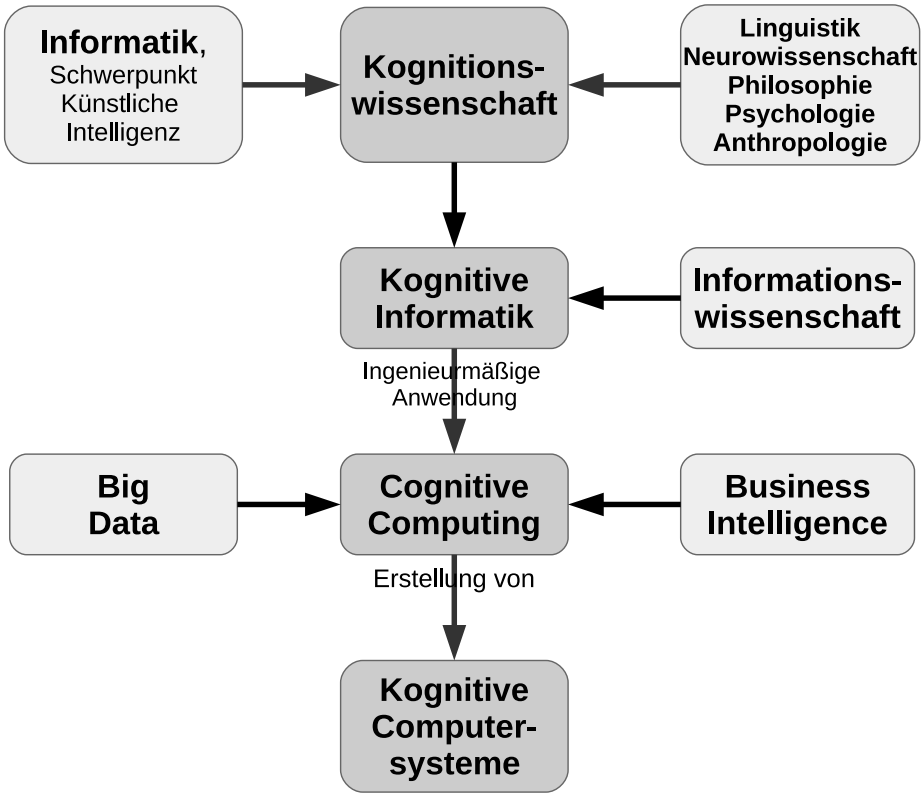
Cognitive Computing kann als die aktuelle Epoche der KI-Forschung, als deren logische und praktische Konsequenz betrachtet werden. Zur Darstellung folgt ein kurzer Abriss der Geschichte der KI-Forschung (basierend auf Haun 2014, S. 107–112; Russell und Norvig 2012, S. 39–52; und Manhart 2017):

- Als Geburtsstunde der KI-Forschung gilt die Dartmouth-Konferenz im Jahre 1956. Die ersten zehn Jahre liegt ihr Schwerpunkt auf Prinzipien und Methoden zur Lösung jedes beliebigen Problems durch Schlussfolgerungen. In diesen Systemen ist jedoch noch kein Wissen hinterlegt.
- Mitte der Sechziger– bis Mitte der Siebzigerjahre wurde dieser Ansatz eines „Universalalgorithmus“ fallen gelassen und spezialisierte Methoden und Techniken entwickelt, mangels Finanzierung durch das Militär oft als Grundlagenforschung in Mikrowelten und ohne realen Bezug („KI-Winter“). Ebenfalls in diesem Zeitraum werden die theoretischen Grundlagen für künstliche Neuronen vorgestellt, finden jedoch aufgrund mangelnder Rechen- und Speicherkapazitäten noch keine praktische Anwendung.
- Mitte der Siebziger– bis Mitte Achtzigerjahre setzt sich die Erkenntnis durch, dass Wissen der Schlüssel zur Problemlösung ist. Folgerichtig werden Techniken zur Wissensrepräsentation und Expertensysteme entwickelt; die Systeme bewegen sich hin zur Lösung „realer“ Probleme.
- Aufgrund der gestiegenen Speicher- und Rechenkapazität erleben KNN Ende der Achtzigerjahre eine erste Renaissance und etablieren sich erstmals als ernstzunehmende Alternative zu Symbol verarbeitender KI, stoßen jedoch wiederum an die Grenzen der verfügbaren Rechen- und Speicherkapazitäten und vor allem der Trainingsdaten.
- Anfang der Neunzigerjahre kommt die verteilte KI auf, ein Ansatz, bei dem viele relativ einfache Agenten miteinander interagieren und gemeinsam komplexe Aufgaben lösen. Ende des Jahrzehntes und Anfang des 21. Jahrhunderts steigt die Anzahl vom kommerziellen Anwendungen von KI-Techniken; zusätzlich werden immer größere Datenmengen verfügbar.
- Etwa 2010 beginnt die aktuelle Epoche des Cognitive Computing. Diese ist durch zwei Faktoren gekennzeichnet. Zum einen erleben KNN ihre zweite Renaissance in Form von „Deep Learning“ – die altbekannte Technik wird durch das Hinzufügen zahlreicher verborgener Schichten erheblich leistungsfähiger. Die hierfür erforderlichen Rechen- und Speicherkapazitäten lassen sich durch den aktuellen Stand der Technik (beispielsweise Parallel Computing, Cloud Computing, verteilte Systeme und In-Memory-Computing) einfach bereitstellen. Darüber hinaus sind auch große Datenmengen zum Trainieren der Netze verfügbar (Big Data). Zum anderen wird die „klassische“ Kognitionswissenschaft im Rahmen

> des Cognitive Computing technologisiert: die Orchestrierung von Erkenntnissen, Werkzeugen und Techniken der beteiligten Disziplinen, um menschliche Intelligenzleistungen per computergestützten Algorithmen als kognitive KI zu modellieren. Auf den Bereich der KI bezogen bedeutet Orchestrierung, dass sowohl Symbol verarbeitende als auch subsymbolische Techniken kombiniert zum Einsatz kommen (vgl. Haun 2014, S. 123).

Andere Autoren sehen Cognitive Computing nicht als direkte Linie der KI-Forschung, sondern als Zusammenführung von KI-Techniken, BI und Big Data (vgl. Hurwitz, Kaufman und Bowles 2015, S. 89; und Gudivada 2016, S. 4) oder als ingenieurmäßige Anwendung der kognitiven Informatik (cognitive informatics). Da die kognitive Informatik wiederum die Disziplinen Informatik, Kognitions- und Informationswissenschaft interdisziplinär vereint (vgl. Wang, Zhang und Kinsner 2010, S. 1–2), lassen sich all diese Ansätze in einer Darstellung zusammenführen (Abbildung 2).

Abbildung 2: Cognitive Computing als Wissenschaftsdisziplin



Quelle: eigene Darstellung, basierend auf Haun 2014, S. 409

2.4.2 Eigenschaften von kognitiven Computersystemen

Computersysteme können in drei Klassen typisiert werden: imperativ, autonom

und kognitiv. Imperative Systeme arbeiten menschengemachte Programme (Algorithmen) ab, welche jeden Verarbeitungsschritt bis ins kleinste Detail festlegen. Autonome Systeme bekommen hingegen lediglich ein Ziel vorgegeben und versuchen selbstständig, dieses zu erreichen (beispielsweise durch Versuch und Irrtum). Kognitive Computersysteme gehen über diese beiden Klassen hinaus, indem sie kognitive Mechanismen wie Wahrnehmung, selbstständige Schlussfolgerung und Lernen beherrschen (vgl. ebenda, S. 5)

Das Cognitive Computing Consortium ist eine 2014 gegründete Arbeitsgemeinschaft, deren Mitglieder größtenteils aus der Industrie stammen und deren Definition von Cognitive Computing beispielsweise von der BITKOM herangezogen wird (vgl. Weber 2015a, S. 14–15). Die Definition geht nicht auf technische Spezifikationen (Innenleben) oder die geschichtlichen Hintergründe ein, sondern verlangt von kognitiven Computersystemen vier zwingende Eigenschaften (vgl. Cognitive Computing Consortium 2014):

- Adaptivität (Lernfähigkeit)
  Kognitive Systeme sind in der Lage, zu lernen und sich an ändernde Informationen, Ziele, und Anforderungen anzupassen; darüber hinaus können sie Mehrdeutigkeiten auflösen und mit Unsicherheiten umgehen. Neue beziehungsweise geänderte Daten werden von ihnen dynamisch in (Beinahe-)Echtzeit verarbeitet.
- Interaktivität
  Kognitive Systeme sind benutzerfreundlich. Sie können darüber hinaus auch mit anderen Systemen (beispielsweise Clouddiensten, Geräten, Sensoren) interagieren.
- Iterativität und Zustandsorientierung
  Kognitive Systeme unterstützen Problemlösungsprozesse, indem sie dem Anwender bei Unklarheiten oder Mehrdeutigkeiten (Rück-)Fragen stellen. Frühere Interaktionen werden gespeichert und fließen in die Problemlösung ein.
- Kontextualität (Kontextabhängigkeit)
  Kognitive Systeme verstehen, identifizieren und extrahieren Kontext und Zusammenhänge (beispielsweise Entitäten und Relationen) aus strukturierten und unstrukturierten Quellen.

Die Definition ist sehr weitgehend, so erfüllt beispielsweise die Internetsuche der Firma Google alle Anforderungen (vgl. Weber 2015a, S. 21). Einen engeren Blick auf die Eigenschaften kognitiver Systeme wirft Hurwitz, Kaufman und Bowles 2015 (S. 1–4):

- Kontinuierliches Lernen
  Kognitive Systeme lernen ohne Neuprogrammierung, indem sie ihr Wissen durch Schlussfolgerungen aus den vorhandenen Daten (strukturiert wie unstrukturiert) erweitern. Sie ahmen Prozesse und/oder Strukturen natürlicher Lernsysteme (insbesondere des menschlichen

Gehirns) nach.

- Modellbildung
  Das System erstellt ein Modell (eine Repräsentation) der Wissensdomäne, auf dem der Lernprozess basiert und welches kontextuelle Erkenntnisse (Zusammenhänge) abbildet. Dazu werden interne und externe Daten mit verschiedensten Algorithmen untersucht, erkannte Muster und Zusammenhänge extrahiert und im Modell abgelegt.
- Hypothesengenerierung
  Hypothesen sind probabilistische Aussagen (Wahrscheinlichkeitsaussagen) und werden auf Grundlage der aktuell vorhandenen Daten und des Modells aufgestellt, überprüft und bewertet. Kognitive Systeme kennen keine absolut richtigen Antworten: ändern sich die Daten und/oder das darauf basierende Modell, verändern sich auch die daraus generierten Hypothesen und deren Bewertung.

Man erkennt viele Gemeinsamkeiten zwischen den beiden Definitionen: kontinuierliches Lernen, Umgang mit Unsicherheiten durch probabilistische Aussagen, Extraktion von Informationen aus strukturierten und unstrukturierten Daten. Letztere Definition verlangt jedoch explizit Modellbildung und die Generierung und Überprüfung von Hypothesen, verzichtet im Gegenzug auf den Aspekt Interaktivität als entscheidendes Merkmal.

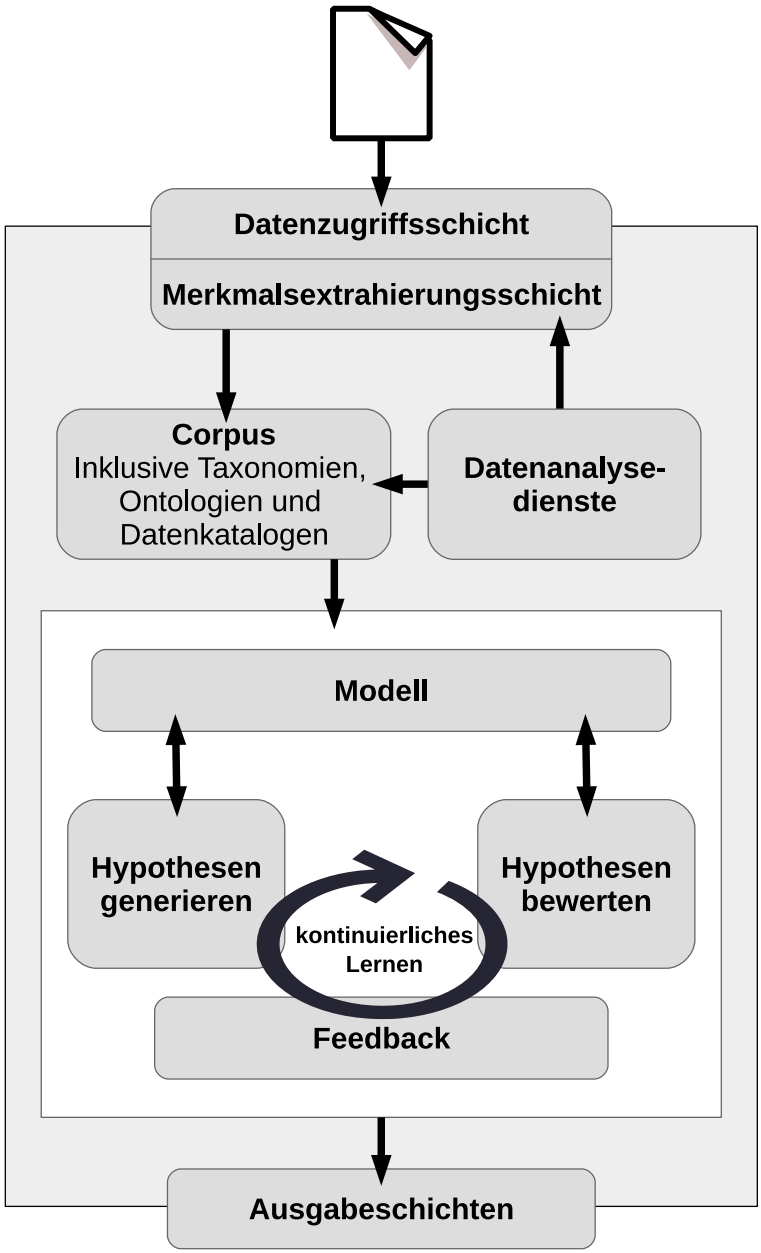
2.4.3 Architektur von kognitiven Computersystemen

Die Architektur eines kognitiven Computersystems ist in Abbildung 3 dargestellt. Nachfolgend werden die einzelnen Elemente näher erläutert (vgl. Hurwitz, Kaufman und Bowles 2015, S. 17–37).

Über die *Datenzugriffsschicht* als Schnittstelle greift das kognitive System auf externe Quellen zu. Unstrukturierte Daten durchlaufen anschließend zwingend die *Merkmalsextrahierungsschicht* (feature extraction) zur Identifizierung, Extraktion und Aufbereitung der relevanten Daten für die Verwendung im Corpus (beispielsweise durch NLP). Erst nach diesem Verarbeitungsschritt sind die Ergebnisse für Datenanalysedienste zugänglich.

Ergebnis ist der Corpus als Wissensbasis beziehungsweise Wissensrepräsentation der aufgenommenen Daten, welcher mit zusätzlichen Ontologien, Taxonomien und/oder Datenkatalogen (data catalogs, enthalten Zeiger auf Daten) erweiterbar ist. Ein kognitives System kann auch durchaus über mehrere spezialisierte Corpora (für verschiedene Domänen oder Themen) verfügen, welche in einem „Meta-Corpus“ aggregiert werden. Im laufenden Betrieb greift das System ausschließlich auf die Corpora, nicht auf externe oder interne Datenquellen zu (Ausnahme: Zeiger, beispielsweise um Speicherplatz zu sparen). Daher ist es möglich, jederzeit neue Datenquellen zu erschließen; der Corpus passt sich dynamisch an.

Abbildung 3: Architektur eines kognitiven Computersystems



Quelle: eigene Darstellung, basierend auf Hurwitz, Kaufman und Bowles 2015, S. 23

Datenanalysedienste (data analytics services) umfassen verschiedene Algorithmen und Verfahren, welche auf die Corpora und aufgenommenen Daten angewendet werden, und ähneln damit den analytischen Informationssystemen der BI. Sie entstammen der Statistik (deskriptiv und inferentiell), dem Data Mining (Techniken, Methoden und Algorithmen zur Datenanalyse und Mustererkennung, vgl. Cleve und Lämmel 2016, S. 3) und dem maschinellen Lernen. Letzteres teilt sich einige Verfahren mit Data Mining, basiert jedoch insbesondere auf iterativer Fehlerminimierung: kontinuierliches, schrittweises Überarbeiten und Verbessern (vgl. Hurwitz, Kaufman und Bowles 2015, S. 92–93).

Auf Grundlage der Informationen im Corpus erzeugt das kognitive System

ein Modell und lernt kontinuierlich. Dies geschieht in drei Schritten:

- Generierung von Hypothesen (Annahmen), entweder aufgrund einer aktiven Aufgabenstellung durch eine Anwenderin oder durch die automatisierte Erkennung von Mustern oder Anomalien im Corpus (Maschinenlernen).
- Bewertung aller aufgestellten Hypothesen durch das Suchen und Bewerten von Hinweisen im Corpus, welche die einzelne Hypothese bestätigen oder widerlegen, und Auswahl der wahrscheinlichsten Hypothese.
- Lernen / Trainieren anhand des Feedbacks, beispielsweise des Anwenders welcher die präsentierte Hypothese bewertet (verstärkendes Lernen). Hat sich eine Hypothese als treffsicher herausgestellt, wird diese übernommen.

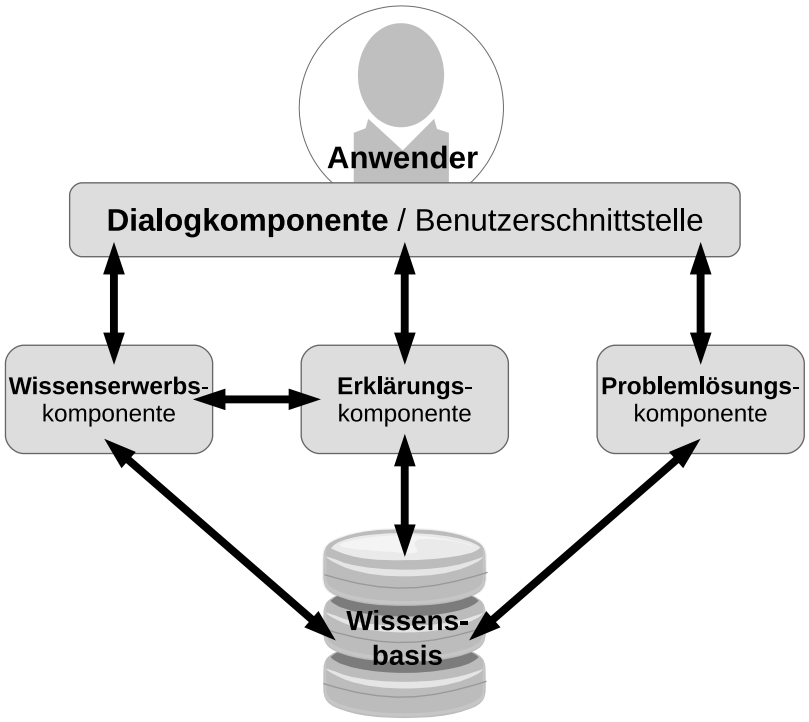
Durch das Hinzufügen neuer und die Wegnahme alter Datenquellen passt sich der Corpus dynamisch an; Erkenntnisse des Modells gehen dadurch aber nicht verloren. Ermittelt das System beispielsweise eine Hypothese, so kann diese durch den Zugriff auf neue Datenquellen vertieft (oder ggf. auch widerlegt) werden.

Als letzte Komponente eines kognitiven Systems sind Ausgabeschichten zu nennen, um mit Anwendern oder abnehmenden Programmen in Kontakt zu treten; beispielsweise über APIs oder Präsentations- und Visualisierungsschichten.

Die skizzierte Architektur folgt dem Aufbau von Expertensystemen (als „klassisches“ Beispiel für wissensbasierte Systeme, vgl. Lämmel und Cleve 2008, S. 31–32, und *Abbildung 4*). Der Corpus stellt die Wissensbasis dar und ist von den Verarbeitungskomponenten klar getrennt. Das Modell inklusive Hypothesengenerierung und -bewertung übernimmt die Aufgaben der Problemlösungskomponente; Datenzugriffsschicht, Merkmalsextrahierung und Datenanalysedienste gleichen der Wissenserwerbskomponente (Aufnahme neuen Wissens) und die Ausgabeschicht(en) der Erklärungs- und Dialogkomponente (Kommunikation mit der Benutzerin).

Darüber hinaus lassen sich Parallelen zum Data Mining ziehen. Das Generieren und Bewerten von Hypothesen ist aus dem Datenanalyse-Zyklus bekannt, welcher zwei Schritte vorsieht (vgl. Neckel und Knobloch 2015, S. 188–190): im ersten Schritt werden auf Grundlage einer Data-Mining-Analyse Zusammenhänge oder Muster in den Daten erkannt (beispielsweise eine Korrelation), also eine Hypothese generiert. Im zweiten Schritt wird diese mit klassischen Analysemethoden detaillierter untersucht und in Konsequenz bestätigt oder widerlegt (Bewertung der Hypothese). Man könnte damit kognitive Computersysteme als die Automatisierung des Data Mining auffassen (vgl. Gudivada u. a. 2016, S. 176–177).

Abbildung 4: Architektur eines Expertensystems



Quelle: eigene Darstellung nach Lämmel und Cleve 2008, S. 31

Eine weitere Parallele ist in der Merkmalsextrahierung zu sehen, welche die Aufgabe des ETL-Prozesses des Data Minings übernimmt: die Umwandlung der operativen Daten, so dass sie dem gewählten Datenanalyseverfahren zugänglich sind (vgl. Kemper, Mehanna und Baars 2010, S. 26–27).

Inwiefern erfüllt die vorgestellte Architektur den Anspruch der kognitiven KI, menschliche Denkprozesse zu imitieren? Das Generieren und Bewerten von Hypothesen orientiert sich an der Theorie der dualen Entscheidungssysteme. Zunächst werden relativ schnell und mit wenig Aufwand zahlreiche Hypothesen generiert (System 1), anschließend in einem aufwändigen Prozess bewertet und verifiziert beziehungsweise falsifiziert (System 2).

2.5 Zusammenfassung

Cognitive Computing beziehungsweise kognitive Computersysteme sind keine neue Erfindung oder gar Revolution – sehr wohl aber eine bedeutende Evolution, welche erreichte Leistungen verschiedener Disziplinen, insbesondere der KI, fortführt. Die entscheidende, kennzeichnende Neuerung ist in der Orchestrierung verschiedenster Werkzeuge, Techniken, Algorithmen und Erkenntnisse zu sehen, mit dem Ziel, dadurch eine neue Qualität von intelligenzähnlichen Leistungen zu erzielen. Als Beispiel sei die Kombination von klassischer und konnektionistischer KI genannt, oder die Verbindung von KI und Business Intelligence (inkl. Big Data und Data Mining). Ob sich der Begriff dauerhaft hält

oder in der nächsten Evolutionsstufe der KI aufgeht, wird die Zukunft zeigen.

3 Cloud Computing

Cloud Computing, übersetzt etwa „Datenverarbeitung in der Wolke“, kann mit dem Schlagwort „Informationstechnologie (IT) aus der Steckdose“ zusammengefasst werden: IT analog zu Strom als jederzeit beziehbares, scheinbar unbegrenzt verfügbares, von Dritten hergestelltes Gebrauchsgut (vgl. Vossen, Haselmann und Hoeren 2013, S. 2). Aus betriebswirtschaftlicher Sicht verspricht Cloud Computing insbesondere Kostenvorteile: Zum einen ermöglicht es Unternehmen durch nutzenbasierte Abrechnung, Fix- in variable Kosten umzuwandeln und so Leerkosten zu vermeiden, zum anderen realisieren Cloud-Service-Provider (CSP) Skaleneffekte und Standortvorteile (vgl. Vossen, Haselmann und Hoeren 2013, S. 32–33; und Sosinsky 2011, S. 15).

3.1 Grundlagen

Die allgemein akzeptierte Definition des Begriffes Cloud Computing stammt vom National Institute of Standards and Technology (NIST) und wurde 2011 veröffentlicht (vgl. Mell und Grance 2011). Danach ist Cloud Computing zusammengefasst ein Modell für jederzeitigen Zugriff auf geteilte Computerressourcen, welche schnell und mit minimalem Aufwand bereitgestellt werden können. Das NIST definiert ausführlicher fünf wesentliche Eigenschaften, drei Servicemodelle und vier Liefermodelle, welche nachfolgend kurz vorgestellt werden.

Cloud Computing zeichnet sich durch fünf wesentliche Merkmale aus (vgl. Mell und Grance 2011, S. 2; und Vossen, Haselmann und Hoeren 2013, S. 21–25).

- On-demand self-service (Selbstbedienung nach Bedarf): Der Kunde kann jederzeit Ressourcen anfordern oder abbestellen. Die Verarbeitung durch den CSP erfolgt unmittelbar und vollautomatisiert.
- Broad Network Access (umfassender Netzwerkzugriff): Die Ressourcen werden über ein Netzwerk bereitgestellt, und der Zugriff erfolgt über standardisierte Protokolle, Formate und Techniken dieses Netzwerkes. Dies bedeutet konkret die Verwendung der Standardprotokolle des Internet, wie zum Beispiel Hypertext Transfer Protocol Secure (HTTPS) und Transmission Control Protocol/Internet Protocol (TCP/IP) und seiner Standardtechniken wie beispielsweise JavaScript Object Notation (JSON), Representational State Transfer (REST) und Extensible Markup Language (XML).
- Ressource Pooling (Zusammenfassen von Ressourcen zu einem gemeinschaftlichen Vorrat): Reale physische Ressourcen werden über

eine Abstraktionsschicht virtualisiert und können als logische Ressourcen beliebig auf die nachfragenden Kunden verteilt werden. Die Virtualisierung ist für den Kunden selbst unsichtbar, aus seiner Perspektive gibt es nur ein System mit scheinbar grenzenlosen Ressourcen.

- Rapid Elasticity (schnelle Elastizität) bezeichnet die jederzeitige Anpassbarkeit an den tatsächlichen Bedarf. In der Realität sind die physischen Ressourcen des CSP selbstverständlich nicht unendlich, können jedoch insbesondere durch horizontale Skalierung (scale out, Hinzufügen weiterer Server) schnell angepasst werden.
- Measured Service (Messung): Clouddienste werden in der Regel nach Nutzung abgerechnet, daher der Ressourcenverbrauch gemessen und protokolliert.

Das NIST definiert drei Servicemodelle, welche zusammen den SPI-Ordnungsrahmen bilden (vgl. Mell und Grance 2011, S. 2–3; und Vossen, Haselmann und Hoeren 2013, S. 27–30). Das Servicemodell legt fest, wie die Verantwortung zwischen Nutzer und CSP verteilt ist (vgl. Sosinsky 2011, S. 9–10).

- Infrastructure-as-a-Service (IaaS): Der CSP stellt ausschließlich Ressourcen wie virtuelle Maschinen, Rechenzeit oder Speicherplatz zur Verfügung, welche der Nutzer nach Belieben verwendet und verantwortet.
- Platform-as-a-Service (PaaS): Der CSP stellt eine standardisierte Plattform bereit, unter der der Cloudnutzer eigenständig Applikationen erstellt und betreibt. Dies umfasst neben den Ressourcen auch Bibliotheken, Hilfsprogramme, Dokumentationen, Frameworks und APIs für häufig benutzte Funktionen wie Persistenz oder Datenbank-Zugriff. Ein Beispiel ist die Google App Engine für Webanwendungen und mobile Applikationen[3].
- Software-as-a-Service (SaaS): Der CSP stellt eine gebrauchsfertige Anwendung für Endbenutzer bereit und kümmert sich vollständig um deren Betrieb (inklusive Wartung und Fehlerbeseitigung).

In der Praxis gibt es noch weitere Varianten (beispielsweise Database as a Service), häufig unter der Bezeichnung Anything-as-a-Service (XaaS) zusammengefasst.

Das Liefermodell (deployment model) legt fest, inwiefern die Cloud nach Außen geöffnet ist, also wer sie nutzen kann (vgl. Mell und Grance 2011, S. 3; und Vossen, Haselmann und Hoeren 2013, S. 30–32).

- Eine öffentliche Cloud (public cloud) wird vom CSP betrieben und steht jedermann kostenlos oder gegen Entgelt zur Verfügung. Die Abrechnung erfolgt in der Regel nach Nutzung.

[3] sh. https://cloud.google.com/appengine/

- Die nicht-öffentliche Cloud (private cloud) wird hingegen ausschließlich durch eine einzige Organisation genutzt. Sie wird entweder von der Organisation selbst betrieben (interne private Cloud) oder von einem CSP dediziert für diese Organisation bereitgestellt.
- Gemeinschaftliche Clouds (community cloud) sind ebenfalls nicht-öffentlich, werden allerdings von mehreren Organisationen mit gleichartigen Anforderungen genutzt. Ein Beispiel ist die Trusted German Insurance Cloud (TGIC), welche vom Gesamtverband der deutschen Versicherungswirtschaft e.V. (GDV) betrieben wird und dem Datenaustausch dient (vgl. Engelage 2015). Auch dieses Liefermodell kann von der Gemeinschaft selbst oder von einem externen CSP bereitgestellt werden.
- Eine hybride Cloud verbindet zwei oder mehr verschiedene Clouds für den Datenaustausch. Beispielsweise wird eine hoch ausgelastete, interne private Cloud mit einer öffentlichen Cloud verbunden, um Lastspitzen abzufedern.

3.2 Datenschutz

In Deutschland sind zwei gesetzliche Vorschriften einschlägig. Zum 25.05.2018 ist die europäische Datenschutz-Grundverordnung (DSGVO, vgl. Europäische Union 2016, Seiten 32 ff.) in Kraft getreten; wo sie Öffnungsklauseln für einzelstaatliche Regelungen vorsieht, ist noch das Bundesdatenschutzgesetz (BDSG, vgl. Bundesministerium der Justiz und für Verbraucherschutz 1990) relevant (vgl. Wybitul 2016, Seiten 5, 27).
Bei Prüfung des Anwendungsbereiches ist zwischen sachlichen und räumlichen Kriterien zu unterscheiden:

- aus sachlicher Sicht gelten die Bestimmungen für nicht-öffentliche Stellen, sobald diese personenbezogene Daten in einem Dateisystem speichern oder ganz oder teilweise automatisiert verarbeiten (vgl. Plath 2016, Art. 2 DSGVO Rz. 1–7); aufgrund der Verbreitung von IT und EDV damit de facto für jedes betriebswirtschaftliche Unternehmen.
- In räumlicher Sicht führt die DSGVO das sogenannte Marktortprinzip ein. Danach spielt es keine Rolle mehr, wo ein Unternehmen seinen Sitz hat oder seine Daten speichert – die Vorschriften sind einschlägig, sobald es seine Waren oder Dienstleistungen innerhalb der Europäischen Union (EU) anbietet (vgl. Buchner 2016, Seite 156).

Grundkonzept der DSGVO ist das Verbot mit Erlaubnisvorbehalt. Jede Verarbeitung personenbezogener Daten ist untersagt – es sei denn, der Betroffene hat wirksam eingewilligt oder es liegt ein Erlaubnistatbestand in Form einer Rechtsvorschrift vor (vgl. ebenda, Seiten 157– 159). Unter Verarbeitung fällt jeder Vorgang im Zusammenhang mit personenbezogenen Daten, sowohl automatisiert als auch manuell; beispielsweise Erheben, Speichern, Verändern und

Übermitteln an Dritte (vgl. Plath 2016, Art. 4 DSGVO Rz. 12).

3.2.1 Welche Daten sind geschützt?

Die DSGVO schützt ausschließlich personenbezogene Daten (vgl. Plath 2016, Art. 4 DSGVO Rz. 4–8). Dabei handelt es sich um alle Informationen, welche sich auf eine identifizierte oder identifizierbare lebende, natürliche Person beziehen. Ein Sonderfall sind die *besonderen personenbezogenen Daten*: sensible Angaben zu rassischer und ethnischer Herkunft, politischen Meinungen, religiösen oder weltanschaulichen Überzeugungen, Gewerkschaftszugehörigkeit, Sexualleben, Gesundheit oder sexueller Orientierung; außerdem genetische oder biometrische Daten. Sie unterliegen einem noch strengeren Schutz, so muss beispielsweise die Einwilligung des Betroffenen diese Daten explizit aufführen (vgl. ebenda, Art. 9 DSGVO Rz. 1–13).

Für nicht-personenbezogene Daten besteht kein Schutz durch DSGVO oder BDSG. Verlieren personenbezogene Daten ihren Bezug durch Anonymisierung, Pseudonymisierung oder Aggregation, verlieren sie ebenfalls den Schutz des Gesetzes und können frei verarbeitet werden.

Bei der *Anonymisierung* werden die Daten derartig verändert, dass die Zuordnung entweder unwiederbringlich verloren geht (zum Beispiel durch Löschen oder Weglassen von Informationen) oder nur unter unverhältnismäßigem Aufwand an Zeit, Kosten und Arbeitskraft wiederherstellbar ist. Die Verschlüsselung von personenbezogenen Daten gilt regelmäßig nicht als Anonymisierung – dies ist nur im Einzelfall unter Betrachtung der konkreten Umstände und der verwendeten Algorithmen und Schlüssellängen denkbar (vgl. Arbeitskreise Technik und Medien der Konferenz der Datenschutzbeauftragten des Bundes und der Länder 2014, Seiten 12–13).

Im Zuge der *Pseudonymisierung* werden Identitätsmerkmale durch fingierte Werte ersetzt (beispielsweise „KundenID4711“ statt „Stefan Schulze“). Im Gegensatz zur Anonymisierung existiert zwischen Identifikationsmerkmal und Pseudonym eine Zuordnungsfunktion, welche die Umkehrung des Prozesses und Wiederherstellung des Personenbezugs erlaubt (vgl. Plath 2016, Art. 4 DSGVO Rz. 18–19).

Bei der *Aggregation* werden die Einzelangaben mehrerer Betroffener zusammengefasst, um beispielsweise Durchschnittswerte oder ähnliche Kennziffern zu errechnen. Sind jedoch trotz Aggregation noch Rückschlüsse auf einzelne Gruppenmitglieder möglich (beispielsweise aufgrund unzureichender Größe oder ungünstiger Zusammensetzung der Gruppe), bleibt der Personenbezug und damit der gesetzliche Schutz erhalten (vgl. Simitis 2014, §3 BDSG Randnr. 14).

3.2.2 Einwilligung und wichtige Ausnahmeregelungen

Aufgrund des Verbotes mit Erlaubnisvorbehalt darf ein Unternehmen personenbezogene Daten nur dann an einen Dritten (hier den CSP) übermitteln, wenn die Betroffene wirksam eingewilligt hat oder ein gesetzlicher Ausnahmetatbestand vorliegt. Wird zu Datenübermittlung das Internet genutzt, ist die Übertragung mit einer geeigneten Transportverschlüsselung (zum Beispiel TLS – Transport Layer Security) abzusichern, da sonst die Gefahr besteht, dass Dritte die Information mitlesen (was eine einwilligungspflichtige Übermittlung darstellt).

Eine wirksame Einwilligung muss ohne Zwang, für den konkreten Fall, in Kenntnis der Sachlage und als ausdrückliche Zustimmung (opt-in) durch den Betroffenen erteilt werden: der Betroffene muss in der Lage sein, eine informierte und freie Entscheidung zu treffen.

Als gesetzliche Ausnahmeregelung kommt insbesondere die *(privilegierte) Auftragsdatenverarbeitung (ADV)* in Betracht. Sie erlaubt, eine dritte Partei (Auftragnehmer, hier der CSP) mit der Verarbeitung personenbezogener Daten zu betreuen, ohne dass die Betroffene einwilligen muss: die Weitergabe der Daten gilt nicht als Übermittlung. An dieses Privileg sind jedoch verschiedene Voraussetzungen geknüpft. Insbesondere ist ein schriftlicher und zahlreichen gesetzlichen Anforderungen genügender Vertrag zu schließen; darüber hinaus muss der Auftragnehmer weisungsgebunden sein und regelmäßig kontrolliert werden (vgl. Wybitul 2016, 69–71 und 127–128). Mit Wirksamwerden der DSGVO ist erstmals die ADV auch in Drittländern außerhalb der Mitgliedstaaten des Europäischen Wirtschaftsraumes (EWR) oder der EU prinzipiell möglich. Da dies jedoch an weitere Bedingungen und Prüfungen geknüpft ist (sh. Artikel 44-50 DSGVO), empfiehlt es sich, mit dem CSP regionale Beschränkungen zu vereinbaren.

Eine weitere Ausnahmeregelung stellt die „General-Klausel“ der DSGVO dar (vgl. Wybitul 2016, Seiten 90–93, 174–175). Diese erlaubt die Nutzung personenbezogener Daten, sofern eine Abwägung zwischen den Interessen der betroffenen Person (insbesondere deren Grundrechte und -freiheiten) gegenüber den berechtigten Interessen des datenverwertenden Unternehmens (zum Beispiel dessen Geschäftsmodell, Senkung von Kosten, Erzielung von Gewinnen) zu Gunsten des Verwerters ausgeht. Dabei sind alle Umstände des jeweiligen konkreten Einzelfalles zu berücksichtigen, wobei verfassungsrechtliche Grundrechte ihrer Natur nach schwerer wiegen als die kommerziellen Interessen eines Unternehmens. Ein mögliches Gegengewicht kann darin bestehen, dass die Betroffene personenbezogene Daten freiwillig selbst veröffentlicht, zum Beispiel öffentlich zugängliche Einträge in sozialen Netzwerken, Foren oder Websites. Für derartig veröffentlichte, besondere personenbezogene Daten sieht die DSGVO hingegen einen expliziten Erlaubnistatbestand vor (vgl. ebenda, Seiten 102–103, 178).

4 Geschäftsprozesse und deren Entwicklung

Dieses Kapitel stellt Grundlagen des operativen Prozessmanagements und dessen Kernaufgabe, die kontinuierliche Verbesserung von Prozessen, dar. Neben Grundbegriffen wird ein allgemeines Vorgehensmodell vorgestellt und dieses anschließend für den Einsatz eines kognitiven Computersystems als Prozessinnovation erweitert.

4.1 Grundlagen

Aus systemtheoretischer Sicht besteht ein Prozess aus verschiedenen Elementen einer Organisation (Aufgaben, Aufgabenträgern, Sachmitteln und Informationen), welche durch mindestens eine der folgenden Beziehungsarten miteinander verbunden sind. Leitungsbeziehungen ordnen Aufgabenträger und Sachmittel Aufgaben zu, Ablaufbeziehungen verknüpfen Aufgaben miteinander und Informations- und Kommunikationsbeziehungen verbinden Aufgabenträger, Sachmittel und Aufgaben über den Austausch von Informationen (vgl. Schallmo und Brecht 2014, S. 14–17). Aus produktionstheoretischer Sicht besteht ein Prozess aus einer Folge von Aktivitäten, welche festgelegte Eingaben (Input) zu einem definierten Arbeitsergebnis (Output) transformieren (vgl. Schmelzer und Sesselmann 2013, S. 51).

Bei einem Geschäftsprozess steht die Erfüllung von Kundenbedürfnissen im Vordergrund: er "*besteht aus der funktions- und organisationsübergreifenden Folge wertschöpfender Aktivitäten, die von Kunden erwartete Leistungen erzeugen und die aus der Geschäftsstrategie und den Geschäftszielen abgeleiteten Prozessziele erfüllen*" (ebenda, S. 52). Geschäftsprozesse beginnen und enden beim (internen oder externen) Kunden, welcher Anforderungen aufstellt und die Ergebnisse (Output beziehungsweise Leistung) in Form von Sach-, Dienst-, Informationsleistungen oder beliebigen Kombinationen daraus (Lösungen) erhält. Zur Ausführung des Prozesses werden Inputs (Produktionsfaktoren) wie beispielsweise Arbeitsleistung, Rohstoffe oder Informationen benötigt (vgl. Schmelzer und Sesselmann 2013, S. 52–55; und Posluschny 2016, S. 21).

Zur Beurteilung der Performance beziehungsweise Leistung eines Prozesses (nicht zu verwechseln mit Prozessleistung im Sinne von Output) ist die Unterscheidung zwischen Effizienz und Effektivität bedeutsam (vgl. Schmelzer und Sesselmann 2013, S. 3–4; und Posluschny 2016, S. 16–17). Prozesseffizienz bedeutet, das gesetzte Prozessziel unter möglichst geringem Ressourceneinsatz zu erreichen (Ökonomisches Prinzip in der Minimal-Ausprägung); in der Literatur oft auf die griffige Formulierung „die Dinge richtig tun" gebracht. Prozesseffektivität liegt hingegen vor, wenn der Prozess die richtigen Ziele verfolgt („die richtigen Dinge tut"), also die Bedürfnisse und Erwartungen des Kunden korrekt erkannt und bedient werden. Prozesse mögen noch so effizient sein:

wenn sich keine oder zu wenige Abnehmer für die Leistung finden, kann das Unternehmen keinen Erfolg haben (vgl. Schmelzer und Sesselmann 2013, S. 56, 273–278).

Geschäftsprozesse können in zwei Kategorien unterteilt werden (vgl. ebenda, S. 65–73). Primäre Geschäftsprozesse stiften unmittelbaren Nutzen für den Kunden, beispielsweise die Fertigung eines Smartphones oder die Schadenabwicklung im Rahmen eines Versicherungsvertrages. Hier erfolgt die originäre Wertschöpfung, da der Kunde bereit ist für diese Leistung zu bezahlen. Sekundäre Geschäftsprozesse stiften dem Kunden zwar keinen direkten Nutzen, sind aber für die Funktion der Primärprozesse unabdingbar. Sie haben entscheidenden Einfluss auf die Qualität der Primärprozesse und der dort erzeugten Leistungen. Typische Beispiele sind die Bereitstellung von Personal und IT in der benötigten Qualität und Quantität (Unterstützungs- oder Bereitstellungsprozesse) beziehungsweise Controlling und Strategieplanung (Unternehmenssteuerungs- oder Managementprozesse).

Darüber hinaus sind drei Typen von Geschäftsprozessen zu unterscheiden (vgl. Schmelzer und Sesselmann 2013, S. 69–72; und Hull und Nezhad 2016, S. 5–7). In der Praxis sind die Übergänge zwischen den drei Typen fließend.

- Typ III (Routine-Prozesse)
  Diese Prozesse sind detailliert beschrieben, transparent, stabil und stark strukturiert, so dass relativ wenige Entscheidungen zu treffen sind. Es bestehen damit vergleichsweise geringe kognitive Anforderungen (das benötigte Wissen weist relativ geringen Umfang und Dynamik auf), dafür laufen die Prozesse sehr häufig ab. Ziel des Prozessmanagements für Typ III-Prozesse ist typischerweise eine hohe Effizienz, also die termin-, kosten- und qualitätsgerechte Erreichung der Prozessziele mit möglichst geringem Ressourceneinsatz. Aufgrund dieser Eigenschaften sind Typ III-Prozesse besonders gut für Automation geeignet. Typische Beispiele sind Lohn- und Gehaltsrechnung oder Lieferverkehr.
- Typ I (Know-how-intensive Prozesse)
  Typ I-Prozesse sind das Gegenteil der Routine-Prozesse: auch wenn es verbreitete Best Practices gibt, ist der Prozessablauf sehr offen (da oft von Zwischenergebnissen abhängig), nur grob strukturiert und planbar. Es bestehen hohe kognitive Anforderungen an die Prozessteilnehmer und deren Erfahrung, Kreativität und Wissen, dafür wird der Prozess nur selten instanziiert. Der Fokus des Prozessmanagements liegt insbesondere auf der Unterstützung und Verbesserung des Wissensaustauschs, der gemeinsamen kreativen Arbeit und der Ideenfindung und somit auf Steigerung der Effektivität zur Erreichung verwertbarer Ergebnisse. Typische Beispiele sind Design oder Strategieplanung.

- Typ II (Entscheidungs-intensive Prozesse)
  Typ II-Prozesse liegen mit ihren Merkmalen zwischen Typ III- und Typ I-Prozessen. Es handelt sich um operationale Tätigkeiten mit mittleren kognitiven Anforderungen, bei denen Urteilsvermögen und Entscheidungen eine Rolle spielen; beispielsweise Verkauf oder Projektmanagement.

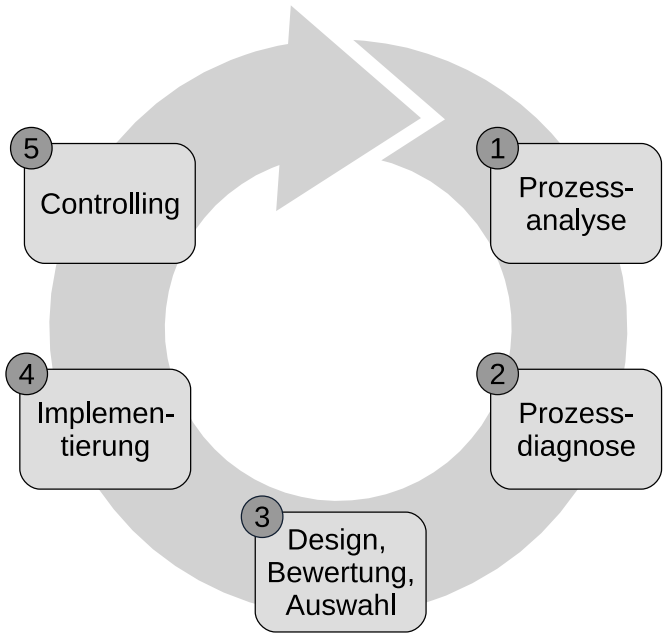
Das Prozessmanagement „umfasst grundsätzlich alle Tätigkeiten zur Planung, Steuerung und Überwachung von Prozessen“ (Schallmo und Brecht 2014, S. 19). Im Rahmen der Prozessgestaltung werden Prozesse aus der Strategie des Unternehmens abgeleitet, identifiziert und modelliert (dessen Ablauf- und Aufbauorganisation festgelegt). Das Prozesscontrolling (alternativ Prozesslenkung) legt die Ziele der Prozesse fest, überwacht anhand von Kennzahlen deren Erfüllung und steuert bei Bedarf gegen. Bei der Prozessentwicklung oder Prozessoptimierung steht hingegen die Identifikation und Bewertung von Verbesserungspotentialen im Mittelpunkt. Die Verbesserung der Prozessleistung im Sinne von nachhaltiger Erhöhung der Effektivität und/oder Effizienz ist das Hauptziel des Prozessmanagements. Die eigentliche Umsetzung der Maßnahmen erfolgt im Rahmen der Prozessgestaltung und -lenkung (vgl. Schmelzer und Sesselmann 2013, S. 8–10; und Schallmo und Brecht 2014, S. 19–21). Im weiteren Verlauf der Arbeit wird der Begriff „Prozessentwicklung“ verwendet: Optimierung kann irreführend sein, da nicht das Erreichen des mathematischen Optimums gemeint ist.

Die Prozessentwicklung kann unterschiedliche Formen annehmen (vgl. Schmelzer und Sesselmann 2013, S. 407–413). Bei der Prozesserneuerung handelt es sich um eine revolutionäre, grundlegende und radikale Veränderung des Prozesses, in der Regel im Rahmen eines Projektes und mit Business Process Engineering als Methode. In der Praxis stößt dieses Vorgehen oft auf starke Widerstände. Der zweite Ansatz ist die Prozessverbesserung, inkrementelle Performancesteigerungen unter Beibehaltung der Grundstruktur des Prozesses als Evolution im Tagesgeschäft. Verbreitete Methoden sind beispielsweise Kaizen und Six Sigma. Manche Autoren sehen noch einen dritten Weg zwischen diesen beiden Polen namens Prozesstransformierung für nicht-radikale Änderungen, welche jedoch für das Tagesgeschäft zu umfangreich sind und im Rahmen kleinerer Projekte stattfinden (vgl. Hierzer 2017, S. 108–109, 191). Im Rahmen dieser Arbeit werden diese unter Prozessverbesserung subsumiert.

Es gibt zahlreiche Vorgehensmodelle zum operativen Prozessmanagement, welche sich zwar in den Begrifflichkeiten, Grafikstilen und der Anzahl der Schritte unterscheiden, aber letztendlich drei Gemeinsamkeiten aufweisen: erstens die Aufteilung in drei aufeinanderfolgende Phasen (Prozessplanung, -realisierung und -einführung), zweitens die rationale Entscheidungsfindung durch Dokumentation des Ist-Zustandes, Definition des Soll-Zustandes sowie Ermitt-

lung, Bewertung und Auswahl von Handlungsalternativen in der Planungsphase; und drittens die Durchführung einer Grobplanung, welche anschließend je nach Bedarf feiner detailliert wird. Das folgende, allgemeine Vorgehensmodell (sh. Abbildung 5) trägt allen diesen Punkten Rechnung (vgl. Fischermanns 2013, S. 216–221; und Amberg, Bodendorf und Möslein 2011, S. 61–66) und wird in den folgenden Abschnitten detaillierter erläutert.

Abbildung 5: Vorgehensmodell für operationales Prozessmanagement



Quelle: eigene Darstellung, basierend auf Amberg, Bodendorf und Möslein 2011, S. 62 und Fischermanns 2013, S. 216.

4.1.1 Prozessanalyse

Ziel der Prozessanalyse ist die wertneutrale Dokumentation seines Ist-Zustandes und seiner Kennzahlen (vgl. Fischermanns 2013, S. 309). Hierzu sind folgende Schritte erforderlich (vgl. Best und Weth 2010, S. 60–82):

- Identifizieren beziehungsweise Ausgrenzen
  Legt die existierenden Prozesse mit deren jeweiligen Start– und Endpunkten, dem benötigten Input und dem gelieferten Output fest. Hierfür stehen zwei grundsätzliche Vorgehensweisen zur Verfügung: der Top Down-Ansatz leitet Soll-Prozesse aus der Geschäftsstrategie am sprichwörtlichen „grünen Tisch“ ab, der Bottom-up-Ansatz basiert hingegen auf der bestehenden Aufbauorganisation (vgl. Schmelzer und Sesselmann 2013, S. 139–143).

- Erheben und Dokumentieren
  Der Prozess wird in seine einzelnen Bearbeitungsschritte aufgebrochen, deren Reihenfolge sowie die jeweils beteiligten Organisationen, genutzte

Hard- und Software, benötigten Inputs und gelieferten Outputs erfasst. Für die weiteren Analysen sind darüber hinaus der Informationsfluss und die statistische Verteilung der Prozessvariationen hilfreich. Der Grad der Detaillierung ist von den Anforderungen und den durchgeführten Prozessdokumentationen und -entwicklungsmaßnahmen abhängig, grundsätzlich ist von grob nach fein vorzugehen.

- Prozessmodellierung
  Die Prozesse werden (in der Regel grafisch) modelliert, also vollständig, formal, präzise und konsistent beschrieben (vgl. ebenda, S. 148), beispielsweise in Business Process Model and Notation (BPMN) 2.0.
- Messung der Prozesse
  Die Auswahl der zu messenden Kennzahlen hängt von den Zielen der Organisation ab. Fünf typische Kernziele sind Prozesszeit, -kosten, -qualität und -termintreue als Kennzahlen für die Prozesseffizienz sowie die Kundenzufriedenheit als Maß für dessen Effektivität. Zwischen diesen Zielen bestehen Interdependenzen (insbesondere das Trilemma zwischen Prozessqualität, Durchlaufzeit und Prozesskosten, vgl. Fischermanns 2013, S. 445) , so dass die Ziele zu gewichten sind (vgl. Schmelzer und Sesselmann 2013, S. 273–276).
  Die *Prozesszeit* wird meist in Form der Durchlaufzeit erfasst. Diese misst die verstrichene Zeit zwischen Start- und Endpunkt des Prozesses, summiert also sämtliche Bearbeitungs-, Warte-, Transport- und Liegezeiten (vgl. Fischermanns 2013, S. 310, 571). Kurze Durchlaufzeiten sind in der Regel komplementär zu Termintreue und Kundenzufriedenheit, jedoch konfliktär zu Prozessqualität und -kosten (vgl. Schmelzer und Sesselmann 2013, S. 280).
  Die *Prozesskosten* bewerten die Personal- und Sachkosten für die Leistungserstellung monetär. Sie sind neben dem erzielbaren Preis der entscheidende Faktor der Wertschöpfung und damit meist hoch priorisiert; können aber gleichwohl konfliktär zu Kundenzufriedenheit und Prozessqualität sein.
  Hohe *Prozessqualität* äußert sich in leichter Beherrschbarkeit des Prozesses (durch hohe Transparenz und niedrige Komplexität, beispielsweise wenigen Variationen) und niedrigen Fehlerraten. Fehler liegen vor, wenn eine Gruppe von geforderten Merkmalen (Qualität) nicht erfüllt wird (vgl. Fischermanns 2013, S. 310–325). Hohe Prozessqualität ist komplementär zu einer hohen Qualität der Leistung und damit der Kundenzufriedenheit, kann jedoch konfliktär zu Prozesskosten oder Durchlaufzeiten sein. Typische Kennzahlen sind Ausschussraten oder der Anteil an Reparaturen während der Gewährleistungspflicht.
  Die *Prozesstermintreue* steht in enger Beziehung zur Prozesszeit, bezieht

sich jedoch nicht auf die Laufzeit des Prozesses sondern auf einen bestimmten, dem Abnehmer zugesagten Zeitpunkt, zu dem der Prozess abgeschlossen sein muss. Einfach beherrschbare, fehlerfreie Prozesse unterstützen die Termintreue, so dass dieses Ziel meist komplementär zur Prozessqualität ist (vgl. Schmelzer und Sesselmann 2013, S. 279–280).
Kundenzufriedenheit liegt vor, wenn die vom Kunden wahrgenommenen Leistung seine Erwartungshaltung übersteigt. Zur Ermittlung bietet sich die Auswertung des Beschwerdemanagements an; da sich aber nur wenige unzufriedene Kunden auch tatsächlich beschweren, stellen regelmäßige Befragungen (beispielsweise mittels Feedbackbögen, im Rahmen telefonischer Kontakte usw.; vgl. ebenda, S. 277–279, 297–302) sinnvolle Ergänzungen dar.

4.1.2 Prozessdiagnose

In diesem Schritt werden die Stärken und Schwächen des Prozesses erkannt und dokumentiert, Ursache-Wirkungs-Analysen durchgeführt, gefundene Probleme dargestellt und priorisiert sowie Ziele für die überarbeiteten Prozesse formuliert (vgl. Best und Weth 2010, S. 84–101; und Fischermanns 2013, S. 379–412).

- Stärken-/Schwächen-Analyse
  Zur Bewertung der Stärken und Schwächen wird ein Maßstab benötigt, letztendlich eine Vorstellung davon, was der jeweilige Soll-Wert ist. Anstöße können beispielsweise aus dem Beschwerdemanagement, dem betrieblichen Vorschlagswesen, Branchenbenchmarks, Kundenbefragungen oder aus Checklisten beziehungsweise Prüfkatalogen (letztendlich Best Practices, Beispiele sh. Fischermanns 2013, S. 385–392; oder Best und Weth 2010, S. 85–92) kommen. Ein ungünstiges Verhältnis von Stärken zu Schwächen spricht für eine radikale Prozesserneuerung; ein günstiges hingegen für die evolutionäre Prozessverbesserung.
- Ursache-/Wirkungsanalyse
  In diesem Schritt werden die festgestellten Schwachstellen (Probleme) in Ursachen und Wirkungen unterschieden und in Beziehung gesetzt. Dies stellt sicher, dass die Ursachen angegangen und nicht nur Symptome kuriert werden. Es stehen zahlreiche grafische Darstellungsmöglichkeiten zur Verfügung, beispielsweise als grafisches Netzwerk, Ishikawa-Diagramm (Fischgräten-Diagramm) oder in Form von Mindmaps.
  - Problemdarstellung und –priorisierung
    Die gefundenen Ursachen und deren Auswirkungen werden dargestellt und festgelegt, welche Probleme in welcher Reihenfolge angegangen werden. Als Kriterien werden Wichtigkeit, Dringlichkeit und Aufwand-/Nutzenverhältnis herangezogen, wobei die Dringlichkeit existierende Termine, Fristen und die Dynamik des Problems im Sinne seiner

Veränderungstendenzen spiegelt und die Wichtigkeit die Priorität der davon betroffenen Unternehmensziele sowie Anzahl und Größe der betroffenen Organisationseinheiten. Für die Untersuchung bietet sich die ABC-Analyse an. Sie basiert auf dem Pareto-Prinzip (80-zu-20-Regel), welches besagt, dass 80% der Ergebnisse mit 20% des Einsatzes erzielt werden, und ordnet Objekte in die drei Klassen A, B und C (mit absteigender Wichtigkeit) ein (vgl. Hierzer 2017, S. 117–120). In der Praxis spielen insbesondere „Quick Wins" eine Rolle, Sofortmaßnahmen mit geringen bis mittleren Gewinnen, welche jedoch schnell und mit geringem Aufwand umsetzbar sind. Auch wenn sie in Krisen schnelle erste Hilfe leisten können, stehen hier insbesondere politische Gründe wie die Einbindung der Betroffenen, Schaffung einer positiven Grundeinstellung oder Demonstration von Schaffenskraft im Vordergrund.

- Formulierung der Ziele (Soll-Werte)
  Die erkannten und priorisierten Probleme werden konkret quantifiziert. Nur so ist die Messung des Erfolgs und darauf aufbauend die Festlegung neuer Ziele möglich.

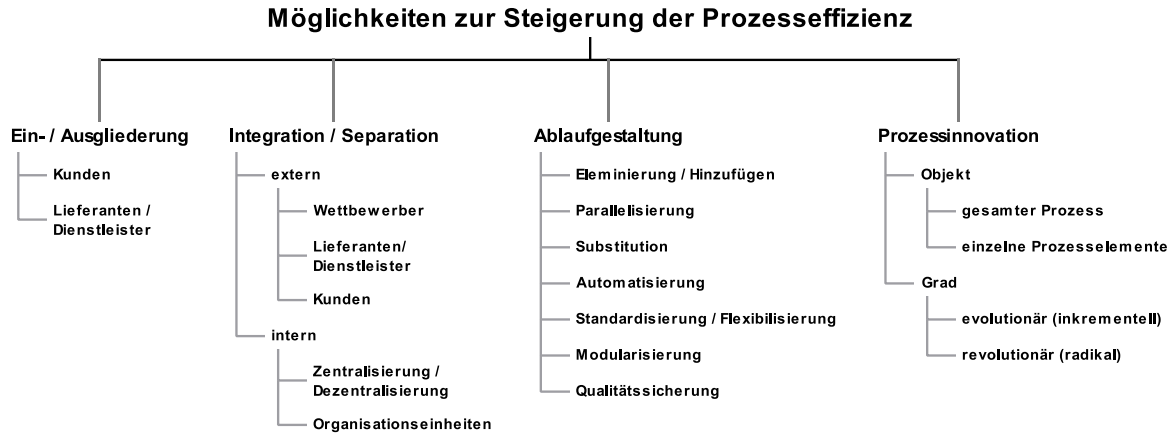
4.1.3 Design, Bewertung, Auswahl eines Prozesses

Zur Effizienzsteigerung eines Prozesses stehen zahlreiche Gestaltungsmöglichkeiten zur Verfügung. Abbildung 6 auf der nächsten Seite zeigt einen Überblick über einige Optionen (vgl. Posluschny 2016, S. 26–51; und Best und Weth 2010, S.136–139).

Unter *Prozessausgliederung* fällt die Verlagerung eines Prozesses nach extern, in Form von Outsourcing an Dritte oder die Verlagerung an den Kunden. Insbesondere das Internet hat für die Einbindung des Kunden völlig neue Möglichkeiten eröffnet, beispielsweise Online-Vertragsabschlüsse, Selbstbedienung (Selfservice) in Online-Kundenbereichen oder zur Informationsbeschaffung (vgl. Posluschny 2016, S. 32–34). Durch die Loslösung von Servicezeiten und die Vermeidung von Wartezeiten kann die Durchlaufzeit stark verkürzt werden. Geeignete Prozesse sind standardisiert, werden oft instanziiert und sind vom Kunden leicht zu durchschauen. Die *Prozesseingliederung* stellt das Gegenstück zur Ausgliederung dar.

Bei der externen *Prozessintegration* werden Lieferanten, Wettbewerber oder Kunden in den Wertschöpfungsprozess einbezogen, um allen Beteiligten Doppelarbeiten, Kontrollkosten, Lagerkosten usw. zu ersparen. Eine Prozessintegration kann auch ausschließlich unternehmensintern stattfinden, durch Zusammenfassen von Prozessschritten in einer Organisationseinheit oder der Konzentration von Aufgaben und Kompetenzen (Zentralisierung). Die *Prozessseparation* ist das jeweilige Gegenteil.

Abbildung 6: Möglichkeiten zur Steigerung der Prozesseffizienz



Quelle: eigene Darstellung, basierend auf Posluschny 2016, S. 26 und Best und Weth 2010, S. 136–139.

Der *Prozessablauf* kann mannigfaltig beeinflusst werden, beispielsweise durch Entfernen von nicht wertschöpfenden Prozesselementen, der Substitution von Elementen durch solche mit höherer Wertschöpfung oder auch dem Hinzufügen völlig neuer Elemente, um dem Abnehmer zusätzliche Mehrwerte zu bieten. Weitere Möglichkeiten sind die Änderung der Reihenfolge oder die Parallelisierung von Elementen, die Standardisierung (beispielsweise durch Nutzung von extern zugekaufter Standardsoftware statt Eigenentwicklungen) und dessen Gegenstück Flexibilisierung, und schließlich die Automatisierung, worunter die Unterstützung oder sogar der Ersatz manueller Tätigkeiten durch IT-Lösungen oder Maschinen verstanden wird.

Prozessinnovationen sind am Markt oder im Unternehmen eingeführte Neuerungen, welche die effizientere Herstellung des Outputs erlauben und somit auf die Verbesserung des eigenen wirtschaftlichen Erfolges abzielen. Prozessinnovationen können sowohl auf einzelne Prozesselemente als auch den gesamten Prozess in einer evolutionären oder radikalen Weise betreffen (vgl. Schallmo und Brecht 2014, S. 21–23). Typisches Beispiel sind Neuerungen im IT-Bereich. Die Konsequenz aus Prozessinnovationen sind in der Regel Veränderungen in der Ablaufgestaltung (beispielsweise erlauben technische Fortschritte die Automatisierung oder Substitution alter Prozesselemente), gleichwohl ist die Ablaufgestaltung auch ohne Innovationen möglich und wird daher als gesonderter Punkt betrachtet.

Abschließend sind die entworfenen Prozessvarianten durch Vergleich mit den Zielvorgaben zu bewerten und der Prozessvorschlag mit dem höchsten Zielerreichungsgrad auszuwählen (vgl. Fischermanns 2013, S. 443–445). Wenn ein Prozess auch nach Überarbeitung keinen ausreichenden Nutzen für die Zielerreichung stiftet, bleibt als letztes Mittel nur noch der Prozessabbau.

4.1.4 Implementierung des neuen Prozesses

Der ausgewählte Prozess wird realisiert, getestet und im Produktivbetrieb implementiert. Die Einführung ist parallel zu bestehenden Prozessen, übergangslos oder in Stufen möglich; letzteres kann sich sowohl auf den Umfang (beispielsweise einzelne Prozesselemente), als auch auf die betroffenen Organisationseinheiten (beispielsweise als Pilotbetrieb innerhalb einer Organisationseinheit) beziehen (vgl. ebenda, S. 446-450).

Der neue Prozess wird anhand der festgelegten Kennzahlen überwacht, welche wiederum als Ausgangspunkt (Ist-Zustand) für weitere Entwicklungsmaßnahmen dienen. Damit schließt sich der Kreis und die kontinuierlich operative Prozessgestaltung beginnt von vorne.

4.2 Ein angepasstes Vorgehensmodell zur Prozessentwicklung mit Cognitive Computing

Grundsätzlich kann die Prozessverbesserung mit Cognitive Computing im Rahmen des in Abschnitt 4.1 vorgestellten Vorgehensmodells (siehe Abbildung 5) durchgeführt werden. Diese Arbeit schlägt jedoch ein erweitertes Vorgehensmodell, welches den besonderen Rahmenbedingungen beim Einsatz einer cloudbasierten, kognitiven Lösung durch zwei Erweiterungen Rechnung trägt.

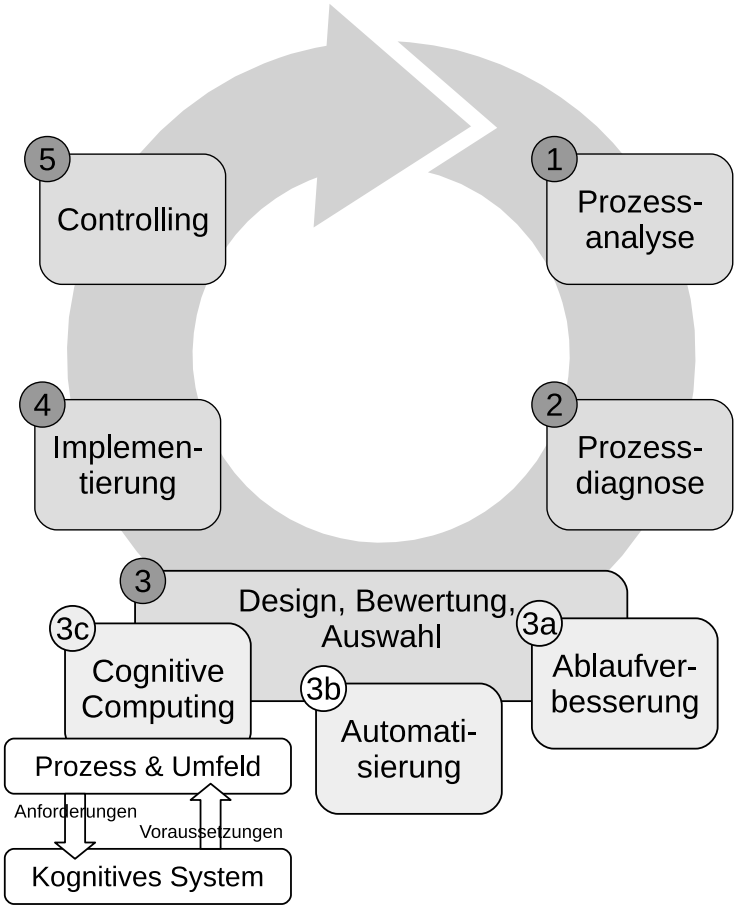
Als erste Erweiterung sieht das erweiterte Vorgehensmodell drei neue Unterschritte in der Designphase vor (vgl. Abbildung 7 auf der nächsten Seite). Die restlichen Phasen des Modelles bleiben unverändert.

Im ersten Schritt Ablaufverbesserung ist der grundsätzliche Prozessablauf und dessen Qualität zu hinterfragen und mit den „klassischen“ Methoden der Ablaufgestaltung (vgl. Abschnitt 4.1.3, beispielsweise Parallelisierung, Reihenfolge der Elemente oder Eliminierung von nicht wertschöpfenden Elementen) zu verbessern. Dies liegt darin begründet, dass der Prozessablauf – analog zum Fundament eines Gebäudes – entscheidend für die Qualität der darauf aufbauenden Konstruktion ist. Beispielsweise können nicht wertschöpfende Genehmigungsschritte im Prozess oder eine hohe Anzahl an Prozessvariationen den Erfolg von Automatisierung oder Cognitive Computing einschränken oder gar verhindern: qualitativ schlechte Prozesse bleiben auch nach Automatisierung qualitativ schlechte Prozesse (vgl. Best und Weth 2010, S. 155–158; und Hierzer 2017, S. 95–96, 112–117).

Aufgrund dieser Überlegungen ist die Automatisierung als zweiter Schritt vorgesehen, obwohl es sich dabei streng genommen ebenfalls um eine „klassische“ Methode der Ablaufgestaltung handelt. Erfolgreiche Automatisierung setzt weiterhin voraus, dass der Prozess häufig instanziiert wird (sonst stehen Aufwand und Ertrag in keinem guten Verhältnis) und wenige Prozessvariationen vorsieht. Es bietet sich an, die automatische Verarbeitung an Regeln zu binden, umso weniger gebräuchliche Varianten auszuschließen (vgl. Best und

Weth 2010, S. 155–158).

Abbildung 7: Erweitertes Vorgehensmodell für operationales Prozessmanagement mit Cognitive Computing



Quelle: eigene Darstellung

Sind diese beiden Schritte erfolgreich absolviert, ist im dritten und letzten Schritt der Einsatz von Cognitive Computing als Prozessinnovation zu prüfen. Hier findet sich die zweite Erweiterung des Vorgehensmodelles: die systematische Erfassung und Prüfung der Anforderungen aus dem Prozessdesign an das kognitive System, und der Anforderungen des kognitiven Systems an den Prozess und dessen Umfeld (nachfolgend als Voraussetzungen bezeichnet). Dies ist insbesondere deswegen erforderlich, weil kognitives Systeme in der Regel als Cloudlösung angeboten werden und nicht im lokalen Rechenzentrum (on premise) installierbar sind. Beim operativen Prozessmanagement handelt es sich um einen zirkulären Prozess. Im Rahmen der Designphase ist unter anderem festzulegen, welche Modifikationen in dieser Iteration umgesetzt und welche zu einem späteren Zeitpunkt angegangen werden. Hierbei ist die im Modell vorgesehene Reihenfolge Ablaufverbesserungen, Automatisierung und Cognitive Computing aus den genannten Gründen einzuhalten. Gleichwohl dürfen diese drei Schritte nicht voneinander losgelöst betrachtet werden. Vielmehr ist

eine umfassende, integrative Analyse im Rahmen der übergeordneten Designphase gefragt, um Mehrfacharbeiten zu vermeiden und spätere Schritte vorzubereiten. Ergibt die Prozessanalyse beispielsweise, dass zunächst zur Ablaufverbesserung ein bestehendes Programmmodul zu erweitern ist, empfiehlt es sich gleich die Erfordernisse möglicher späterer Ausbaustufen zu berücksichtigen (beispielsweise die Einführung von REST-Schnittstellen) oder zumindest Vorbereitungen zu treffen.

Folgende Anforderungen aus dem Prozessdesign an das kognitive System sind zu berücksichtigen:

1. Funktionalität
   Neben dem Funktionsumfang der einzelnen Module ist hierbei auch deren potenzielle Zusammenarbeit von Interesse und wie diese zur Behebung der diagnostizierten Prozessschwäche beitragen können.
2. Sprachunterstützung
   Viele Werkzeuge unterstützen entweder ausschließlich englischsprachige Texte oder bieten nur für diese Sprache den vollen Leistungsumfang. Da der Umfang cloudbasierter Dienste ständiger Veränderung unterliegt, empfiehlt sich ein Blick in die jeweils aktuellen Dokumentationen.
3. (Laufende) Verfügbarkeit
   CSP sagen in der Regel vertraglich eine Mindestverfügbarkeit zu (Service Level Agreement SLA, beispielsweise 99,5% Verfügbarkeit bei 24h-Betrieb). Die Folgen von Verletzungen sind in den Allgemeinen Geschäftsbedingungen (AGB) geregelt, oft handelt es sich dabei lediglich um geringe Maluszahlungen. Daher ist eine systematische Risikobetrachtung für den Fall erforderlich, dass der Dienst temporär ausfällt.
4. Kritikalität
   Dieser Punkt ist eng mit dem Vorgänger verwandt: hier wird das Risiko eines dauerhaften Ausfalls des Dienstes betrachtet, sei es, weil der Dienst eingestellt wird, oder der CSP das Geschäft aufgibt oder insolvent wird.

Kritikalität und Verfügbarkeit sind gemeinsam unter Einsatz einer Risikomatrix zu bewerten. Diese sammelt erkannte Risiken, schlüsselt sie nach Eintrittswahrscheinlichkeit und Wirkung auf und erlaubt so eine Bewertung und Steuerung der Risiken (vgl. Schmelzer und Sesselmann 2013, S. 390-392). Wichtige Bewertungskriterien stellen dar: ist der Ausfall durch den Kunden wahrnehmbar? Existieren Überbrückungsmöglichkeiten, wenn ja: mit welchem Aufwand an Zeit und Geld sind sie verbunden? Wie kritisch ist der Prozess für das Unternehmen?

Folgende Voraussetzungen an Prozess und Umfeld sind für den Einsatz eines kognitiven Systems relevant:

1. Technische Rahmenbedingungen

Hierunter fallen alle Bedingungen und Anforderungen zum Datentausch mit dem kognitiven System, insbesondere die verwendeten Schnittstellen, Formate und Programmiersprachen. Im Gegensatz zur einer Outsourcing-Lösung bieten CSP standardisierte Dienste an, an welche sich das Unternehmen anpassen muss. Neben erhöhten Aufwänden birgt dies auch die Gefahr von Lock-in-Effekten: Abhängigkeiten, welche den Wechsel des Anbieters verkomplizieren oder unmöglich machen, beispielsweise durch aufwändigen Export / Import der Daten (vgl. Vossen, Haselmann und Hoeren 2013, S. 27, 104–106).

2. Benötigte Daten
   Die benötigten Daten sind zu erfassen und zu bewerten: unterliegen diese den Datenschutzvorschriften? Wenn ja, liegt die Einwilligung der Betroffenen vor (beziehungsweise kann dies arrangiert werden) oder ist die Anwendung einer Ausnahmeregelung möglich, insbesondere die Vereinbarung einer ADV wie in Abschnitt 3.2.2 beschrieben? Sofern die Daten Geschäftsgeheimnisse oder Unternehmensinterna bergen (beispielsweise beim Einlesen von Dokumenten für eine Wissensdatenbank), ist das damit verbundene Risiko zu bewerten und zu beurteilen.
3. Benötigtes Training (Anlernen des Systems)
   Sofern das System mit eigenen Daten trainiert werden muss, sind zahlreiche Punkte zu beachten. Diesen wird in Abschnitt 6.1 mit einem eigenen Vorgehensmodell Rechnung getragen.

Das folgende Kapitel stellt die Module der Watson Developer Cloud vor. Dabei werden Anforderungen (Funktionalität, Sprachunterstützung und SLAs) und Voraussetzungen (technische Rahmenbedingungen, benötigte Daten, benötigtes Training) zunächst allgemeingültig aufgezeigt; die Konkretisierung auf den jeweiligen Einsatzzweck erfolgt in Kapitel 6.

5 Das kognitive System IBM Watson

IBM Watson (benannt nach dem IBM-Gründer Thomas J. Watson, vgl. Ferrucci u. a. 2010, S. 78) ist das wahrscheinlich bekannteste kognitive Computersystem. Es nahm 2011 an dem Fernsehquiz „Jeopardy!“ teil (für einen Überblick über das Quiz und dessen Regeln vgl. ebenda, S. 61) und besiegte die beiden Spitzenspieler Ken Jennings und Brad Ruttner (vgl. Kramer 2011). Das Event und die Gegner wurden dabei von IBM bewusst als gut vermarktbare und öffentlichkeitswirksame „Grand Challenge“ ausgewählt, analog zu IBMs DeepBlue, welcher 1997 als erster Schachcomputer einen Schachweltmeister besiegte (vgl. Kelly und Hamm 2014, S. 27–30).

Dieses Kapitel wirft zunächst einen Blick auf Architektur und Funktionsweise dieses Systems und zeigt auf, inwiefern es sich um ein kognitives System im Sinne der Definition aus Kapitel 2 handelt. Anschließend stellt es den Funk-

tionsumfang und die technischen Anforderungen der heute kommerziell verfügbaren Variante „Watson Developer Cloud“ dar.

5.1 DeepQA: Watson als Forschungsprojekt

Die Idee zu Jeopardy als „Grand Challenge“ wurde innerhalb IBM erstmals 2004-2005 diskutiert. Ernsthaft angegangen wurde das Projekt unter dem Namen DeepQA im Sommer 2007, allerdings mit eher geringen Erwartungen: in zwei Jahren sollte das System mittelmäßige Spieler schlagen, in vier Jahren eine 25%ige Chance gegen Spitzenspieler haben (vgl. Baker 2012, S. 20–22, 49, 80). Um die beiden ausgewählten Spitzenspieler zu besiegen, muss das System ca. 70% der Fragen innerhalb von drei Sekunden beantworten und in 80% der Fälle richtigliegen. Dabei sind lediglich 2% der Fragen durch eine Tripel-Suche lösbar (vgl. Ferrucci u. a. 2010, S. 60,65,70).

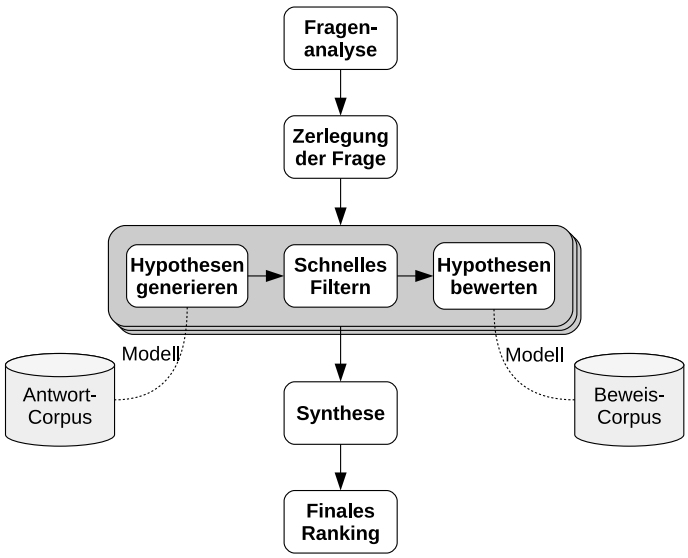
Das System setzt insbesondere auf die Parallelisierung von Expertenmodulen; hierfür stehen 2.880 Prozessorkerne (90 IBM Power 750 Server à 4 Prozessoren à 8 Kerne) und 15 Terabyte (TB) Arbeitsspeicher bereit. Während des Quiz hat das System keinen Zugriff auf das Internet (vgl. Kramer 2011, S. 55; und Hurwitz, Kaufman und Bowles 2015, S. 140).

Ein Blick auf den Systemaufbau (siehe Abbildung 8 auf der nächsten Seite) zeigt die Parallelen zu der in Abschnitt 2.4.3 auf Seite 11 dargestellten Architektur kognitiver Systeme: Im Kern generiert und bewertet Watson Hypothesen (hier „candidate answers“ genannt) auf Grundlage eines Modells, welches wiederum auf einem Corpus basiert. Als Besonderheit verfügt das System über jeweils einen Corpus zur Generierung und zur Bewertung von Hypothesen. Quizfragen werden wie folgt verarbeitet (vgl. Ferrucci u. a. 2010, S. 69–74):

- Analyse der Frage durch verschiedene, parallellaufende Algorithmen. Neben typischen linguistischen Analysen (NLP), unter anderem der Erkennung von Entitäten (Named Entity Recognition – NER) und Relationen, kommen auch spezielle Module für Jeopardy! zum Einsatz, beispielsweise zur Erkennung besonderer Fragenarten wie Puzzles oder Rechenaufgaben.
- Fragenzerlegung: Verschachtelte oder kombinierte Fragen werden in ihre Bestandteile zerlegt. Die folgenden drei Schritte werden für jeden Bestandteil parallel durchlaufen und die Ergebnisse ggf. anschließend wieder zusammengefasst.
- Hypothesengenerierung: Hypothesen werden durch den Einsatz mehrerer Suchwerkzeuge (information retrieval) auf den Corpus gewonnen. Es kommen unter anderem Apache Lucene für unstrukturierte Dokumente und SPARQL für strukturierte Wissensdatenbanken zum Einsatz (vgl. Gliozzo u. a. 2017, S. 35).
- Schnelle Filterung: mittels schnell zu berechnender Verfahren werden die etwa 100 erfolgversprechendsten Hypothesen ausgesiebt und der

Rest verworfen.

Abbildung 8: Systemarchitektur DeepQA



Quelle: eigene Darstellung, basierend auf Ferrucci u. a. 2010, S. 69

- Hypothesenbewertung: das System sucht nach Beweisen für die verbliebenen Hypothesen und bewertet diese. Hierzu stehen über 50 verschiedene Expertenmodule mit jeweils eigenen Ansätzen und Prüfungen (beispielsweise temporale oder räumliche Nähe) bereit, welche als Ergebnis jeweils einen Konfidenzwert liefern.
- Synthese: mehrfach vorkommende Hypothesen, beispielsweise von Form von Synonymen, werden zusammengeführt.
- Finales Zusammenführen und Bewerten: Als letzter Schritt werden die errechneten Konfidenzen aller Module für jede Hypothese gewichtet zusammengefasst (auf Grundlage eines maschinell angelernten Modells) und so das Gesamt-Ranking ermittelt.

Andere Komponenten des Systems, wie Jeopardy!-spezifische Spieltaktiken, werden an dieser Stelle nicht weiter vertieft.

Die Befüllung der Corpora erfolgte sowohl manuell als auch automatisiert über eine Datenzugriffsschicht (vgl. Ferrucci u. a. 2010, S. 69). Im ersten Schritt wurden manuell Fragen aus vergangenen Folgen analysiert und passende strukturierte und unstrukturierte Quellen für die Antworten (Problemraum) in den Corpus aufgenommen. Diese dienten als Ausgangsbasis (seed documents) für den automatisch ablaufenden zweiten Schritt, der Suche nach verwandten Dokumenten im Internet. Aus diesen wurden kleine Textauszüge extrahiert (Merkmalszugriffsschicht), bewertet und dem Corpus hinzugefügt (Datenanalyseschicht). Dieses zweistufige Vorgehen (kleine Startmenge plus automatisierte Erweiterung) wird als Bootstrapping bezeichnet (vgl. Heyer, Quasthoff und Wittig 2012, S. 260). Im dritten Schritt wurde der Corpus nochmals manuell

überarbeitet und umstrukturierte und unstrukturierte Daten wie Taxonomien ergänzt.

Zusammengefasst orchestriert IBM Watson in seiner ersten Inkarnation als DeepQA verschiedene Techniken und Algorithmen der Sprachverarbeitung, Informationserschließung (information retrieval) und der KI (beispielsweise Wissensrepräsentation und Maschinenlernen), um damit intelligenzähnliche Leistungen zu erzielen (vgl. Schmidt 2013, S. 46). Die Generierung und Bewertung von Hypothesen auf Grundlage eines Modells ist an die im Kapitel 2 vorgestellten Dual-Process-Theorien der menschlichen Kognition angelehnt. Sowohl in seiner Architektur als auch in seinen Eigenschaften erfüllt DeepQA alle Anforderungen an ein kognitives Computersystem.

Gleichwohl handelt es sich um eine schwache KI: das System ist nicht in der Lage, tatsächlich zu denken oder seine Aktionen zu verstehen. DeepQA ist ein Hochleistungsrechner („Number Cruncher"), der sehr schnell sehr viele Berechnungen parallel ausführen kann und daher in der Lage ist zu agieren, als ob er intelligent wäre (vgl. Russell und Norvig 2012, S. 1176; und Baker 2012, S. 148–153).

5.2 Watson heute: als kommerzielles Produkt

Watson wurde jahrelang entwickelt, ohne Einsatzmöglichkeiten jenseits der „Grand Challenge" Jeopardy in Betracht zu ziehen. Erst im August 2010 begann die Suche nach neuen Aufgabengebieten für das System (vgl. Baker 2012, S. 192). Konzentrierte sich IBM zunächst auf prestigeträchtige Mammutaufgaben und Großprojekte wie beispielsweise die Krebsforschung oder das Gesundheitswesen (vgl. Ferrucci u. a. 2013), erweiterte es seine Geschäftsstrategie 2014 und bietet seitdem die Watson Developer Cloud an, eine cloudbasierte Sammlung von SaaS und APIs (vgl. Gliozzo u. a. 2017, S. 42–43). Im Laufe der Zeit sammelten sich noch zahlreiche weitere Watson-Produkte, welche jedoch in dieser Arbeit nicht eruiert werden.

Kritiker werfen IBM vor, das Potential der Technologie nicht auszuschöpfen und die Marke Watson zu verwässern. Das kommerzielle Produkt hätte nur wenig mit dem Jeopardy! spielenden System gemeinsam (vgl. Waters 2016); IBMs Marketing wecke überzogene Erwartungen, während der Markt noch nach sinnvollen Anwendungsfeldern suche (vgl. Kroker 2017).

IBM sieht drei grundsätzliche Einsatzmöglichkeiten für ihr Produkt (vgl. Gliozzo u. a. 2017, S. 12–13):

- „Discovery": Zusammenhänge und Erkenntnisse in großen Informationsmengen entdecken. Das gleiche Ziel verfolgt Data Mining als Bestandteil der BI.
- „Decision": Entscheidungsunterstützung aufgrund von Wissen und erkannten Zusammenhängen. Auch hier lassen sich Parallelen zur BI ziehen, da diese ebenfalls MIS umfasst.

- „Engagement“: neue Interaktionswege zwischen Maschinen und Menschen eröffnen und Fähigkeiten der Anwender durch kognitive Unterstützung ergänzen/ausbauen.

Ein wichtiger Unterschied beim kommerziellen Einsatz ist die Begrenzung auf eine Wissensdomäne. Dies erlaubt, häufig gestellte Fragen (Frequently Asked Questions – FAQ) fest modelliert in der Wissensdatenbank zu hinterlegen, so dass nicht erst Hypothesen generiert und bewertet werden müssen. Auch hier greift wieder das Pareto-Prinzip, wonach sich etwa 80 % der Fragen mit weniger als zwanzig Antworten erledigen lassen (vgl. Gliozzo u. a. 2017, S. 46–51).

5.3 Technische Rahmenbedingungen der Watson Developer Cloud

Die IBM Watson Developer Cloud wird als SaaS im Rahmen der IBM Cloud (ehemals Bluemix) angeboten, welche verschiedene IaaS und SaaS umfasst. IBM bietet drei Liefermodelle an. Bei „Shared“ und „Premium“ handelt es sich um öffentliche Clouds, welche auf der gleichen Hardware betrieben werden. „Premium“ räumt dem Kunden getrennte virtuelle Maschinen und Container ein, kostet dafür jedoch mehr und bietet nur einen verringerten SLA von 99,5% (im Vergleich zu 99,95%) im 24h-Betrieb. Bei Verletzung des SLA ist eine Rückerstattung der Gebühren bis zu 25% möglich, jedoch keine darüber hinausreichenden Schadenersatzansprüche (vgl. IBM Deutschland GmbH 2017e). Bei dem Liefermodell „Dedicated“ handelt es sich um eine dediziert private Cloud, deren Konditionen nur auf Anfrage erhältlich sind.

Die Installation der Watson Developer Cloud on premise (lokal im eigenen Rechenzentrum) ist derzeit nicht möglich. Als Transportverschlüsselung kommt TLS (Transport Layer Security) zum Einsatz, ungeschützte Verbindungen werden nicht akzeptiert (vgl. IBM Deutschland GmbH 2016d, S. 1–4). Die APIs der Watson Developer Cloud basieren auf der REST-Architektur und setzen JSON für den Datenaustausch ein.

Representational State Transfer (REST) ist ein Architekturstil für verteilte Systeme, welcher auf bereits existierende Standards des Internets (HTTPS und URI – Uniform Resource Identifier) aufsetzt. Er findet insbesondere in Webdiensten Anwendung und weist sechs zentrale Eigenschaften auf (vgl. Fielding 2000, 45–48 und 76–97).

- Client-Server-Architektur: Aktionen werden von Clients ausgelöst, und von einem kontinuierlich laufenden Serverprozess beantwortet (vgl. Tanenbaum und van Stehen 2008, S. 55–56). Dies hat den Vorteil, dass beliebige Clients zum Einsatz kommen und Server und Client unabhängig voneinander entwickelt werden können.
- Zustandslosigkeit: Jede Transaktion zwischen Client und Server steht für sich. Der Server verwaltet keine Sitzungen, folglich muss die An-

frage des Clients sämtliche benötigten Daten enthalten. Die daraus resultierenden Vorteile – geringe Serverbelastung, einfache horizontale Skalierbarkeit, Stabilität – werden mit einer höheren Netzwerkbelastung durch den multiplen Transport der Daten erkauft.

- Verwendung von Caches (Zwischenspeichern) auf Client-Seite, um Zugriffe auf den Server zu minimieren; dazu markiert der Server in seiner Antwort cachbare Daten.
- Einheitliche Schnittstellen: jeder REST-Dienst ist unter einer festgelegten Adresse erreichbar, versteht einen einheitlichen Satz von Befehlen und liefert in Abhängigkeit von den Parametern der Anfrage verschiedene Repräsentationen (Formate) der Information aus.
- Verwendung von Schichten: REST-Dienste basieren auf mehreren übereinanderliegenden Schichten, welche sauber voneinander getrennt sind. Jede Schicht kommuniziert ausschließlich mit ihren unmittelbaren Nachbarn, nur die höchste Schicht kommuniziert mit dem Client. Tiefere Schichten können damit einfach ausgetauscht und verändert werden, im Gegenzug entsteht höherer Ressourcenbedarf durch die erforderliche Verwaltung der Schichten und ihrer Kommunikation.
- Code on demand (optional): Der Client kann Code (beispielsweise JavaScript) vom Server herunterladen und selbst ausführen, um Latenzprobleme zu vermeiden.

JavaScript Object Notation (JSON) ist ein textbasierendes Datenformat zum plattformübergreifenden Austausch strukturierter Daten, insbesondere in der Maschine-zu-Maschine-Kommunikation. Das Format ist jedoch auch für Menschen lesbar. Es weist damit starke Ähnlichkeiten zu XML auf, ist jedoch schlanker und eignet sich insbesondere für die Kommunikation mit REST-Services oder im mobilen Umfeld (vgl. Nurseitov u.a. 2009).

Für JSON existieren zwei relevante Standards. Der Standard 404 der Ecma International (einer ursprünglich europäischen, jetzt internationalen Organisation zur Normung von Informations– und Kommunikationssystemen) spezifiziert insbesondere die Syntax des Formates (vgl. ECMA International 2013) und wird durch das Request for Comments (RFC) 7159 der Internet Engineering Task Force (IETF) ergänzt (vgl. Internet Engineering Task Force IETF 2014), welches zusätzliche Facetten wie zum Beispiel Kompatibilitätsprobleme zwischen verschiedenen Implementierungen beleuchtet (vgl. Brutzman 2015).

JSON sieht zwei Datenstrukturen vor. Objekte sind ungeordnete Mengen von beliebig vielen Name-/Werte-Paaren (name/value-pairs), während Arrays geordnete Listen von beliebig vielen Werten darstellen. Als Werte erlaubt JSON Zeichenketten (strings), Zahlen, Objekte, Arrays sowie NULL (leere Menge) und die booleschen TRUE und FALSE (wahr beziehungsweise falsch).

Für jede gängige Programmiersprache ist mindestens eine Bibliothek zur Interaktion mit JSON-Dateien verfügbar, um beispielsweise die Eigenschaften

von Objekten nach JSON und wieder zurück zu konvertieren.

5.4 APIs und Applikationen der Watson Developer Cloud

Dieser Abschnitt stellt alle Module der Watson Developer Cloud vor, basierend auf den online verfügbaren Dokumentationen und Redbooks. Die Internas der Systeme werden vom Hersteller nicht veröffentlicht, insofern handelt es sich durchweg um „Black Boxes".

Per Voreinstellung protokollieren alle Watson-Module eingehende Anfragen und ihre Antworten, um diese zur Verbesserung des Dienstes, insbesondere zum Training, heranzuziehen. Um dies für sensible oder personenbezogene Daten zu verhindern, muss im Kopf jeder einzelnen Anfrage ein Parameter gesetzt werden (opt-out, vgl. IBM Deutschland GmbH 2017d).

5.4.1 Watson Conversation

Watson Conversation (Gespräch, Unterhaltung) erlaubt die Erstellung von Chatbots und virtuellen Agenten, welche eine simulierte Unterhaltung mit einem Anwender zu definierten Themen führen. Unterhaltungen in Watson Conversation bestehen aus vier verschiedenen, in Workspaces zusammengefassten Bausteinen (vgl. Azraq u. a. 2017, S. 4–10; und IBM Deutschland GmbH 2016a). Zur Erläuterung dient folgendes Beispiel: der Chatbot soll in der Lage sein, Fragen des Benutzers zu Versandkosten zu beantworten. Diese hängen von dem Zielland ab, möglich sind Deutschland, Österreich und die Schweiz.

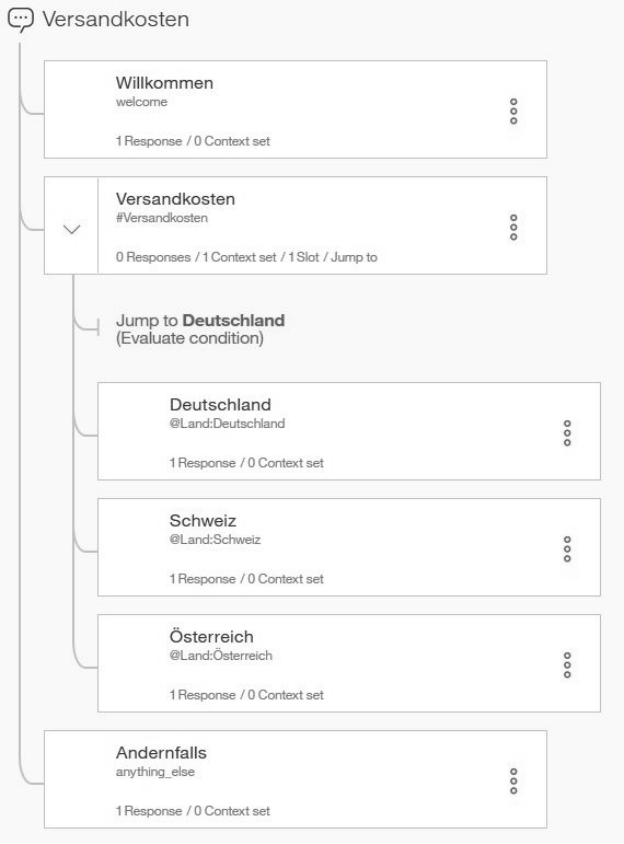
- Ein *#Intent* (Absicht, Bedeutung) beschreibt, was der Benutzer mit einer Eingabe erreichen möchte und damit oft einen Anwendungsfall (Use Case, beispielsweise Reklamation einer Lieferung). Das #Intent besteht aus einem frei gewählten Namen (beginnend mit „#") und mindestens fünf verschiedenen, typischen Fragen oder Sätzen. Für unseren Anwendungsfall wird das #Intent „#Versandkosten" mit den typischen Fragen „Wie viel kostet der Versand?", „Wie hoch sind die Versandkosten?", „Wie teuer ist das Porto?", „Was muss ich für das Paket bezahlen?" und „Was berechnen Sie mir für die Zustellung?" erstellt. Der Dienst trainiert die angegebenen Fragen, so dass er auch neue, noch unbekannte Variationen erkennen kann.
- Ein *@Entity* (Entität) spezifiziert ein #Intent näher; es beschreibt beispielsweise, wie die Benutzerin das Ziel erreichen möchte oder worauf es sich bezieht. Es besteht aus einem frei gewählten Namen (beginnend mit „@") und den möglichen Ausprägungen und Synonymen; dabei sind auch reguläre Ausdrücke erlaubt. In unserem Beispiel sind die Versandkosten abhängig vom Zielland, abgebildet über das @Entity @Land mit den Ausprägungen „Deutschland", „Österreich" und „Schweiz" mit dem jeweiligen Synonymen „BRD", „AUT" und „CH".

Watson Conversation stellt darüber hinaus Systementitäten bereit, um Geldbeträge, Zahlen, Prozentwerte, Zeit– und Datumsangaben zu erkennen und zu extrahieren.

- *$Contexts* (Zusammenhänge) speichern beliebige Informationen, analog zu Variablen in Programmiersprachen. Ihr Name kann frei gewählt werden, beginnt aber immer mit „$“. In unserem Beispiel könnte man aus der Kundendatenbank das Heimatland des Kunden auslesen und diese Information in die Analyse einfließen lassen.
- Der *Dialog* besteht aus verschiedenen Knoten (nodes), welche in einer hierarchischen Baumstruktur angeordnet sind und von oben nach unten und links nach rechts abgearbeitet werden (sh. Abbildung 9 auf der nächsten Seite).
  Knoten verfügen über beliebig viele, logisch und-/oder-verknüpfte Eingangsbedingungen aus #Intents, @Entities und $Contexts. Mögliche Aktionen bei Aktivierung sind unter anderem Ausgabe einer Antwort, Durchführung weiterer Prüfungen, Sprung zu anderen Knoten im Baum oder Rückgabe des Gespräches an die menschliche Benutzerin. Darüber hinaus existieren spezielle Knoten für den erstmaligen Aufruf des Dialoges ("welcome“) und wenn kein anderer Knoten aktiviert („anything_else“). Im einfachsten Fall sind alle Eingangskriterien in einer Dialogzeile vorhanden, beispielsweise enthält die Frage „Wie hoch sind die Versandkosten nach Deutschland?“ sowohl #Intent als auch @Entity. Zum gleichen Ergebnis führt folgender Dialogverlauf: „Wie hoch sind die Versandkosten?“ – „Das hängt davon ab, in welches Land die Sendung geht.“ – „Nach Deutschland“. Da aus dem Verlauf der Konversation der #Intent bereits bekannt ist, muss der Kunde ihn bei der zweiten Eingabe nicht nochmals vorgeben.

Alle Bausteine können über eine im Browser bereitgestellte SaaS konfiguriert werden. Es stehen Import– und Exportfunktionen für JSON- und CSV-Dateien (Comma-separated values) bereit. Der Service bietet ausschließlich Backend-Funktionalität, die steuernde Applikation ist vom Unternehmen selbst zu entwicklen (vgl. IBM Deutschland GmbH 2016a). Der Ablauf gestaltet sich wie folgt:

Abbildung 9: Beispiel für einen Dialog in Watson Conversation



Quelle**:** selbst erstellter Screenshot

1. Die steuernde Applikation beschafft sich die aktuelle Eingabe der Anwenderin.
2. Die steuernde Applikation ruft Watson Conversation über die REST-Schnittstelle auf. Neben der aktuellen Dialogzeile übergibt sie auch den gespeicherten Context, falls vorhanden (siehe Schritt 5).
3. Watson Conversation extrahiert aus der Eingabe @Entities und #Intents.
4. Watson Conversation prüft für jeden Knoten im Dialog, ob die Eingangskriterien erfüllt sind. Wenn Ja, wird der Knoten aktiv.
5. Aktive Knoten werden vom Service abgearbeitet, bis ein Knoten die Kontrolle über das Gespräch an den Anwender zurückgibt. In diesem Fall gibt Watson Conversation die Ausgaben und den „Context“ als Rückgabewert an das Steuerprogramm zurück. Er enthält neben den $Context-Variablen, @Entities und #Intents auch Metadaten wie ID und Zustand des aktuellen Knotens und die Anzahl der Seitenwechsel im laufenden Gespräch. Die steuernde Applikation muss ihn zwischenspeichern und beim nächsten Aufruf des Dienstes übergeben.
6. Die steuernde Applikation sendet die Antwort an den Benutzer. Der Ablauf beginnt wieder von vorne.

Vorteil dieser Architektur ist ihre Flexibilität. Es stehen Software Development Kits (SDKs) für zahlreiche Programmiersprachen wie Node, Java, Python und .NET bereit, und selbst ohne SDK sind REST-APIs unter jeder Sprache ansprechbar. Die Applikation kann beliebige Drittsysteme und Eingangskanäle (beispielsweise Chatfenster auf der Website, Messagingdienste, mobile Applikationen) einbinden. Nachteilig sind die erforderlichen Aufwände für die Eigenentwicklung.

5.4.2 Watson Virtual Agent

Watson Virtual Agent ist ein auf Watson Conversation beruhender, weitgehend vorkonfigurierter Chatbot beziehungsweise virtueller Agent. Er bietet Funktionspacks (capabilities), welche für verschiedene Branchen und in verschiedenen Sprachen (Allgemein, Energie, Banking, Telekommunikation; unter anderem deutsch und englisch) jeweils 50-100 typische Funktionen enthalten, beispielsweise Fragen zur nächstgelegenen Filiale oder Vertragsanliegen wie Adressänderungen (vgl. IBM Deutschland GmbH 2017b). Die Reaktion der einzelnen Funktionen ist konfigurierbar: so können sie einfache Antworten ausgeben, über integrierte, vorgefertigte Dialoge Daten vom Kunden einholen und gesammelt an firmeneigene Backend-Systeme übergeben, Dialoge in Watson Conversation aufrufen, den Dialog an menschliche Kundenberater eskalieren oder deaktiviert werden. Darüber hinaus gehört ein Chatbot-Widget zum Lieferumfang, das System kann aber auch über eine selbst erstellte Chat-Schnittstelle in JavaScript (SDK vorhanden) oder REST-API-Aufrufe angesprochen werden.

Im Vergleich zu Watson Conversation können sich Unternehmen Entwicklungsaufwände für den Dialog, die Steuerkomponente und evtl. den Chatbot sparen; einzurechnen sind hingegen Aufwände für Schnittstellen in die firmeneigenen Backend-Systeme für Eskalation und vorgefertigte Dialoge. Virtual Agent wird ausschließlich mit zwölfmonatiger Mindestlaufzeit und monatlicher Grundgebühr[4] angeboten, während Conversation rein nach Nutzung (Anzahl der API–Aufrufe) abgerechnet wird.

5.4.3 Watson Natural Language Classifier

Watson Natural Language Classifier ist ein Service zur Klassifizierung von kurzen Texten bis maximal 1.024 Zeichen; sie sollten aus weniger als 60 Wörtern bestehen (vgl. Manhaes u. a. 2017, S. 61; und IBM Deutschland GmbH 2015a).

Vor der produktiven Nutzung des Dienstes muss zunächst ein Klassifizierer mit einer Trainingsmenge angelernt werden, diese besteht aus einer Menge von Texten und dem jeweils zugeordneten, frei wählbaren Klassennamen. Jeder

[4] Sh. https://www.ibm.com/de-de/marketplace/cognitive-customer-engagement/purchase

Klasse müssen mindestens fünf, maximal 15.000 Texte zugeordnet sein. Das Trainieren ist über eine SaaS im Browser oder über die REST-API möglich.

Der Aufruf eines Klassifizierers erfolgt ebenfalls über die REST-API, dabei muss das Steuerprogramm zunächst die Verfügbarkeit des Klassifizierers prüfen und ihm anschließend den Text übergeben. Die Antwort im JSON-Format enthält unter anderem die erkannten Kategorien mit ihrem Konfidenzgrad.

5.4.4 Watson Natural Language Understanding

Watson Natural Language Understanding ist eine Sammlung von Werkzeugen zur oberflächlichen Analyse und Extraktion von Metadaten aus unstrukturierten Fließtexten. Dabei kann es sich um reine Textinformationen (plain text), Texte im Format Hypertext Markup Language (HTML) oder die URL einer Webseite handeln (vgl. IBM Deutschland GmbH 2017a; und Vergara u. a. 2017, S. 1–2).

Für die einzelnen Werkzeuge stellt IBM fertige, nicht konfigurierbare Modelle für unterschiedliche Sprachen bereit. Alle folgenden Funktionen sind für englischsprachige Texte verfügbar; für deutsche nur, wo dies ausdrücklich vermerkt ist[5] (vgl. Vergara u. a. 2017, S. 5–54).

- *Concepts* erkennt generelle Themen und abstrakte Konzepte des Textes und liefert Links in die DBPedia, einem semantischen Netz welches strukturierte Informationen aus Wikipedia extrahiert (vgl. Auer u. a. 2007). Die Funktionalität bietet sich beispielsweise zum Clustern von Artikeln an.
- *Emotions* bewertet im die im Text ausgedrückten Emotionen Ärger, Empörung/Abscheu, Angst, Freude und Traurigkeit auf einer Skala von 0 (nicht vorhanden) bis 1, optional auch auf ein bestimmtes Ziel bezogen (beispielsweise ein Unternehmen oder eine Person). Potenzielle Anwendungsgebiete sind die Analyse von Social Media-Einträgen oder Kundenbewertungen; aber auch die Erkennung von Beschwerden.
- *Keywords* extrahiert die wesentlichen Schlüsselbegriffe im Text und deren Relevanz. Die Informationen können beispielsweise für Suchen oder Indizes benutzt werden. Auch für deutsche Texte.
- *Entities* erkennt Entitäten im Text (beispielsweise Personen, Firmen, Organisationen, Länder), zählt wie oft diese auftreten und bewertet ihre Relevanz. Ein Modell für deutsche Texte steht bereit, das Werkzeug unterstützt zudem in Watson Knowledge Studio selbst erstellte Modelle.
- *Relations* erkennt Relationen (Beziehungen) zwischen zwei Entitäten. Für deutsche Texte gibt es kein fertiges Modell, es ist jedoch möglich,

[5] Vollständige, stets aktuelle Übersicht sh. https://console.bluemix.net/docs/services/natural-language-understanding/language-support.html

ein in Watson Knowledge Studio selbst erstelltes Modell einzubinden.
- *Metadata* extrahiert Autor, Titel und Veröffentlichungsdatum aus dem Kopf (Header) von HTML-Dateien. Da der eigentliche Inhalt (Body) nicht analysiert wird, funktioniert dies für alle Sprachen.
- *Semantic Roles* erkennt Subjekt, Prädikat und Objekt eines Satzes (Part-of-Speech-Tagging).
- *Sentiment* bewertet die Stimmung eines Textes (im Sinne der Einstellung des Verfassers) und bewertet sie auf einer Skala von -1 (negativ) bis +1 (positiv), optional auch auf bestimmte Entitäten (beispielsweise ein Unternehmen oder eine Person) bezogen. Es steht zwar ein Modell für deutsche Texte bereit, dieses wird jedoch ausdrücklich nur für soziale Medien aufgrund deren Tendenz zu überschwänglicher Sprache empfohlen.
- *Categories* kategorisiert den Text anhand einer von IBM festgelegten, fünfstufigen Taxonomie[6].

Diese Liste macht deutlich, dass die Unterstützung für deutsche Texte schon rein quantitativ hinter der englischen Sprachunterstützung zurückbleibt. Experimente des Autors haben zudem gezeigt, dass die Unterstützung für deutsche Texte auch qualitativ nicht so gut wie die für englische Texte ist.

5.4.5 Watson Discovery

Watson Discovery liest unstrukturierte Texte ein, ergänzt diese um enrichments (Anreicherungen) und ermöglicht Suchanfragen (information retrieval) auf beides. Bei den Enrichments handelt es sich um die gleichen Module wie bei Watson Natural Language Understanding (sh. Abschnitt 5.4.4), mit den gleichen Einschränkungen in der Sprachunterstützung: Keywords, Sentiments, Concepts, Metadata, Categories, Semantic Roles, Emotions, Entities und Relationen (vgl. IBM Deutschland GmbH 2017c).

Als erster Schritt ist eine Data Collection einzurichten und zu konfigurieren. Sie sammelt alle aufgenommenen Dokumente als JSON-Dateien, welche wiederum hauptsächlich aus einer ID, HTML- und Text-Versionen des Dokumenteninhaltes und den ermittelten Enrichments bestehen. Importierbar sind HTML-, JSON-, Portable Document Format (PDF)- und Microsoft Word-Dateien. PDF- und Word-Dokumente werden zunächst intern in HTML umgewandelt, wobei ein konfigurierbares Regelwerk festlegt, welche Schriftgrößen beziehungsweise -formatierungen als Überschrift (HTML-Tag H1, H2 usw.) zu behandeln sind. HTML-Dateien werden intern nach JSON umgewandelt, wobei ein ebenfalls konfigurierbares Regelwerk über ein- und auszuschließende Tags entscheidet.

[6] Jeweils aktuelle Version verfügbar unter https://console.bluemix.net/docs/services/natural-language-understanding/categories.html

Die Data Collection legt auch fest, welche Enrichments anzuwenden und ob beim Dokumenten-Import weitere Bearbeitungen der JSON-Felder (löschen, zusammenfassen usw.) vorzunehmen sind.

Im zweiten Schritt werden Dokumente zu einem beliebigen Zeitpunkt importiert, beispielsweise manuell über die Weboberfläche der SaaS, über REST-APIs oder über einen bereitgestellten Crawler für Linux. Da die Einstellungen für alle Dokumente der Sammlung in Schritt eins festgelegt wurden, sind neben dem Dokument keine weiteren Parameter erforderlich.

Im dritten Schritt stehen die eingelesenen und angereicherten Dokumente über die REST-API für Abfragen bereit. Bei der „Natural Language Query" werden Fragen im Klartext gestellt, während die „Query Language" Filterung und Aggregation anhand der angereicherten, extrahierten Daten erlaubt. Die folgende Abfrage sucht beispielsweise alle Dokumente mit mindestens leicht positiver Sentiment-Einstufung (> 0,3), in denen die Entität „Media Markt" als Firma vorkommt[7]:

```
enriched_text:(sentiment.document.score>=0.3,entities:(text::Media
                Markt,type::Company) )
```

5.4.6 Watson Knowledge Studio

Watson Knowledge Studio erlaubt das Erstellen eigener, domänenspezifischer Modelle für Entitäten und Relationen, um diese in Watson Discovery und Watson Natural Language Understanding einzusetzen. Sie können auf Regeln oder Maschinenlernen basieren (vgl. IBM Deutschland GmbH 2016b).

Um ein Maschinenlernen-Modell zu trainieren, ist zunächst ein jederzeit modifizierbares Typensystems (type system) festzulegen, bestehend aus den für die Domäne relevanten Entitäten und den verbindenden Relationen. Im zweiten Schritt werden Dokumente als Text- oder CSV-Dateien importiert, in Arbeitspakete (annotation sets) aufgeteilt und an die Trainerinnen verteilt, welche über eine SaaS im Browser Entitäten im Text markieren und Relationen zuordnen. Danach wird das Modell auf Basis der Annotationen trainiert.

Auch regelbasierte Modelle basieren auf einem Typensystem aus Entitäten und Relationen, sehen aber zusätzlich noch Klassen (classes) für Regelergebnisse vor. Diese werden anschließend entweder auf andere Klassen oder auf Entitäten abgebildet. Ist beispielsweise eine Postanschrift regelbasiert als Entität zu erkennen, bietet sich die Einrichtung der drei Klassen Straße, Postleitzahl und Ort an; die Entität ergibt sich als Kombination der drei Klassen.

Regeln werden direkt aus Dokumenten im Arbeitspaket abgeleitet. Klickt der Anwender beispielsweise auf eine Postleitzahl, werden deren Eigenschaften (exakt fünf Stellen, ausschließlich Ziffern usw.) im Dialog vorbelegt, sind aber

[7] Eine vollständige Übersicht der Sprache befindet sich unter https://console.bluemix.net/docs/services/discovery/query-reference.html.

modifizierbar. Das System unterstützt darüber hinaus auch reguläre Ausdrücke und Wortlisten.

5.4.7 Watson Tone Analyzer

Watson Tone Analyzer bewertet den Tonfall eines Textes in zwei verschiedenen Modi (vgl. IBM Deutschland GmbH 2016c).

Der allgemeine Modus (general purpose) bewertet die im Text ausgedrückten Emotionen Wut/Ärger (anger), Angst (fear), Freude (joy), und Traurigkeit (sadness) sowie den Tonfall analytisch (analytical), zuversichtlich/souverän (confident) und zaghaft (tentative) auf einer Skala von 0 bis 1. Die Bewertung bezieht sich dabei ausschließlich auf den Text (wahlweise im Ganzen oder Satz für Satz), nicht auf den Autoren selbst. Verarbeitet werden Dokumente in englischer und französischer Sprache im HTML-, JSON- oder Textformat.

Der Modus Kundenansprache (customer engagement) ist hingegen auf die Bewertung der Kommunikation zwischen Unternehmen und Kunden ausgelegt (beispielsweise Chat, Social Media, transkribierte Telefongespräche) und bewertet jede einzelne Dialogzeile hinsichtlich Aufregung (excited), Frust (frustrated), Unhöflichkeit (impolite), Höflichkeit (polite), Traurigkeit (sad), Zufriedenheit (satisfied) and Sympathie (sympathetic). Der Dienst akzeptiert ausschließlich englischsprachige Texte im JSON-Format.

5.4.8 Watson Personality Insights

Im Gegensatz zu Watson Tone Analyzer, welches den Text an sich bewertet, erstellt Watson Personality Insights Persönlichkeitsprofile eines Autoren auf Grundlage seiner Texte (vgl. IBM Deutschland GmbH 2015b). Es basiert auf der Annahme, dass die alltägliche Wortwahl einer Person deren Gedanken reflektiert und dadurch Rückschlüsse auf ihre Persönlichkeit zulässt und eignet sich für Texte, bei denen ein Autor seine Worte frei wählen kann (beispielsweise Korrespondenz, Blog-, Social Media- und Forenbeiträge). Nicht geeignet sind hingegen Texte mit eingeschränkten oder festgelegten Ausdrucksmöglichkeiten, beispielsweise wissenschaftliche Arbeiten oder literarische Werke (Geschichte und Wortwahl fiktionaler Charaktere lassen keine Rückschlüsse auf den Autor zu).

Die API unterstützt die Analyse von HTML-, Text- und JSON-Dateien in den Sprachen Englisch, Spanisch, Japanisch, Koreanisch und Arabisch. Texte müssen mindestens 600, besser 1.200 Zeichen lang sein.

Der Dienst nimmt folgende Einschätzungen vor:

- Consumption Preferences (Konsumvorlieben) liefert 42 verschiedene Bewertungen in acht Kategorien, unter anderem bezüglich Musik–, Bücher– und Kleidungsgeschmack und Beeinflussbarkeit durch sozi-

ale Medien[8], auf einer Skala von 0 (kein Interesse) bis 1 (hohes Interesse).

- Needs (Bedürfnisse) bewertet das Bedürfnis der Person nach Aufregung (excitement), Harmonie (harmony), Neugierde (curiosity), Ideale (ideal), Nähe (closeness), Selbstdarstellung (self-expression), Freiheit (liberty), Liebe (love), Zweckmäßigkeit (practicality), Stabilität (stability), Herausforderung (challenge) und Struktur (structure). Dieses und alle folgenden Module bewerten in Form eines Perzentils. Ein Wert von 0,78 bedeutet beispielsweise, dass das Bedürfnis der Person stärker ausgeprägt ist als bei 77% der Stichprobe.
- Values (Werte) gewichtet die motivierenden Faktoren einer Person: Selbsttranszendenz (self-transcendence), Bewahrung (conservation/ tradition), Hedonismus (hedonism), Selbstaufwertung (self-enhancement) und Offenheit gegenüber Veränderungen (openness to change).
- Das Fünf-Faktoren-Modell der Persönlichkeitsanalyse (FFM, im englischen Sprachraum „Big Five“) ist ein Referenzmodell der Persönlichkeitspsychologie zur einheitlichen Beschreibung und Messung der Persönlichkeit (vgl. Herzberg und Roth 2014, S. 40–45). Die Faktoren sind Neurotizismus (neuroticism), Extraversion, Verträglichkeit (agreeableness), Gewissenhaftigkeit (conscien-tiousness) und Offenheit für Erfahrungen (openess).

Der Kerneinsatzbereich des Moduls liegt im Marketing, insbesondere zur Analyse von Social-Media-Aktivitäten und als Grundlage für ziel(gruppen)gerichtete Werbung und Kundenakquise.

5.4.9 Watson Visual Recognition

Watson Visual Recognition umfasst zwei Funktionen: die Klassifikation von Bildern auf Grundlage von selbst definierten Klassen und Trainingsmengen, und die Bildbeschreibung auf Basis unveränderlicher, von IBM bereitgestellter Modelle (vgl. IBM Deutschland GmbH 2015e).

Das Erstellen, Trainieren und Aktualisieren eigener Klassifizierer erfolgt über die API oder eine Weboberfläche. Dazu werden die (positiven) Klassen frei festgelegt und für jede Klasse passende Beispielbilder als Trainingsmenge hochgeladen. Eine Klasse muss mindestens zehn Beispielbilder beinhalten, empfohlen sind jedoch mehrere Hundert oder gar tausende Bilder pro Klasse. Neben den positiven Klassen ist auch eine optionale Klasse für Negativbeispiele (also Bilder, welcher keiner der positiven Klassen zuzuordnen sind) vorgesehen (vgl. Elhassouny u. a. 2017, S. 29–34). Auf das Training selbst hat der Anwen-

[8] Vollständige aktualisierte Auflistung unter https://console.bluemix.net/docs/services/personality-insights/preferences.html

der keinen direkten Einfluss, das Ergebnis kann nur indirekt durch Überarbeitung der Trainingsmengen beeinflusst werden.

Die Bildbeschreibung listet erkannte Gegenstände, Tiere, Farben, Personen und anderes in Bildern oder in Videos (in diesem Falle analysiert die API jedes einzelne Frame) auf. Die Erkenner werden fortlaufend überarbeitet, eine Auflistung der möglichen Kategorien wird nicht veröffentlicht. Erkennt das System Gesichter, gibt es deren Position im Bild und das vermutete Geschlecht und Alter der Person aus; bei erkannten Persönlichkeiten (beispielsweise Politikern) darüber hinaus auch deren Namen (vgl. ebenda, S. 35–40).

5.4.10 Unterstützende APIs

Diese APIs dienen primär zur Unterstützung der bereits beschriebenen APIs und werden daher nur kurz skizziert.

Watson Language Translation übersetzt Texte. Es werden drei vorkonfigurierte Modelle bereitgestellt, welche durch Listen mit gewünschten/präferierten Übersetzungen (zur Einhaltung von Unternehmensvorgaben) individualisierbar sind (vgl. Lampkin u. a. 2017, S. 1–7; und IBM Deutschland GmbH 2014).

- News ist das Standard-Modell mit Schwerpunkt Nachrichten und Artikel. Es übersetzt englische Texte nach Arabisch, brasilianischem Portugiesisch, Französisch, Deutsch, Italienisch, Japanisch, Koreanisch, Spanisch und jeweils umgekehrt.
- Das Modell Conversational (Umgangssprache) übersetzt von Englisch nach Arabisch, brasilianischem Portugiesisch, Französisch, Italienisch, Spanisch und umgekehrt.
- Patents ist ein auf technisches und rechtliches Vokabular spezialisiertes Modell und übersetzt unidirektional von brasilianischem Portugiesisch, Chinesisch, Koreanisch und Spanisch nach Englisch.

Watson Text to Speech setzt Text in gesprochene Sprache um (Sprachsynthese). Insgesamt stehen 13 Stimmen in acht Sprachen bereit, darunter eine männliche und eine weibliche Stimme für deutsche Texte (vgl. Santiago, Singh und Sri 2017, S. 7–8; und IBM Deutschland GmbH 2015d).

Watson Speech to Text geht den umgekehrtem Weg, erzeugt also aus gesprochener Sprache Text. Unterstützte Sprachen sind Englisch, Spanisch, Arabisch, Chinesisch, Japanisch, Französisch und brasilianisches Portugiesisch (vgl. Santiago, Singh und Sri 2017, S. 1–4; und IBM Deutschland GmbH 2015c).

Tabelle 1 zeigt alle Module der Watson Developer Cloud im Überblick, inklusive deren Sprachunterstützung für deutsche Texte.

Tabelle 1: IBM Watson Developer Cloud: API-Übersicht

Name API bzw. Applikation	**Kurzbeschreibung**	**Sprachunterstützung Deutsch**
Watson Conversation	Erstellen von Chatbots und Assistenten.	Ja.
Watson Discovery	Einlesen unstrukturierter Texte, Anreicherung mit Metadaten, Suche über Volltext und Metadaten.	Stark eingeschränkt: Entitäten (vorgefertigte und selbsterstellte Modelle), HTML-Metadaten, Schlüsselwörter, Relationen (selbst erstellte Modelle), Meinungen/Gefühle (nur für Social Media geeignet).
Watson Knowledge Studio	Erstellen von Modellen für Entitäten und Relationen.	Ja.
Watson Language Translator	Übersetzung und Spracherkennung.	Deutsch – Englisch und umgekehrt.
Watson Natural Language Classifier	Klassifikation kurzer Texte (bis 1.024 Buchstaben) in frei wählbare Klassen.	Ja.
Watson Natural Language Understanding	Oberflächliche Analyse und Extraktion von Metadaten aus unstrukturierten Fließtexten.	Stark eingeschränkt, analog Watson Discovery.
Watson Personality Insight	Persönlichkeitsanalyse anhand von Texten: Fünf-Faktoren-Modell, Werte, Bedürfnisse, Konsumverhalten.	Nein.
Watson Text to Speech	Sprachsynthese.	Zwei deutsche Stimmen.
Watson Tone Analyzer	Analyse eines Textes (Emotionen, Sprachstil) oder eines Kundengespräches.	Nein.
Watson Speech to Text	Spracherkennung.	Nein.
Watson Virtual Agent	Vorgefertigter Chatbot beziehungsweise Assistent, mit Vertragslaufzeit und monatlicher Grundgebühr.	Ja.
Watson Visual Recognition	Klassifizierung von Bildern in selbst gewählte Klassen. Gesichtserkennung. Bildbeschreibung (vorgefertigte Klassen).	Nicht relevant.

Quelle: eigene Darstellung

6 Betriebswirtschaftliche Anwendung von Cognitive Computing

Zu Beginn dieses Kapitels wird CRISP-DM als Vorgehensmodell zum Trainieren von kognitiven Systemen vorgestellt, bevor vier Szenarien zur Prozessverbesserung von Routine-Prozessen durch das kognitive System IBM Watson Developer Cloud aufgezeigt und jeweils der Funktionsumfang, die Anforderungen und die Voraussetzungen (sofern nicht bereits allgemeingültig in Kapitel 5 aufgeführt) behandelt werden. Abschließend untersucht die Arbeit den Einsatz des kognitiven Systems für Know-how- und Entscheidungs-intensive Prozesse (Typ I / II).

6.1 Vorgehensmodell zum Trainieren kognitiver Anwendungen

Die Zusammensetzung der Trainingsmenge und die Qualität des Trainings sind entscheidende Einflussgrößen für den Erfolg eines anzulernenden KNN oder kognitiven Systems. Daher bietet sich ein Vorgehensmodell an, um den Prozess zu systematisieren. Wie bereits in Abschnitt 2.4.3 dargestellt, weisen Data Mining und Maschinenlernen Parallelen auf: Beides sind Verfahren zur Datenanalyse, insbesondere der Mustererkennung, und Cognitive Computing kann als eine Automatisierung des Data Minings aufgefasst werden. Daher bietet es sich an, auf ein bewährtes Modell des Data-Mining-Prozesses zurückzugreifen: dem CRISP-DM (Cross Industry Standard Process for Data Mining, vgl. Cleve und Lämmel 2016, S. 6–9). Das Modell kann ohne strukturelle Änderungen, lediglich in Details modifiziert (basierend auf Hurwitz, Kaufman und Bowles 2015, S. 157–173), für diese Aufgabenstellung verwendet werden (s. Abbildung 10).

1. Verstehen der Aufgabe
   Ziel ist ein grundlegendes Verständnis des Fachgebietes und der Aufgabe selbst, insbesondere der mit dem kognitiven System zu erreichenden Ziele, der bestehenden Risiken und der Anwender des fertigen Systems (beispielsweise externe Kunden, Sachbearbeiter oder Management).
2. Verständnis der Daten
   In dieser Phase steht die Definition, das Verständnis und die Beschaffung der benötigten Trainingsdaten im Vordergrund. Hier ist zunächst zu klären, welche Trainingsdaten benötigt werden, woher diese stammen und welche Expertinnen bei Auswahl geeigneter Daten unterstützen und das Training durchführen. Die Trainingsdaten sind insbesondere von der Wissensdomäne und dem Benutzer abhängig: von dessen Kenntnisstand und der Art und Formulierung seiner Fragen beziehungsweise Anliegen. Ein fachlicher Laie (wie beispielsweise ein Privatkunde) erwartet verständliche Antworten und wird sein Anliegen

anders formulieren als ein Experte, der an möglichst pointierten Detailinformationen interessiert ist. Sind die benötigten Daten identifiziert, ist zu prüfen, in welchem Format diese vorliegen und welche Konsequenzen sich daraus ergeben. E-Mails sind beispielsweise semistrukturiert, verfügen über irrelevante Daten (Werbung, Fußzeilen) und können in verschiedensten Codierungen vorliegen; Schreiben in Papierform sind hingegen unstrukturiert und müssen eingescannt werden, so dass mit Fehlern in der Zeichenerkennung (OCR, Optical Character Recognition) zu rechnen ist.

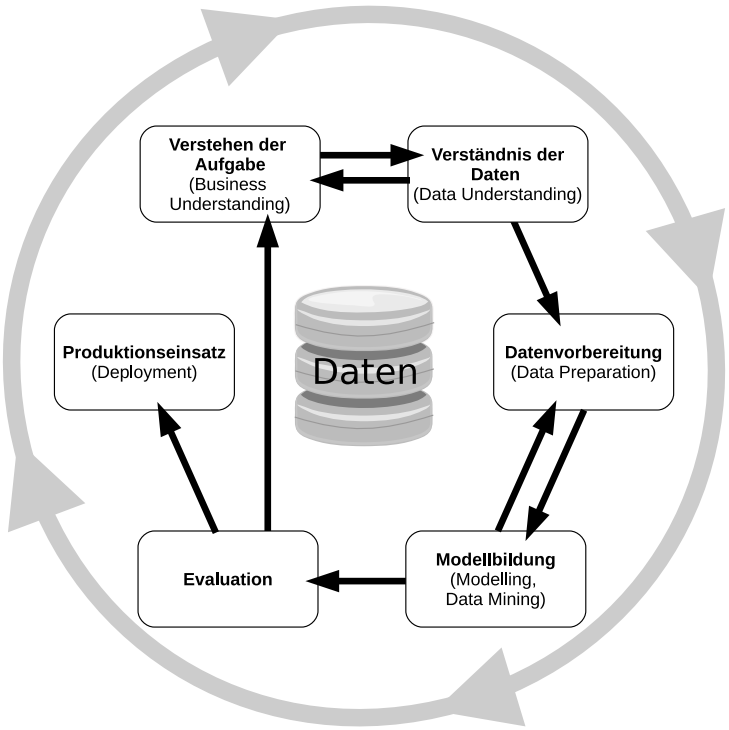
3. Datenvorbereitung
   Dieser Schritt dient der Vorbereitung der Modellbildung. Die Trainingsmengen werden beschafft, extrahiert und in der vom kognitiven System benötigten Form aufbereitet (Datenformate, Codierung usw.). Anschließend ist in der Regel eine Form des manuellen Trainings durch Menschen erforderlich, beispielsweise die Einteilung von Dokumenten in Klassen oder das Annotieren der zu extrahierenden Bestandteile (tagging) in unstrukturierten Dokumenten. Für qualitativ hochwertige Ergebnisse ist dazu von vornherein ausreichend Zeit und Kapazität einzuplanen, ausreichend für mehrere Iterationen mit Anpassungen der Trainingsmengen und Klassen. Ein weiterer wichtiger Aspekt ist die Homogenität des Trainings: gleichartige Sachverhalte müssen von verschiedenen Trainern auf die selbe Art interpretiert und trainiert werden (vgl. IBM Deutschland GmbH 2016b). Um dies zu erreichen, wird in den ersten Iterationen nur ein kleiner Ausschnitt der Trainingsmenge bearbeitet; zudem nach jedem Durchgang die Ergebnisse verglichen und einheitliches Vorgehen durch Abstimmungen und Richtlinien (guidelines) sichergestellt.
4. Modellbildung
   Die eigentliche Datenanalyse (Erstellung eines Modelles) erfolgt automatisiert durch das kognitive System.
5. Evaluation der Ergebnisse
   Die Ergebnisse werden anhand der in Schritt 1 festgelegten Ziele überprüft. Sind sie nicht zufriedenstellend, kommen folgende, typische Probleme in Betracht:
   - Die Trainingsdaten sind nicht repräsentativ für ihre Klasse, decken also nicht die typischen Anfragen dieser Klasse und/oder des Benutzerkreises ab oder sind veraltet, vgl. Schritt 2.
   - Die Trainingsdaten weisen einen Bias (eine Verzerrung) auf, sind also nicht repräsentativ für den Gesamtbestand. Beispielsweise enthält die Trainingsmenge nur Datensätze mit bestimmten Eigenschaften. Siehe Schritt 2.
   - Es stehen noch nicht genug Trainingsdaten zur Verfügung, oder

es sind weitere Datenquellen zu erschließen (siehe Schritt 2).

- Fehler beim manuellen Training, beispielsweise sind Datensätze den falschen Klassen zugeordnet oder gleiche Sachverhalte unterschiedlichen Klassen. Dieses Problem wird insbesondere durch den Einsatz mehrerer Anwenderinnen zum Training aufgrund deren Idiosynkrasien verschärft, vgl. Schritt 3.
- Klassen ähneln sich zu sehr (fehlende Trennschärfe), so dass keine zuverlässige Klassifikation möglich ist. Lösung: Klassen(struktur) überarbeiten, siehe Schritt 2.
- Es besteht die Gefahr, dass Trainingsdaten vom System anders interpretiert werden als erwartet. Ein Beispiel aus der Literatur: ein KNN soll Bilder mit Zügen erkennen und ist dabei sehr erfolgreich; allerdings ist das wichtigste Kriterium des Netzes nicht das eigentliche Gefährt, sondern die Schienen auf denen es steht – es wurde also unbemerkt ein „Schienenerkenner“ trainiert (vgl. Knight und Wolfangel 2017, S. 33–34). Die Lösung besteht im Überarbeiten der Trainingsmengen (Schritt 2).

6. Produktionseinsatz
   Vorbereitung und Durchführung des produktiven Einsatzes. Dabei müssen auch Möglichkeiten zum Controlling geplant und umgesetzt werden.

Abbildung 10: Das CRISP-Modell für Data Mining (CRISP-DM)



Quelle: eigene Darstellung nach Cleve und Lämmel 2016, S. 7

6.2 Routine-Prozesse (Typ III)

In diesem Abschnitt werden vier Anwendungsmöglichkeiten für Routine-Prozesse dargestellt: die Prozessausgliederung an Kunden, die Verbesserung von Prozessabläufen inklusive Automatisierung und die Unterstützung von BI. Wie im Abschnitt 4.2 dargestellt wird bei den folgenden Ausführungen davon ausgegangen, dass Prozessabläufe und Automatisierungsgrad bereits ohne den Einsatz kognitiver Systeme eine zufriedenstellende Qualität erreicht haben.

Ein detaillierterer Blick wird hingegen auf die Anforderungen (Funktionalität, Sprachunterstützung, Verfügbarkeit, Kritikalität) und Voraussetzungen (technische Rahmenbedingungen, benötigte Daten, Vorarbeiten) für den Einsatz des kognitiven Systems Watson Developer Cloud geworfen, sofern diese nicht bereits in Kapitel 5 behandelt wurden.

6.2.1 Prozessausgliederung an den Kunden

Die Ausgliederung von Prozesselementen an den Kunden setzt voraus, dass dieser im Laufe des Prozesses in irgendeiner Form kontaktiert wird. Sie bietet sich damit insbesondere für Dienstleistungen an, deren wesentliches Merkmal die Integration des Kunden als externer Faktor in den Produktionsprozess ist (vgl. Haller 2017, S. 7–9). Damit der Kunde die Prozessausgliederung akzeptiert, muss sie ihm Vorteile bringen. Neben einem niedrigeren Preis (das Unternehmen spart durch die Ausgliederung Kosten und gibt diese Einsparungen an seine Kunden weiter) können auch nicht-monetäre Faktoren eine Rolle spielen: Bequemlichkeit, Verringerung der Wartezeiten (und damit auch der Durchlaufzeit), Loslösung von den Servicezeiten der Kundenbetreuung, Freude an der eigenen Leistung. Als nachteilig können hingegen der erforderliche Eigenaufwand und die erhöhte Eigenverantwortung (und damit Unsicherheit) angesehen werden (vgl. Posluschny 2016, S. 32–34).

Lässt sich der Kunde auf die Prozessausgliederung ein, bieten ihm Chatbots und Avatare darüber hinaus potenziell weiteren Nutzen (vgl. Bruhn u. a. 2013, S. 447–449): Komplexitätsnutzen (Reduzierung der Komplexität, beispielsweise durch Beratung, Information oder schrittweises “an die Hand nehmen“), Benutzerfreundlichkeitsnutzen (Vereinfachung der Bedienung), Bestätigungsnutzen (Bestätigung des Kunden, die richtige Entscheidung getroffen zu haben), Erlebnisnutzen (die Interaktion mit dem Avatar als Erlebnis) und sozialen Interaktionsnutzen (die Kommunikation mit dem Chatbot kann als soziale Interaktion empfunden, der Avatar als Gastgeber, persönlicher Helfer oder Berater wahrgenommen werden).

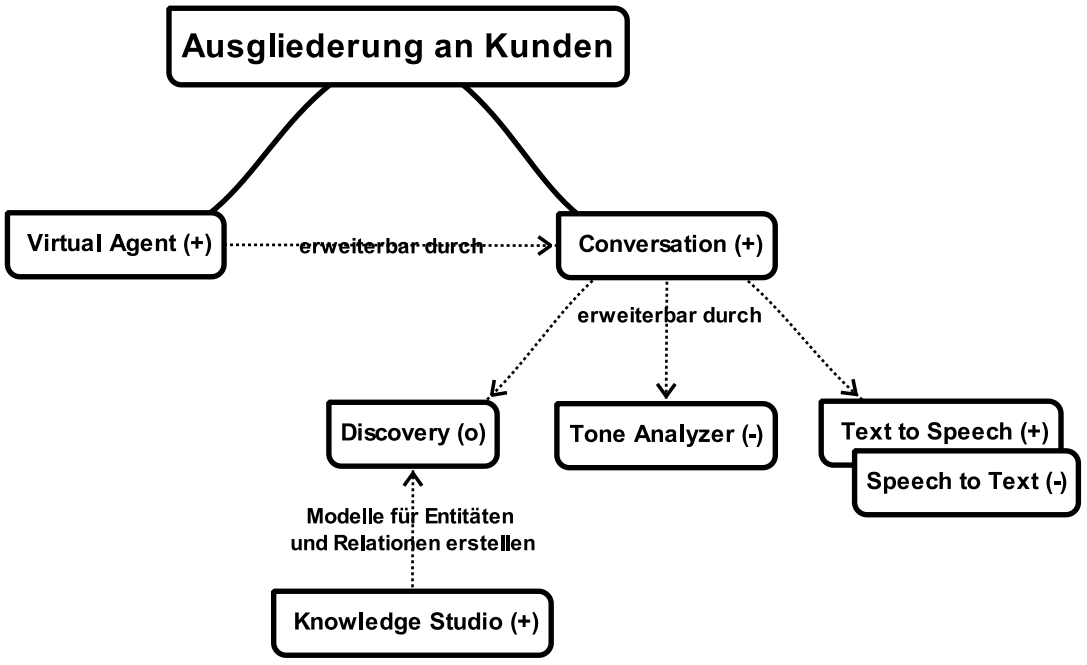
Abbildung 11 zeigt die für eine Prozessauslagerung an Kunden in Frage kommenden Module der Watson Developer Cloud und deren Zusammenspiel. Das

Symbol in Klammern spiegelt die Sprachunterstützung der Module für deutsche Texte wider (+ = uneingeschränkt, o = eingeschränkt, - = nicht vorhanden). Als Einstieg beziehungsweise Basis fungiert entweder Watson Conversation (vgl. Abschnitt 5.4.1) oder das darauf basierende, vorgefertigte Watson Virtual Agent (vgl. Abschnitt 5.4.2). Beide Module führen als Chatbot beziehungsweise Avatar strukturierte Dialoge mit der Kundin, nehmen unstrukturierte Anfragen entgegen und beantworten diese. Als Quelle kommt jeder Kanal in Betracht, über den das unternehmensintern entwickelte Steuerprogramm kommunizieren kann; beispielsweise Chatfenster, Messagingdienste, Online-Kontaktformulare, E-Mails oder Smartphone-Applikationen.

Es sind drei Arten von Antworten zu unterscheiden: allgemeine Informationen, spezifische Informationen und die Steuerung (einfacher) Geschäftsvorfälle.

Allgemeine Informationen sind beispielsweise Frequently Asked Questions (FAQs), Hilfetexte, Erklärungen, Links zu weiterführenden Webseiten oder Frage-Antwort-Paare wie beispielsweise „Liefern Sie nach Spanien?“ – „Ja“. Diese lassen sich in der Regel relativ schnell umsetzen, da sie keinen Zugriff auf Backend-Systeme benötigen und für häufig nachgefragte Sachverhalte in der Regel bereits Textbausteine vorhanden sind. Sie bieten sich daher für erste Schritte / Phasen oder „Quick Wins“ an.

Abbildung 11: Anwendung der Watson Dienste zur Prozessauslagerung an Kunden



Quelle: eigene Darstellung

Spezifische Informationen sind hingegen kontextsensitiv, also abhängig vom Fragesteller; beispielsweise Fragen wie „Wann endet mein Vertrag?“ oder „Ist mein Kundenkonto ausgeglichen?“. Hierzu sind Schnittstellen in die Backend-Systeme des Unternehmens erforderlich, welche vom Steuerprogramm aufgerufen und als $Contexts ausgetauscht werden.

Für Geschäftsvorfälle werden mehrere Angaben von der Kundin benötigt, beispielsweise die neue Anschrift (bestehend aus Straße, Hausnummer, Postleitzahl und/oder Ort) und das Gültigkeitsdatum für eine Adressänderung. Die Art und Anzahl der Daten kann dabei je nach Gesprächsverlauf variieren. Conversation beziehungsweise Virtual Agent extrahiert die benötigten Daten aus den Eingaben, fragt fehlende Angaben gezielt nach und gibt diese gesammelt und strukturiert an das Steuerprogramm zurück, welches sie wiederum an die unternehmenseigenen Backend-Systeme übermittelt. Letztendlich handelt es sich um Geschäftsvorgänge, welche typischerweise über Webformulare (frei zugänglich oder im Rahmen eines Kundenbereiches) angeboten werden. Ein konkretes Beispiel für die Anwendung von Watson zur Prozessausgliederung stellt der Chatbot EVA („Empathische Versicherungsassistentin“) der INTER Krankenversicherung AG dar. Er unterstützt Kunden und Interessenten beim Online-Abschluss von Zahnzusatzversicherungen (vgl. Dinzler 2017). EVA beantwortet allgemeine Fragen, beispielsweise zum Leistungsumfang oder zahnmedizinischem Fachvokabular, und nimmt Rückrufwünsche auf (Geschäftsvorfall, benötigte Daten sind unter anderem Telefonnummer und Rückrufzeitpunkt)[9]. Man erkennt hier den Komplexitäts– und Benutzerfreundlichkeitsnutzen.

An diesem konkreten Fall lässt sich zudem die Sinnhaftigkeit der in Abschnitt 4.2 vorgeschlagenen Umsetzungsreihenfolge aufzeigen. Bevor der Einsatz des kognitiven Systems an der Kundenschnittstelle in Frage kommt, muss der zugrundeliegende, eigentliche Prozess (Antragsbearbeitung) entwickelt und automatisiert sein. Es wäre beispielsweise wenig zielführend, einen online gestellten Antrag zu drucken, in Papierform zu verteilen und von einem Kundenbetreuer neu eingeben zu lassen – ein qualitativ hochwertiger Prozess sähe stattdessen eine direkte Schnittstelle in das bestandsführende System vor (Maschine-zu-Maschine-Kommunikation). Er würde darüber hinaus einfache Standardfälle regelbasiert erkennen und automatisieren (beispielsweise keine Vorerkrankungen, keine Zahlungsausfälle im Bestand usw.), so dass dem Kunden unmittelbar nach Antragstellung die Vertragsannahme per E-Mail bestätigt wird.

Das Basissystem auf Grundlage von Watson Conversation und/oder Virtual Agent ist durch den Aufruf weiterer Module der Watson Developer Cloud aus

[9] EVA ist unter der URL https://schoenezaehne.inter.de/ aufrufbar.

dem Steuerprogramm erweiterbar. Durch Hinzufügen von Watson Text to Speech und Speech to Text (vgl. Abschnitt 5.4.10) sind sprachgesteuerte, digitale Assistenten à la Siri (Apple), Cortana (Microsoft) oder Google Now denkbar (vgl. Azraq u. a. 2017, S. 1–3). Watson Tone Analyzer (vgl. Abschnitt 5.4.7) erlaubt Frustration oder Ärger des Kunden zu erkennen und darauf zu reagieren; beispielsweise durch eine Eskalation an den menschlichen Kundendienst und/oder der Priorisierung aufgenommener Anliegen. Allerdings unterstützt von den genannten Modulen nur Watson Text to Speech deutsche Texte.

Bezüglich der Anforderungen und Voraussetzungen im Sinne des Vorgehensmodells ergeben sich für dieses Anwendungsszenario folgende Besonderheiten.

Je nach gewähltem Liefermodell ist mit potenziellen Ausfallzeiten von 43:48 Stunden (Premium, 99,5% SLA) beziehungsweise 4:23 Stunden (Shared, SLA 99,95%) im Jahr zu rechnen, welche durch den Einsatz an der Kundenschnittstelle von diesem negativ wahrgenommen werden könnte. Um zumindest eine rudimentäre Verfügbarkeit gegenüber dem Kunden zu gewährleisten, hat das Steuerprogramm die Nicht-Verfügbarkeit des Dienstes abfangen und den Kunden auf andere Kontakt– beziehungsweise Informationsmöglichkeiten (beispielsweise: Telefon, E-Mailadresse, statische FAQ auf der Seite usw.) zu verweisen.

Bei Einstellung dies Dienstes hätte das Unternehmen die Möglichkeit, auf die weiterhin vorhandenen Kontaktmöglichkeiten (beispielsweise Chat mit Anwender, Telefonie usw.) zurückfallen. Der Export des Dialoges im JSON-Format ist möglich. Auch wenn der direkte Import in ein anderes System fraglich erscheint, so sind die Daten zumindest menschenlesbar, so dass die darin enthaltene Struktur und das Know-how erhalten bleiben. Steuerprogramm, Oberfläche und Backend-Systeme liegen ohnehin in der Verantwortung des Unternehmens.

Die analysierten Eingaben der Kunden können Personenbezug aufweisen (insbesondere im Rahmen des gespeicherten $Contexts) und damit den datenschutzrechtlichen Vorschriften unterliegen. Es ist entweder eine ADV mit IBM abzuschließen, welche die in Abschnitt 3.2.2 genannten Kriterien erfüllt, oder das explizite Einverständnis der Kundin zur Übermittlung ihrer Daten einzuholen, beispielsweise durch Klick oder Checkbox zu Beginn des Dialoges.

Bevor das Training des Systems angegangen wird, ist zunächst eine ABC-Analyse über die Anfragen und Anliegen der Kunden durchzuführen – sofern nicht bereits bei den Kundenbetreuern als heranzuziehende Experten vorhanden. Diese können auch Musterfragen (als Trainingsmenge der Intents, müssen als CSV bereitgestellt werden) sowie die passenden Antworten bereitstellen; in der Regel werden für häufige Anliegen bereits Textbausteine, FAQ usw. bereitstehen und können als Basis für den Dialog dienen.

Die Ergebnisse können über Tests evaluiert werden. Für das dauerhafte Con-

trolling nach Produktionseinsatz steht in der IBM Cloud eine Komponente namens „Improve“ als SaaS bereit, welche sämtliche Gesprächsverläufe des Agenten protokolliert und so erlaubt, fehlende #Intents, @Entities, Formulierungen usw. zu erkennen und zu ergänzen. Watson Conversation beginnt unmittelbar den Trainingsprozess und verwendet die neuen Ergebnisse sofort nach dessen Abschluss (vgl. Azraq u. a. 2017, S. 185–202).

Als letzte Erweiterung ist die Anbindung von Watson Discovery (vgl. Abschnitt 5.4.5) für Sachverhalte denkbar, die nicht explizit in Conversation als #Intent modelliert sind, insbesondere für C-Fälle der ABC-Analyse. Dazu wird in Discovery eine Wissensdatenbank aus unstrukturierten Texten aufgebaut. Aktiviert in Watson Conversation der „anything_else“–Knoten, startet das Steuerprogramm eine Suche mit den Kundeneingaben in Discovery und zeigt diese an. Hierbei ist jedoch zu beachten, dass der Funktionsumfang für deutsche Texte eingeschränkt ist und manuelle Modelle für Entitäten und Relationen zu modellieren sind; auf die Voraussetzungen des Trainings wird im nächsten Abschnitt eingegangen.

6.2.2 Prozessablaufgestaltung und Automatisierung

Zur Prozessgestaltung und Automatisierung mit kognitiven Systemen kommen drei Einsatzmöglichkeiten in Betracht: Klassifikation (inklusive Priorisierung und Routing), Extrahieren und Anreichern von Informationen (information extraction) und das Zugänglichmachen von im Unternehmen vorhandenem Wissen für interne Anwender wie beispielsweise Sachbearbeiter in Form von Wissensdatenbanken (information retrieval).

Vor Betrachtung der Besonderheiten der jeweiligen Szenarios und Module werden zunächst allgemeingültige Punkte dargestellt.

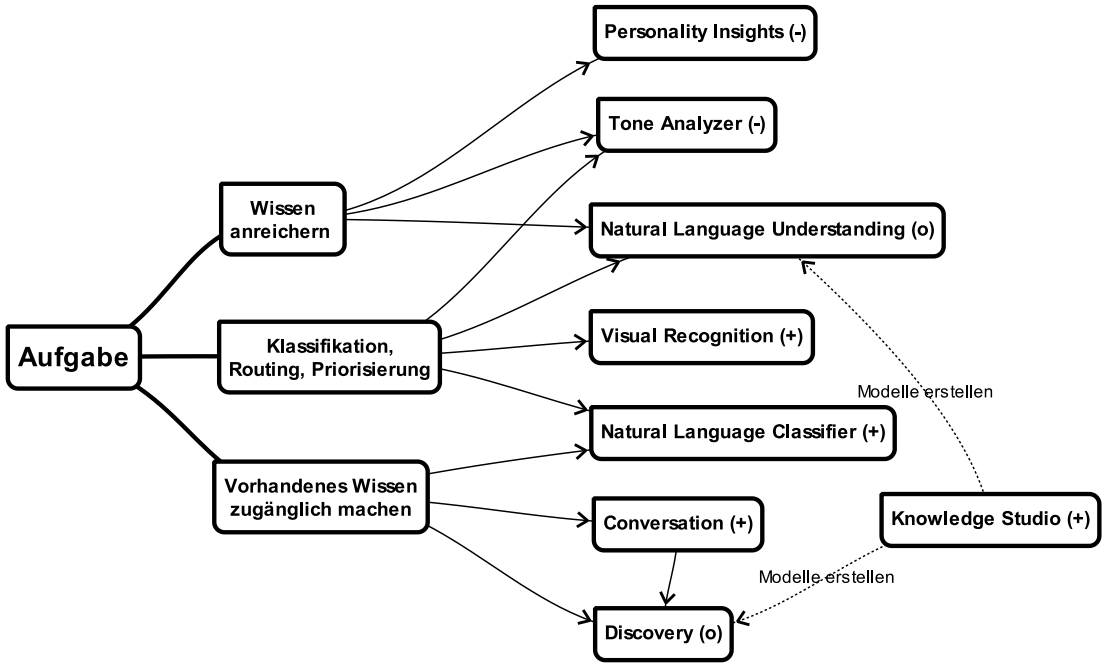
- Für die Übermittlung personenbezogener Daten an einen CSP ist grundsätzlich die Einwilligung des Betroffenen oder der Abschluss einer ADV mit dem CSP erforderlich (sh. Abschnitt 3.2.2).
- Zum Trainieren eines Dienstes sind sauber anonymisierte (und damit nicht mehr dem Datenschutz unterliegende) Datensätze ausreichend.
- Unternehmensinterna (Textkonserven, Arbeitsanweisungen, Richtlinien, Dokumentationen, interne Abläufe usw.) sind hingegen in der Regel datenschutzrechtlich unbedenklich – hier ist jedoch eine Analyse der mit dem Außer-Haus-geben verbundenen Risiken erforderlich. Dabei handelt es sich nicht nur um eine Vertrauensfrage ggü. dem CSP, sondern auch um Risiken welche aus der Virtualisierung in der Cloud resultieren. So wurden Anfang 2018 unter den Namen „Meltdown“ und „Spectre“ gravierende Sicherheitslücken in Prozessoren bekannt, welche erlauben aus der Virtualisierung „auszubrechen“ und unberechtigt auf fremde Daten auf der gleichen Hardware zuzugreifen (vgl. Lipp u. a. 2018; und Kocher u. a. 2018).

- Dienstleistungsprozesse weisen im Vergleich zur Produktion materieller Güter noch viel Potential für Automatisierung auf, insbesondere aufgrund ihrer Abhängigkeit von unstrukturierten Daten (vgl. Best und Weth 2010, S. 155).

Abbildung 12 zeigt die Einsatzmöglichkeiten der Watson Developer Cloud zur Ablaufgestaltung.

Klassifikation teilt Daten in festgelegte Klassen ein. Eine typische Anwendung ist die Klassifizierung des zentralen Eingangs eines Unternehmens (Post, Faxe, zentrale Emailadressen à la info@... usw.), um die zahlreichen, eingehenden Schriftstücke direkt an den jeweils Zuständigen zu steuern (Routing). Korrekte Klassifikation verringert Liege-, Bearbeitungs- und damit Durchlaufzeiten und erlaubt darüber hinaus die Priorisierung wichtiger Anliegen (beispielsweise Beschwerden, Gerichtspost, Terminsachen usw.). Sie ist ein typischer Kandidat für die (Teil–)Automatisierung, abhängig von der Qualität und dem Konfidenzgrad.

Abbildung 12: Anwendung der Watson Dienste zur Prozessablaufgestaltung und Automatisierung



Quelle: eigene Darstellung

Für dieses Einsatzgebiet kommen vier Module der Watson Developer Cloud in Frage. Watson Natural Language Classifier (vgl. Abschnitt 5.4.3) klassifiziert kurze Texte bis 1.024 Zeichen in frei wählbare Klassen, kommt damit insbesondere für E-Mails und Social Media in Betracht; längere Schriftstücke nur, wenn diese in einzelne zu klassifizierende Bestandteile (Absätze oder Sätze)

zerlegt werden. Um die Repräsentativität der Trainingsmenge sicherzustellen, ist sie realen Kundenanliegen des betroffenen Kommunikationskanales zu entnehmen und von den Expertinnen der unternehmenseigenen Kundenbetreuung manuell zu klassifizieren. Die Trainingsmenge ist entweder direkt im Browser (SaaS) einzugeben oder über eine Unicode-codierte CSV-Datei bereitzustellen (vgl. Manhaes u. a. 2017, S. 1–4); es empfiehlt sich die Pflege der Trainingsmenge in einer Tabellenkalkulation wie Microsoft Excel oder LibreOffice Calc, welche in das benötigte Format exportieren kann und gleichzeitig als lokale Sicherung dient. Eingriffe in den Lernprozess sind nur indirekt durch das Hinzufügen und Entfernen von uneindeutigen Texten oder dem Korrigieren der Klassenzuordnung eines Textes möglich (vgl. ebenda, S. 73–85).

Watson Visual Recognition (vgl. Abschnitt 5.4.9) erlaubt die Klassifikation von Bildern in frei wählbare Klassen. Es sind von den passenden Experten reale Fälle zu sammeln und manuell zu klassifizieren, empfohlen sind mehrere tausend Bilder pro Klasse (vgl. IBM Deutschland GmbH 2017a). Auch Bilder können personenbezogene Daten darstellen (beispielsweise Bilder eines Autowracks mit amtlichen Kennzeichen), so dass zumindest für die eigentliche Klassifikation entweder die Einwilligung des Betroffenen oder die Vereinbarung einer ADV erforderlich ist.

Watson Tone Analyzer (vgl. Abschnitt 5.4.7) kann zur Erkennung von Beschwerden eingesetzt werden, unterstützt jedoch keine deutschen Texte. Training des Moduls ist nicht vorgesehen. Watson Natural Language Understanding (vgl. Abschnitt 5.4.4) arbeitet anders als die bislang vorgestellten Module: Es bewertet nicht das Gesamtdokument, sondern extrahiert aus deutschen Texten Entitäten, Relationen, Schlüsselwörter, Metadaten und eingeschränkt Einstellungen/Ansichten.

Ein konkretes Beispiel hierfür ist die Erkennung, Priorisierung und Steuerung von Beschwerden bei der Versicherungskammer Bayern beziehungsweise Union Krankenversicherung AG[10] (vgl. Scheerer 2016; ergänzende Details sh. Fromme 2015). Hier wurden unter anderem drei Entitäten trainiert: der Auslöser der Beschwerde (beispielsweise Kürzungen der Versicherungsleistung, fehlende Antwort, Bearbeitungszeiten usw.), die Unmutsäußerung des Beschwerdeführers und dessen Forderung (beispielsweise die Bezahlung des Restbetrages oder Abschluss der Schadenabwicklung).

Für die Anreicherung von Wissen kommen drei Module in Frage. Die per Natural Language Understanding aus dem Dokument extrahierten Informationen sind nicht nur zur Kategorisierung nutzbar, sondern insbesondere zur Anreicherung als strukturierte Metadaten zur weiteren Verwendung. Ziehen wir als Beispiel nochmals den oben geschilderten Anwendungsfall der Versicherungskammer Bayern heran. In einem Schreiben oder einer E-Mail zu einer

[10] Die Union Krankenversicherung ist ein 100%iges Tochterunternehmen der Versicherungskammer Bayern, vgl. https://www.vkb.de/content/ueber-uns/unternehmen/konzern/

KFZ-Schadenangelegenheit finden sich zahlreiche Informationen, welche der Anwender manuell in die versicherungstechnischen Systeme übernehmen muss, beispielsweise amtliches Kennzeichen, Name und Postanschrift des Unfallgegners und der Zeugen. In Knowledge Studio können passende, domänenspezifische Modelle erstellt und per Natural Language Understanding auf eingehende Texte angewandt werden. Die so erzeugten Metadaten sind vielseitig nutzbar: von der farblichen Markierung der Stellen im Dokument für den Anwender, über das Vorbelegen dieser Daten in den Erfassungsmasken der versicherungstechnischen Anwendungen, bis hin zur automatisierten Anlage von Datensätzen für Zeugen und Anspruchsteller. Ein weiteres Beispiel bietet der japanische Lebensversicherer Fukoku Mutual Life Insurance, welcher mit Watson Informationen aus medizinischen Berichten extrahiert und damit den Leistungsprozess teilweise automatisiert (vgl. Welter 2017).

Ob Watson Natural Language Understanding zur Informationsextraktion oder zur Klassifizierung genutzt wird: Die benötigten Entitäten und Relationen sind in der SaaS Watson Knowledge Studio (vgl. Abschnitt 5.4.6) unter Berücksichtigung des skizzierten Vorgehensmodelles zu trainieren. Dazu sind zunächst auf Grundlage der Ziele die benötigten Entitäten, die möglichen Relationen zwischen diesen Entitäten (type system) sowie die Ermittlungsmethode festzulegen und passende Dokumente zu sammeln. Die regelbasierende Erkennung bietet sich beispielsweise für Listen (Wochentage, Monatsnamen usw.), reguläre Ausdrücke und ähnliche, einfach beschreibbare Entitäten an; für komplexere Fälle empfiehlt sich ein maschinell angelerntes Modell. Erst nach diesen Vorarbeiten werden die Dokumente als Text– oder CSV–Dateien importiert, Typsystem, Anwender, Arbeitspakete in der SaaS angelegt und das Training begonnen. Zur Überprüfung der Homogenität der Annotationen stellt Knowledge Studio einen „Inter-Annotator Agreement Score“ bereit.

Die Ergebnisse von Watson Tone Analyzer könnten analog zum Einsatz als Klassifizierer als Metadaten am Dokument angereichert werden, um diese beispielsweise Kundenbetreuern anzuzeigen oder Dokumente mit auffälligen Bewertungen hervorzuheben.

Watson Personality Insights ist zur Erstellung von Persönlichkeitseinschätzungen gedacht und benötigt daher mehrere Texte eines Autors (beispielsweise Social Media-Einträge). Die Ergebnisse könnten ebenfalls in der Kundendatenbank angereichert und für Marketing-Kampagnen (Selektion) oder zur Unterstützung der Mitarbeiterin im Beratungs- oder Verkaufsgespräch dienen. Aus datenschutzrechtlicher Sicht handelt es sich um einwilligungspflichtiges Profiling. Als Ausnahmeregelung ist die „Generalklausel“ mit der Interessenabwägung zwischen der Betroffenen und den Unternehmen denkbar, insbesondere wenn es sich um Beiträge in den sozialen Medien handelt, welche der Betroffene für die Allgemeinheit veröffentlicht hat (vgl. Wybitul 2016, 162 und 170; und Eschholz 2017, S. 183–184). Für die Übermittlung der Daten an IBM

darüber hinaus wie üblich die Einwilligung des Betroffenen oder die Vereinbarung einer ADV erforderlich. Sowohl Watson Tone Analyzer als auch Watson Personality Insight unterstützen derzeit keine deutschen Texte.

Für das Zugänglichmachen von vorhandenem Wissen (information retrieval) in Form einer Wissensdatenbank bietet sich insbesondere Watson Discovery an: Es liest Dokumente (beispielsweise Arbeitsanweisungen oder Informationsseiten) ein und reichert diese analog zu Natural Language Understanding mit Metadaten an (sh. Abschnitt 5.4.5).

Für den Aufbau der Wissensdatenbank sind zwei Faktoren relevant: erstens, wem welches Wissen bereitzustellen ist (Anwenderkreis, Quellen für Dokumente, deren Format usw.) und zweitens, welches Typensystem für Entitäten und Relationen herangezogen wird. Zwar sind in Watson Knowledge Studio erstellte Modelle sowohl in Watson Natural Language Understanding als auch in Watson Discovery einsetzbar (der Trainingsvorgang ist unabhängig vom Zielsystem und wurde bereits behandelt), abhängig vom Einsatzzweck kann jedoch eine modifizierte Kopie oder sogar ein völlig neues Typensystem angebracht sein. Ein Typensystem zur Beschwerdeerkennung eignet sich beispielsweise nicht zur Dokumentation von Vertragsklauseln und deren Zusammenspiel.

Die APIs von Watson Discovery sehen nur Backend-Funktionalität vor. Daher muss eine eigene Benutzeroberfläche geschaffen werden, welche Anfragen der Benutzerin entgegennimmt und sie bei der Syntax der Abfragesprache unterstützt.

Als Erweiterung ist denkbar, die häufigsten Fragen (A-Fragen) in Conversation zu modellieren und der Suche in Discovery vorzuschalten (analog zum Einsatz Conversation/Discovery an der Kundenschnittstelle; statt eines Dialoges steht hier aber der schnelle Zugriff im Vordergrund). Conversation deckt hierbei die A-Fragen bzw- A-Informationen ab, die restlichen Anfragen werden durch Discovery übernommen. Vorteil dieser Lösung ist, dass die A-Fälle (welche in der Regel ca. 80% der Anfragen verursachen) explizit modellierbar sind.

Auch Watson Natural Language Classifier kann als einfache Wissensdatenbank verwendet werden, indem Wissen passend zur erkannten Klasse abgelegt und verfügbar gemacht wird. Beispiel: eine Versicherung trainiert u .a. die Klasse „Kündigung“ für E-Mails an. Bearbeitet der Kundenbetreuer eine solche E-Mail, werden passende Arbeitsanweisungen, Checklisten und Textvorschläge zu dieser Kategorie eingeblendet: beispielsweise die einzuhaltenden Kündigungsfristen und Formvorschriften für diesen Vertrag.

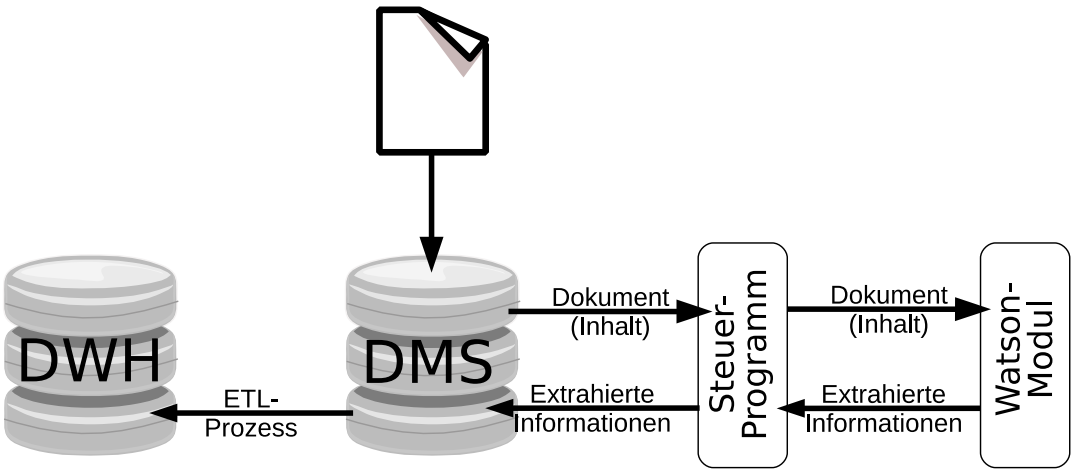
6.2.3 Unterstützung Business Intelligence

Im letzten Abschnitt wurden die Fähigkeiten der kognitiven Watson Developer Cloud genutzt, aus unstrukturierten Daten strukturierte Informationen zu entnehmen (information extraction) und sofort operativ im Prozess zu nutzen. Eine

weitere Anwendungsmöglichkeit besteht darin, diese Daten dauerhaft zu speichern und im Rahmen der BI einzubeziehen – beispielsweise für Kundenanalysen im Rahmen des Customer-Relationship-Management (CRM), Wettbewerbs– und Marktanalysen oder im unternehmenseigenen Wissensmanagement. Die Verfahren des klassischen Data Minings benötigen strukturierte Daten, und auch DWHs sind primär auf strukturierte Daten ausgerichtet (vgl. Cleve und Lämmel 2016, S. 38–39; und Kemper, Mehanna und Baars 2010, S. 117–118). Das kognitive System übernimmt insoweit die Aufgabe des Text Minings.

Abbildung 13 zeigt die Aufbereitung dieser Daten für das DWH. Während das unstrukturierte Dokument (beispielsweise eine E-Mail oder ein Schreiben) inklusive der dazugehörigen Metadaten im Dokumentenmanagement-System (DMS) abgelegt wird und dort verbleibt, werden die von Watson Developer Cloud angereicherten Daten zusammen mit einem Verweis auf das Dokument (beispielsweise eine eindeutige DokumentenID) und ggf. weiteren, benötigten Daten per ETL–Werkzeuge in das DWH kopiert (dispositive Datenhaltung). Der Vorteil gegenüber der Nutzung eines Text-Mining-Werkzeuges im Rahmen des ETL–Prozesses ist, dass die Daten zusätzlich operativ genutzt werden können (wie im letzten Abschnitt dargestellt, beispielsweise zur Klassifizierung).

Abbildung 13: Prozess zur Aufbereitung unstrukturierter Inhalte für DWHs



Quelle: eigene Darstellung, basierend auf Baars und Kemper 2008, S. 139

Führen wir den Anwendungsfall aus dem letzten Abschnitt beispielhaft fort. Watson Natural Language Understanding klassifiziert Beschwerden und extrahiert dabei unter anderem den Auslöser und die Forderung des Kunden. Diese Informationen werden im DMS als Metadaten an das Dokument gehängt und sind damit den BI–Analysesystemen für Auswertungen zugänglich. Die Managerin ist so in der Lage, Schwachstellen in den jeweiligen Prozessen bezüglich der Effektivität (Kundenzufriedenheit) zu erkennen und kann sogar bis die Ebene einzelner Dokumente vordringen (drill-down).

Der skizzierte Prozess eignet sich auch für das nachträgliche Überarbeiten bereits vorhandener Dokumentenbestände. Hierzu ist ein Programm zu entwickeln, welches im Stapelbetrieb den Inhalt alter Dokumente extrahiert (beispielsweise durch OCR), diesen vom Watson-Modul auswerten lässt und die zurückgegebenen Ergebnisse im DMS ergänzt. Falls die Daten ausschließlich im dispositiven Datenbestand benötigt werden, kann dieses Programm auch im Rahmen des ETL-Prozesses als zusätzliche Komponente betrieben werden.

Als Module kommen die im letzten Abschnitt genannten APIs zur Anreicherung von Wissen in Frage (Tone Analyzer, Personal Insight, Natural Language Understanding), mit den dort bereits genannten Vorbedingungen. Insbesondere Personal Insight, welches Bedürfnisse und Konsumvorlieben des Kunden erkennen soll und Persönlichkeitsanalysen durchführt, wirkt geeignet für zielgruppenorientiertes Marketing und CRM. Watson Tone Analyzer bietet einen zusätzlichen Modus, welcher den Gesprächsverlauf zwischen Kundenbetreuer und Kunde bewertet und so ermöglicht, die Leistung des Mitarbeiters einzuschätzen. Der Einsatz beider Module scheitert abermals an der fehlenden Unterstützung für deutschsprachige Texte. Neben Dokumenten kommen auch Texte aus Social-Media-Kanälen, Foren und der unternehmenseigenen Website (beispielsweise Produktbewertungen) für Auswertungen in Frage, sofern diese zweifelsfrei zuordenbar sind.

In diesem Kapitel wurden Module der Watson Developer Cloud eingesetzt, um bestehende BI–Lösungen durch das Zugänglichmachen von unstrukturierten Informationen zu unterstützen beziehungsweise deren Fertigkeiten zu erweitern. Betrachtet man nochmals die Architektur eines kognitiven Systems in Abschnitt 2.4.3 und die dort aufgezeigten Parallelen zum Data Mining, fällt auf, dass man diese Schritte in einem kognitiven System vereinigen könnte: Einlesen unstrukturierter Dokumente, Extraktion der benötigten Daten (feature extraction), Speicherung der Features in einen Corpus, um sie dort mit Datenanalysediensten auszuwerten (unter anderem mit Verfahren des Data Mining) und dem Anwender verfügbar zu machen. Einen solchen Ansatz verfolgt IBM Watson Explorer (vgl. Chen u. a. 2014, S. 5–9), welches jedoch nicht Bestandteil der in dieser Arbeit eruierten Watson Developer Cloud ist.

6.3 Know-how- und entscheidungsintensive Prozesse (Typ I und II)

Der Einsatz kognitiver Computersysteme wird oft in der Unterstützung anspruchsvoller, menschlicher Wissensarbeit gesehen, wie sie insbesondere in Prozessen des Typ I, aber auch Typ II vorkommt (vgl. Hull und Nezhad 2016, S. 4). Dabei sollen Beschränkungen und Schwachpunkte des menschlichen Denkens überwunden werden (vgl. Kelly und Hamm 2014, S. 11–16; Weber 2015a, S. 31–40; und Gliozzo u. a. 2017, S. 4–6,12):

1. Menschen sind nicht in der Lage, große Datenmengen oder komplexe Systeme mit interdependanten Akteuren vollständig zu durchdringen.

Kognitive Systeme können dadurch völlig neue Lösungsansätze erschließen („kognitive Sonde“, vgl. Weber 2015a, S. 39). Beispiel: die KI AlphaGo basiert unter anderem auf mehreren orchestrierten KNN und besiegte 2016 den Spitzenspieler Lee Sedol in Go. Dabei zeigte das System neuartige Strategien und innovative Zugfolgen – in einem der ältesten Brettspiele der Menschheit (vgl. Bögeholz 2016).

2. Kognitive Systeme sollen fehlende Expertise ausgleichen können, beispielsweise bei bereichs– beziehungsweise domänenübergreifenden Problemen oder Daten.
3. Menschen unterliegen zahlreichen Biasen (Wahrnehmungsverzerrungen) und Gruppendynamik, welche Entscheidungen irrational beeinflussen. Bekannte Beispiele sind der Framing-Effekt (die Präsentation einer Information beeinflusst die Entscheidung), die Neigung zum Status Quo und die systematische Fehleinschätzung von Risiken (vgl. Magrabi und Bach 2013, S. 275–277). Darüber hinaus prägen Ausbildung, beruflicher Werdegang und Lebensweg die Wahrnehmung jedes einzelnen Individuums.

Die Zusammenarbeit zwischen Mensch und kognitivem Computersystem spiegelt dabei die Theorien der dualen Entscheidungsfindung: Während Menschen als System 1 fungieren und Hypothesen generieren, analysiert und bewertet System 2 in Gestalt des kognitiven Computersystems diese in einer für Menschen unerreichbaren Rationalität.

Punkt 1 ist wohl zuzustimmen, auch wenn es sich nicht um ein Alleinstellungsmerkmal des Cognitive Computing handelt, sondern analog für KNN und den Themenkomplex Data Mining, BI und Big Data gilt.

Punkt 2 widerspricht dem in Abschnitt 6.1 dargestellten Vorgehensmodell, welches in Phase 2 explizit das Hinzuziehen von Experten der jeweiligen Wissensdomäne vorsieht. Art, Zusammensetzung und Qualität der Trainingsmenge haben entscheidenden Einfluss auf die Ergebnisse des kognitiven Systems, gleichzeitig kann dieses seine Resultate nicht begründen[11] und dadurch auf Probleme in der Trainingsmenge aufmerksam machen. Die Zusammenstellung und Untersuchung der Trainingsmengen durch Experten scheint daher zwingend erforderlich.

Punkt 3 birgt die Gefahr, die Ergebnisse des kognitiven Computersystems als objektive, unumstößliche Wahrheit aufzufassen, was jedoch aufgrund deren Abhängigkeit von den Trainingsmengen nicht angemessen ist. Wahrnehmungsverzerrungen, denen auch Experten unterliegen, können zu Biasen in der Auswahl der Trainingsmenge, deren Klassen und der jeweiligen Datensätze führen und damit wiederum zu verzerrten Ergebnissen des kognitiven Computersystems (Waters 2016). Um Biase auszugleichen, empfiehlt sich ein breit aufgestelltes Team mit unterschiedlichen persönlichen und fachlichen Hintergründen

[11] Es existieren jedoch erste Ansätze zur Plausibilisierung durch Visualisierung, beispielsweise https://www.research.ibm.com/cognitive-computing/watson/watsonpaths.shtml.

und Werdegängen.

Untersucht man die Module der Watson Developer Cloud auf Eignung für den Einsatz als kognitive Sonde, fällt auf, dass keines davon selbstständig Schlussfolgerungen aus seinem Input generiert. Watson Discovery und Natural Language Understanding extrahieren zwar Informationen, und Watson Tone Analyzer und Personal Insight nehmen Bewertungen vor; jedes dieser Module überlässt jedoch die weitere Verwertung dem aufrufenden Programm beziehungsweise der abfragenden Anwenderin.

Watson Developer Cloud kann daher Wissensarbeit in Typ I//II-Prozessen nur auf zwei (eingeschränkte) Arten unterstützen: durch Erstellung einer Wissensdatenbank zur Entscheidungsunterstützung in Watson Discovery (analog zu Abschnitt 6.2.2); und durch Erschließen von strukturierten Informationen aus unstrukturierten Dokumenten für weitergehende Analysen im Rahmen der BI (analog zu Abschnitt 6.2.3). Der Unterschied zu dem bereits aufgezeigten Vorgehen besteht darin, dass beliebige – nicht zwingend operativ genutzte – Dokumente ausgewertet werden.

7 Schlussfolgerung und Ausblick

Cognitive Computing ist keine Revolution, sondern eine Evolution der beteiligten Disziplinen: die Technologisierung der Kognitionswissenschaft, verbunden mit der Fortführung der KI-Forschung und Einbeziehung von Data Mining, BI und Big Data. Das entscheidende Merkmal ist die Orchestrierung der beteiligten Werkzeuge, Techniken, Erkenntnisse und Disziplinen um intelligenzähnliche Leistungen zu erzielen.

Für den Einsatz cloudbasierter, kognitiver Computersysteme zur Prozessentwicklung wurden zwei Vorgehensmodelle aufgezeigt. Das erste Modell wird im operativen Prozessmanagement angewendet und verfolgt zwei Ziele: Sicherstellen einer ausreichenden Prozessqualität als Voraussetzung und Ausgangspunkt für den Einsatz kognitiver Computersysteme; und die systematische Betrachtung der mit kognitiven Systemen verbundenen Voraussetzungen, Anforderungen und Risiken. Als zweites wurde CRISP-DM, ein Modell für Data Mining-Prozesse, mit leichten Modifikationen zum systematischen Trainieren eines kognitiven Systems (oder auch KNN) herangezogen. Dies ist erforderlich, da das Training entscheidenden Einfluss auf die Ergebnisse derartiger Systeme hat; diese aber gleichzeitig ihre Ergebnisse nicht begründen können.

Anschließend wurde das Potential kognitiver Computersysteme zur Prozessverbesserung anhand eines konkreten Systems, IBM Watson Developer Cloud, untersucht (unter Berücksichtigung der beiden Vorgehensmodelle). Im Ergebnis bestehen bei Routine-Prozessen (Typ III) vier Anwendungsmöglichkeiten.

- Prozessauslagerung an den Kunden: virtuelle Agenten sind in der Lage, über verschiedenste Eingangskanäle (Mobilapplikationen, E-

Mail, Chats, Messaging usw.) Dialoge zu führen, um allgemeine und kundenspezifische Fragen zu beantworten und Geschäftsvorgänge zu moderieren.
- Prozessablaufgestaltung sowie Automatisierung durch Klassifikation, Routing und Priorisierung von unstrukturierten Dokumenten; dem Verfügbarmachen von Wissen (information retrieval) und dem Extrahieren von Informationen aus unstrukturierten Dokumenten (information extraction).
- Informationsextraktion zur Aufbereitung von Daten für Data Mining im Rahmen bereits vorhandener BI.

Dies gilt insbesondere für Dienstleistungsprozesse, aufgrund deren Abhängigkeit von unstrukturierten Daten; es ist kein Zufall, dass alle aufgeführten Praxisbeispiele aus der Versicherungsbranche stammen. Das größte Hindernis für den produktiven Einsatz stellt die teilweise fehlende beziehungsweise eingeschränkte Unterstützung deutscher Sprache dar.

Der Einsatz der Watson Developer Cloud ist auch im Rahmen von Know-how- beziehungsweise entscheidungsintensiven Prozessen (Typ I/II) denkbar: zur Erstellung von Wissensdatenbanken und Erschließung unstrukturierter Dokumente für BI, analog zu Routine-Prozessen. Das System kann jedoch nicht automatisiert aus Datenbeständen lernen und so menschliche Expertise ersetzen oder menschliche Biase ausgleichen. Der in der Einleitung zitierten Prognose von Gartner Inc. (vgl. Panetta 2017), welche diese Art von Einsatz erst in fünf bis zehn Jahren erwartet, kann nicht widersprochen werden.

Ziel dieser Arbeit war, einen Überblick über das Trendthema Cognitive Computing zu geben und dessen betriebswirtschaftliche Einsatzmöglichkeiten in Geschäftsprozessen aufzuzeigen. Aufgrund des begrenzten Umfanges hat sie sich dabei auf ein kognitives System, die Watson Developer Cloud, beschränkt. Es existieren jedoch weitere kognitive Plattformen, beispielsweise Microsoft Azure AI[12], WiPro Holmes[13] und Google Cloud Machine Learning Engine[14].

Ein systematischer Vergleich von kognitiven Plattformen, beispielsweise unter den Kriterien Funktionsumfang (inkl. Sprachunterstützung), technische Rahmenbedingungen und insbesondere Leistungsfähigkeit könnte einen lohnenden Aspekt zur Vertiefung in weiteren wissenschaftlichen Arbeiten darstellen.

[12] Vgl. https://azure.microsoft.com/de-de/overview/ai-platform/
[13] Vgl. https://www.wipro.com/holmes/
[14] Vgl. https://cloud.google.com/ml-engine/

8 Literaturverzeichnis

Amberg, Michael, Freimut Bodendorf und Kathrin M. Möslein (2011). Wertschöpfungsorientierte Wirtschaftsinformatik. Band 4. Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg. ISBN: 978-3-642-16755-3. DOI: 10.1007/978-3-642-16756-0.

Arbeitskreise Technik und Medien der Konferenz der Datenschutzbeauftragten des Bundes und der Länder (2014). Orientierungshilfe Cloud Computing Version 2.0. URL: https://www.datenschutz-bayern.de/technik/orient/oh_cloud.pdf (letzter Zugriff 01. 02. 2018).

Auer, Sören u. a. (2007). »DBpedia: A Nucleus for a Web of Open Data«. In: The Semantic Web. Herausgegeben von David Hutchison u. a. Band 4825. Lecture notes in computer science. Berlin, Heidelberg: Springer Berlin Heidelberg, S. 722–735. ISBN:978-3-540-76297-3. DOI: 10.1007/978-3-540-76298-0_52.

Azraq, Ahmed u. a. (2017). Building Cognitive Applications with IBM Watson Services: Volume 2 Conversation. IBM Redbooks. Poughkeepsie, NY: IBM Corporation, International Technical Support Organization. ISBN: 978-0-7384-4256-3. URL: https://www.redbooks.ibm.com/Redbooks.nsf/RedbookAbstracts/sg248394.html?Open (letzter Zugriff 01. 02. 2018).

Baars, Henning und Hans-George Kemper (2008). »Management Support with Structured and Unstructured Data—An Integrated Business Intelligence Framework«. In: Information Systems Management 25.2, S. 132–148. ISSN: 1058-0530. DOI: 10.1080/10580530801941058.

Bager, Jo (2016). »Die KI-Revolution. Vom Siegeszug der lernenden Software«. In: c't magazin für computertechnik 06/2016, S. 124–129.

Baker, Stephen (2012). Final jeopardy. The story of Watson, the computer that will transform our world. Boston, MA: Mariner Books. ISBN: 978-0-547-74719-4.

Best, Eva und Martin Weth (2010). Process Excellence. Praxisleitfaden für erfolgreichesProzessmanagement. 4. Auflage. Wiesbaden: Gabler Verlag. ISBN: 978-3-8349-2211-3. DOI: 10.1007/978-3-8349-8950-5.

Bögeholz, Harald (2016). »Jubel und Ernüchterung. Google AlphaGo schlägt Top-Profi 4:1 im Go«. In: c't magazin für computertechnik 07/2016, S. 44.

Bruhn, Manfred u. a. (2013). »Qualität und Nutzen von Avataren als Dienstleister im Social Web – Messung und Konsequenzen«. In: Dienstleistungsmanagement und Social Media. Herausgegeben von Manfred Bruhn und Karsten Hadwich. Wiesbaden: Springer Gabler, S. 443–466. ISBN: 978-3-658-01247-2. DOI: 10.1007/978-3-658-01248-9_20.

Brutzman, Don (2015). ACTION-732: Investigate relationship between ECMA's JSON spec and IETF RFC 7159. URL: https://www.w3.org/2005/06/tracker/exi/actions/732 (letzter Zugriff 01. 02. 2018).

Buchner, Benedikt (2016). »Grundsätze und Rechtmäßigkeit der Datenverarbeitung unter der DS-GVO«. In: Datenschutz und Datensicherheit (DuD) 40.3, S. 155–161. DOI: 10.1007/s11623-016-0567-0.

Bundesministerium der Justiz und für Verbraucherschutz (1990). Bundesdatenschutzgesetz (BDSG). URL: https://www.gesetze-im-internet.de/bdsg_1990/ (letzter Zugriff 01.02.2018).

Carstensen, Kai-Uwe u. a., Herausgeber (2010). Computerlinguistik und Sprachtechnologie. Eine Einführung. 3. Auflage. Heidelberg: Spektrum Akademischer Verlag. ISBN: 978-3-8274-2023-7.

Chen, Whei-Jen u. a. (2014). Building 360-degree information applications. IBM Redbooks. IBM Corporation, International Technical Support Organization. ISBN: 978-0-7384-3992-1. URL: https://www.redbooks.ibm.com/abstracts/sg248133.html (letzter Zugriff 01. 02. 2018).

Cleve, Jürgen und Uwe Lämmel (2016). Data Mining. 2. Auflage. Berlin und Boston: De Gruyter Studium. ISBN: 978-3-11-045675-2.

Cognitive Computing Consortium (2014). Cognitive Computing Defined. URL: https://cognitivecomputingconsortium.com/resources/cognitive-computing-defined/ (letzter Zugriff 01. 02. 2018).

Dinzler, André (2017). Pressemitteilung „INTER goes digital: Schritt für Schritt zum Abschluss im Internet - Neuer Ansatz in der Kundenkommunikation“. Herausgegeben von INTER Versicherungsgruppe AG. URL: https://www.inter.de/fileadmin/user_upload/inter/dokumente/pdf/Presse/Pressemitteilung_IQMZ_EVA_2017.pdf (letzter Zugriff 09. 01. 2018).

ECMA International (2013). Standard ECMA-404: The JSON Data Interchange Format. URL: https://www.ecma-international.org/publications/standards/Ecma-404.htm (letzter Zugriff 01. 02. 2018).

Elhassouny, Azeddine u. a. (2017). Building cognitive applications with IBM Watson Services: Volume 3 Visual Recognition. IBM Redbooks. Poughkeepsie, NY: IBM Corporation, International Technical Support Organization. ISBN: 978-0-7384-4257-0. URL: https://www.redbooks.ibm.com/Redbooks.nsf/RedbookAbstracts/sg248393.html?Open (letzter Zugriff 01. 02. 2018).

Engelage, Henning (2015). Trusted German Insurance Cloud: Neue Wege in der Kommunikation. Herausgegeben von Gesamtverband der deutschen Versicherungswirtschaft e.V. URL: http://www.gdv.de/2015/03/neue-wege-in-der-kommunikation/ (letzter Zugriff 01. 02. 2018).

Eschholz, Stefanie (2017). »Big Data-Scoring unter dem Einfluss der Datenschutz-Grundverordnung«. In: Datenschutz und Datensicherheit (DuD) 41.3, S. 180–185. DOI: 10.1007/s11623-017-0752-9.

Europäische Union (2016). Verordnung (EU) 2016/679 des europäischen Parlaments und des Rates zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG (Datenschutz-Grundverordnung DSGVO). URL: http://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=CELEX:32016R0679&from=DE (letzter Zugriff 01. 02. 2018).

Evans, Jonathan St B. T. (2008). »Dual-processing accounts of reasoning, judgment, and social cognition«. In: Annual review of psychology 59, S. 255–278. ISSN: 0066-4308. DOI: 10.1146/annurev.psych.59.103006.093629.

Fasel, Daniel und Andreas Meier (2016). »Was versteht man unter Big Data und NoSQL?« In: Big Data. Grundlagen, Systeme und Nutzungspotentiale. Herausgegeben von Daniel Fasel und Andreas Meier. HMD Praxis der Wirtschaftsinformatik. Wiesbaden: Springer Vieweg, S. 3–16. ISBN: 978-3-658-11588-3. DOI: 10.1007/978-3-658-11589-0_1.

Ferrucci, David u. a. (2010). »Building Watson: An Overview of the DeepQA Project«. In: AI Magazine 31, S. 59–79. DOI: 10.1609/aimag.v31i3.2303.

Ferrucci, David u. a. (2013). »Watson: Beyond Jeopardy!« In: Artificial Intelligence 199-200, S. 93–105. ISSN: 00043702. DOI: 10.1016/j.artint.2012.06.009.

Fielding, Roy Thomas (2000). »Architectural Styles and the Design of Network-based Software Architectures«. Dissertation. Irvine: University of California. URL: https://www.ics.uci.edu/~fielding/pubs/dissertation/fielding_dissertation.pdf (letzter Zugriff 01. 02. 2018).

Fischermanns, Guido (2013). Praxishandbuch Prozessmanagement. 11. Auflage. Band 9. ibo-Schriftenreihe. Gießen: Verlag Dr. Götz Schmidt. ISBN: 978-3-921313-89-3.

Fromme, Herbert (2015). Ärger für Watson. Herausgegeben von Süddeutsche Zeitung. URL:https://www.sueddeutsche.de/wirtschaft/kuenstliche-intelligenz-aerger-fuer-watson-1.2772927 (letzter Zugriff 15. 01. 2018).

Gliozzo, Alfio u. a. (2017). Building Cognitive Applications with IBM Watson Services: Volume 1. Getting Started. IBM Redbooks. IBM Corporation, International Technical Support Organization. ISBN: 978-0-7384-4264-8. URL: https://www.redbooks.ibm.com/Redbooks.nsf/RedbookAbstracts/sg248387.html?Open (letzter Zugriff 01. 02. 2018).

Gudivada, Venkat N. (2016). »Cognitive Computing: Concepts, Architectures, Systems, and Applications«. In: Handbook of Statistics. Band 35. Amsterdam: Elsevier, S. 3–38. DOI: 10.1016/bs.host.2016.07.004.

Gudivada, Venkat N. u. a. (2016). »Cognitive Analytics: Going Beyond Big Data Analytics and Machine Learning«. In: Handbook of Statistics. Band 35. Amsterdam: Elsevier, S. 169–205. DOI: 10.1016/bs.host.2016.07.010.

Haller, Sabine (2017). Dienstleistungsmanagement. Wiesbaden: Springer Fachmedien Wiesbaden. ISBN: 978-3-658-16896-4. DOI: 10.1007/978-3-658-16897-1.

Haun, Matthias (2014). Cognitive Computing: Steigerung des systemischen Intelligenzprofils. Springer Science and Business Media. ISBN: 978-3-662-44074-2. DOI: 10.1007/978-3-662-44075-9.

Herzberg, Philipp Yorck und Marcus Roth (2014). Persönlichkeitspsychologie. Wiesbaden: Springer Fachmedien Wiesbaden. ISBN: 978-3-531-17897-4. DOI: 10.1007/978-3-531-93467-9.

Heyer, Gerhard, Uwe Quasthoff und Thomas Wittig (2012). Text Mining: Wissensrohstoff Text. Konzepte, Algorithmen, Ergebnisse. 2. Nachdruck. Herdecke und Witten: W3L-Verlag. ISBN: 978-3-937137-30-8.

Hierzer, Rupert (2017). Prozessoptimierung 4.0. Den digitalen Wandel als Chance nutzen. 1. Auflage. Freiburg, München und Stuttgart: Haufe Gruppe. ISBN: 978-3-648-09519-5.

Hull, Richard und Hamid R. Motahari Nezhad (2016). »Rethinking BPM in a Cognitive World: Transforming How We Learn and Perform Business Processes«. In: Business Process Management 14th International Conference BPM 2016 - Proceedings. Herausgegeben von Marcello La Rosa, Peter Loos und Oscar Pastor. Cham: Springer International Publishing, S. 3–19. ISBN: 978-3-319-45347-7. DOI: 10.1007/978-3-319-45348-4_1.

Hummeltenberg, Wilhelm (2014). Enzyklopädie der Wirtschaftsinformatik: Business Intelligence. URL: http://enzyklopaedie-der-wirtschaftsinformatik.de/lexikon/daten-wissen/Business-Intelligence/index.html (letzter Zugriff 01. 02. 2018).

Hurwitz, Judith, Marcia Kaufman und Adrian Bowles (2015). Cognitive computing and big data analytics. Indianapolis: Wiley. ISBN: 978-1-118-89662-4.

IBM Deutschland GmbH (2014). IBM Cloud Docs: Watson Language Translator. URL: https://console.bluemix.net/docs/services/language-translator/index.html (letzter Zugriff 16. 01. 2018).

IBM Deutschland GmbH (2015a). IBM Cloud Docs: Watson Natural Language Classifier. URL: https://console.bluemix.net/docs/services/natural-language-classifier/natural-language-classifier-overview.html (letzter Zugriff 16. 01. 2018).

IBM Deutschland GmbH (2015b). IBM Cloud Docs: Watson Personality Insights. URL: https://console.bluemix.net/docs/services/personality-insights/index.html (letzter Zugriff 28. 01. 2018).

IBM Deutschland GmbH (2015c). IBM Cloud Docs: Watson Speech to Text. URL: https://console.bluemix.net/docs/services/speech-to-text/index.html (letzter Zugriff 16. 01. 2018).

IBM Deutschland GmbH (2015d). IBM Cloud Docs: Watson Text to Speech. URL: https://console.bluemix.net/docs/services/text-to-speech/index.html (letzter Zugriff 16. 01. 2018).

IBM Deutschland GmbH (2015e). IBM Cloud Docs: Watson Visual Recognition. URL: https://console.bluemix.net/docs/services/visual-recognition/index.html (letzter Zugriff 16. 01. 2018).

IBM Deutschland GmbH (2016a). IBM Cloud Docs: Watson Conversation. URL: https://console.bluemix.net/docs/services/conversation/index.html (letzter Zugriff 16. 01. 2018).

IBM Deutschland GmbH (2016b). IBM Cloud Docs: Watson Knowledge Studio. URL:https://console.bluemix.net/docs/services/watson-knowledge-studio/index.html (letzter Zugriff 23. 01. 2018).

IBM Deutschland GmbH (2016c). IBM Cloud Docs: Watson Tone Analyzer. URL: https://console.bluemix.net/docs/services/tone-analyzer/index.html (letzter Zugriff 28. 01. 2018).

IBM Deutschland GmbH (2016d). IBM: Watson Developer Cloud Security Overview Stand Juli 2016. URL: https://www.ibm.com/watson/assets/pdfs/Watson_Developer_Cloud_Security_Overview_July-2016_1.pdf (letzter Zugriff 01. 02. 2018).

IBM Deutschland GmbH (2017a). IBM Cloud Docs: Natural Language Understanding. URL: https://console.bluemix.net/docs/services/natural-language-understanding/index.html#about (letzter Zugriff 23. 01. 2018).

IBM Deutschland GmbH (2017b). IBM Cloud Docs: Virtual Agent. URL:https://console.bluemix.net/docs/services/virtual-agent/index.html#about (letzter Zugriff 01. 02. 2018).

IBM Deutschland GmbH (2017c). IBM Cloud Docs: Watson Discovery. URL: https://console.bluemix.net/docs/services/discovery/index.html (letzter Zugriff 01. 02. 2018).

IBM Deutschland GmbH (2017d). IBM Cloud Docs: Watson Services. URL: https://console.bluemix.net/docs/services/watson/index.html (letzter Zugriff 29. 01. 2018).

IBM Deutschland GmbH (2017e). IBM Cloud Service Description. URL: https://www-03.ibm.com/software/sla/sladb.nsf/sla/bm-6605-12 (letzter Zugriff 06. 01. 2018).

Internet Engineering Task Force IETF (2014). RFC 7159: The JavaScript Object Notation Data Interchange Format. URL: https://tools.ietf.org/html/rfc7159 (letzter Zugriff 01. 02. 2018).

Kahneman, Daniel (2015). Schnelles Denken, langsames Denken. 17. Auflage. München: Pantheon Verlag. ISBN: 978-3-570-55215-5.

Kelly, John E. und Steve Hamm (2014). Smart machines: IBM's Watson and the era of cognitive computing. New York [u. a.]: Columbia Business School Publishing. ISBN: 978-0-231-16856-4.

Kemper, Hans-Georg, Walid Mehanna und Henning Baars (2010). Business Intelligence - Grundlagen und praktische Anwendungen. Eine Einführung in die IT-basierte Managementunterstützung. 3. Auflage. Wiesbaden: Vieweg + Teubner Verlag. ISBN: 978-3-8348-0719-9. DOI: 10.1007/978-3-8348-9727-5.

Knight, Will und Eva Wolfangel (2017). »Was denkt sich die KI?« In: Technology Review August 2017, S. 31–34.

Kocher, Paul u. a. (2018). »Spectre Attacks: Exploiting Speculative Execution«. In: ArXiv e-prints. arXiv: 1801.01203. URL: https://meltdownattack.com/ (letzter Zugriff 01. 02. 2018).

Kramer, André (2011). »Maschine: 1, Menschheit: 0: Supercomputer gewinnt Quizshow«. In: c't magazin für computertechnik 06/2011, S. 55.

Kroker, Michael (2017). »Dumm gelaufen«. In: Wirtschaftswoche 39/2017, S. 55–57. URL: http://www.wiwo.de/unternehmen/it/watson-ibms-supercomputer-stellt-sich-dumm-an/20325548-all.html (letzter Zugriff 01. 02. 2018).

Lämmel, Uwe und Jürgen Cleve (2008). Künstliche Intelligenz. 3. Auflage. München: Hanser. ISBN: 978-3-446-41398-6.

Lampkin, Valerie u. a. (2017). Building cognitive applications with IBM Watson Services: Volume 5 Language Translator. IBM Redbooks. Poughkeepsie, NY: IBM Corporation, International Technical Support Organization. ISBN: 978-0-7384-4258-7. URL: https://www.redbooks.ibm.com/Redbooks.nsf/RedbookAbstracts/sg248392.html?Open (letzter Zugriff 01. 02. 2018).

Lipp, Moritz u. a. (2018). »Meltdown«. In: ArXiv e-prints. arXiv: 1801.01207. URL: https://meltdownattack.com/ (letzter Zugriff 01. 02. 2018).

Magrabi, Armadeus und Joscha Bach (2013). »Entscheidungsfindung«. In: Handbuch Kognitionswissenschaft. Herausgegeben von Achim Stephan und Sven Walter. Darmstadt: Wissenschaftliche Buchgesellschaft, S. 274–289. ISBN: 978-3-534-26352-3.

Manhaes, Marcelo Mota u. a. (2017). Building cognitive applications with IBM Watson Services: Volume 4 Natural Language Classifier. IBM Redbooks. Poughkeepsie, NY: IBM Corporation, International Technical Support Organization. ISBN: 978-0-7384-4259-4. URL: https://www.redbooks.ibm.com/Redbooks.nsf/RedbookAbstracts/sg248391.html?Open (letzter Zugriff 01. 02. 2018).

Manhart, Klaus (2017). Eine kleine Geschichte der Künstlichen Intelligenz. Herausgegeben von Computerwoche. URL: https://www.computerwoche.de/a/eine-kleine-geschichte-der-kuenstlichen-intelligenz,3330537 (letzter Zugriff 03. 01. 2018).

Mell, Peter und Timothy Grance (2011). The NIST Definition of Cloud Computing, Special Publication 800-145. Herausgegeben von National Institute of Standards and Technology (NIST). URL: http://dx.doi.org/10.6028/NIST.SP.800-145 (letzter Zugriff 01. 02. 2018).

Moore, Susan (2017). Gartner Says AI Technologies Will Be in Almost Every New Software Product by 2020. Herausgegeben von Gartner Inc. URL: https://www.gartner.com/newsroom/id/3763265 (letzter Zugriff 01. 02. 2018).

Neckel, Peter und Bernd Knobloch (2015). Customer Relationship Analytics. Praktische Anwendung des Data Mining im CRM. 2. Auflage. Heidelberg: dpunkt.verlag. ISBN: 978-3-86490-090-7.

Nurseitov, Nurzhan u.a. (2009). Comparison of JSON and XML data interchange formats: A case study. 22nd International Conference on Computer Applications in Industry and Engineering 2009, CAINE 2009. S. 157-162. URL: https://www.cs.montana.edu/izurieta/pubs/caine2009.pdf

Panetta, Kasey (2017). Top Trends in the Gartner Hype Cycle for Emerging Technologies 2017. Herausgegeben von Gartner Inc. URL: https://www.gartner.com/smarterwithgartner/top-trends-in-the-gartner-hype-cycle-for-emerging-technologies-2017/ (letzter Zugriff 07. 01. 2018).

Pfuhl, Markus (2012). Enzyklopädie der Wirtschaftsinformatik: Taxonomien. URL: http://www.enzyklopaedie-der-wirtschaftsinformatik.de/lexikon/daten-wissen/Wissensmanagement/Wissensmodellierung/Wissensreprasentation/Semantisches-Netz/Taxonomien (letzter Zugriff 01. 02. 2018).

Plath, Kai-Uwe, Herausgeber (2016). BDSG/DSGVO. Kommentar zum BDSG und zur DSGVO sowie den Datenschutzbestimmungen des TMG und TKG. 2. Auflage. Köln: Verlag Dr. Otto Schmidt. ISBN: 978-3-504-56074-4.

Posluschny, Peter (2016). Praxishandbuch Prozessmanagement. Kundenorientierung, Modellierung, Optimierung. 2. Auflage. Konstanz, München: UVK Verlagsgesellschaft mbH. ISBN: 978-3-86764-687-1.

Russell, Stuart J. und Peter Norvig (2012). Künstliche Intelligenz: Ein moderner Ansatz. 3. Auflage. Pearson Studium - IT. München: Pearson Deutschland. ISBN: 978-3-86894-098-5.

Santiago, Felipe, Pallavi Singh und Lak Sri (2017). Building cognitive applications with IBM Watson Services: Volume 6 Speech to Text and Text to Speech. IBM Redbooks. Poughkeepsie, NY: IBM Corporation, International Technical Support Organization. ISBN: 978-0-7384-4260-0. URL: https://www.redbooks.ibm.com/Redbooks.nsf/RedbookAbstracts/sg248388.html?Open (letzter Zugriff 01. 02. 2018).

Schallmo, Daniel R. A. und Leo Brecht (2014). Prozessinnovation erfolgreich anwenden. Grundlagen und methodisches Vorgehen. Gabler Verlag. ISBN: 978-3-642-55242-7.

Scheerer, Claudia (2016). Pressemitteilung „Wieder kam keine Reaktion von Ihnen!“ Herausgegeben von Union Krankenversicherung AG. URL: https://www.vkb.de/export/sites/vkb/_resources/pdf/ueber-uns/presse/pressemitteilungen-2016/20161125-Watson-UKV.pdf (letzter Zugriff 15. 01. 2018).

Schmelzer, Hermann J. und Wolfgang Sesselmann (2013). Geschäftsprozessmanagement in der Praxis. Kunden zufrieden stellen - Produktivität steigern - Wert erhöhen. 8. Auflage. München: Carl Hanser Verlag. ISBN: 978-3-446-43460-8.

Schmidt, Ute (2013). »Künstliche-Intelligenz-Forschung«. In: Handbuch Kognitionswissenschaft. Herausgegeben von Achim Stephan und Sven Walter. Stuttgart und Weimar: Verlag J.B. Metzler, S. 44–46. ISBN: 978-3-476-02331-5.

Schröder, Bernhard (2013). »Computerlinguistik«. In: Handbuch Kognitionswissenschaft. Herausgegeben von Achim Stephan und Sven Walter. Stuttgart und Weimar: Verlag J.B. Metzler, S. 71–75. ISBN: 978-3-476-02331-5.

Shahd, Maurice (2016). Neue digitale Technologien vor dem Durchbruch. URL: https://www.bitkom.org/Presse/Presseinformation/Neue-digitale-Technologien-vor-dem-Durchbruch.html (letzter Zugriff 01. 02. 2018).

Simitis, Spiros, Herausgeber (2014). Bundesdatenschutzgesetz. 8. Auflage. Baden-Baden: Nomos Verlagsgesellschaft. ISBN: 978-3-8487-0593-1.

Sosinsky, Barrie (2011). Cloud Computing Bible. Indianapolis: Wiley. ISBN: 978-0-470-90356-8.

Stephan, Achim und Sven Walter, Herausgeber (2013). Handbuch Kognitionswissenschaft. Stuttgart und Weimar: Verlag J.B. Metzler. ISBN: 978-3-476-02331-5.

Studer, Rudi (2016). Enzyklopädie der Wirtschaftsinformatik: Ontologien. URL: http://www.enzyklopaedie-der-wirtschaftsinformatik.de/lexikon/daten-wissen/Wissensmanagement/Wissensmodellierung/Wissensreprasentation/Semantisches-Netz/Ontologien (letzter Zugriff 01. 02. 2018).

Tanenbaum, Andrew S. und Maarten van Steen (2008). Verteilte Systeme. Prinzipien und Paradigmen. 2. Auflage. It Informatik. München: Pearson Studium. ISBN: 978-3-8273-7293-2.

Trinkwalder, Andrea (2016). »Netzgespinste. Die Mathematik neuronaler Netze: einfache Mechanismen, komplexe Konstruktion«. In: c't magazin für computertechnik 06/2016, S. 130–135.

Unland, Rainer (2012). Enzyklopädie der Wirtschaftsinformatik: Wissensrepräsentation. URL: http://www.enzyklopaedie-der-wirtschaftsinformatik.de/lexikon/daten-wissen/Wissensmanagement/Wissensmodellierung/Wissensreprasentation (letzter Zugriff 01. 02. 2018).

Varian, Hal R. (2007). Grundzüge der Mikroökonomik. 7. Auflage. München: Oldenbourg Verlag. ISBN: 978-3486583113.

Vergara, Sebastian u. a. (2017). Building cognitive applications with IBM Watson Services: Volume 7 Natural Language Understanding. IBM Redbooks. Poughkeepsie, NY: IBM Corporation, International Technical Support Organization. ISBN: 978-0-7384-4262-4. URL: https://www.redbooks.ibm.com/Redbooks.nsf/RedbookAbstracts/sg248398.html?Open (letzter Zugriff 01. 02. 2018).

Vossen, Gottfried, Till Haselmann und Thomas Hoeren (2013). Cloud-Computing für Unternehmen. Technische, wirtschaftliche, rechtliche und organisatorische Aspekte. Heidelberg: dpunkt.verlag. ISBN: 978-3-89864-808-0.

Wang, Yingxu, Du Zhang und Witold Kinsner (2010). »Advances in the Fields of Cognitive Informatics and Cognitive Computing«. In: Advances in Cognitive Informatics and Cognitive Computing. Herausgegeben von Yingxu Wang, Du Zhang und Witold Kinsner. Band 323. Studies in Computational Intelligence. Berlin, Heidelberg: Springer Berlin Heidelberg, S. 1–11. ISBN: 978-3-642-16082-0. DOI: 10.1007/978-3-642-16083-7_1.

Waters, Richard (2016). »Artificial intelligence: Can Watson save IBM?« In: Financial Times. URL: https://www.ft.com/content/dced8150-b300-11e5-8358-9a82b43f6b2f (letzter Zugriff 01. 02. 2018).

Weber, Mathias (2015a). Kognitive Maschinen. Meilenstein in der Wissensarbeit. Herausgegeben von BITKOM e.V. URL:https://www.bitkom.org/Bitkom/Publikationen/Kognitive-Maschinen-Meilenstein-in-der-Wissensarbeit.html (letzter Zugriff 01. 02. 2018).

Weber, Mathias (2015b). Rasantes Wachstum für Cognitive Computing. URL: https://www.bitkom.org/Presse/Presseinformation/Rasantes-Wachstum-fuer-Cognitive-Computing.html (letzter Zugriff 01. 02. 2018).

Weber, Mathias (2017a). Weltmarkt für Cognitive Computing vor dem Durchbruch. URL:https://www.bitkom.org/Presse/Presseinformation/Weltmarkt-fuer-Cognitive-Computing-vor-dem-Durchbruch.html (letzter Zugriff 01. 02. 2018).

Weber, Volker (2017b). IBM Connect 2017: Watson über alles. Herausgegeben von Heise-Medien. URL: https://heise.de/-3633407 (letzter Zugriff 01. 02. 2018).

Welter, Patrick (2017). Versicherer ersetzt zahlreiche Mitarbeiter durch künstliche Intelligenz. Herausgegeben von Frankfurter Allgemeine Zeitung. URL: http://www.faz.net/aktuell/wirtschaft/japan-versicherer-ersetzt-mitarbeiter-durch-ki-ibm-watson-14605854.html (letzter Zugriff 15. 01. 2018).

Wybitul, Tim (2016). EU-Datenschutz-Grundverordnung im Unternehmen: Praxisleitfaden. Kommunikation & Recht. Frankfurt: Fachmedien Recht und Wirtschaft in Deutscher Fachverlag GmbH. ISBN: 978-3-8005-1634-6.Zaino, Jennifer (2014). Another Face of Cognitive Computing. URL: http://www.dataversity.net/another-face-cognitive-computing/ (letzter Zugriff 01. 02. 2018).

WDP - Wismarer Diskussionspapiere / Wismar Discussion Papers

Heft 01/2012: Robin Rudolf Sudermann, Arian Middleton, Thomas Frilling: Werteorientierung als relevanter Erfolgsfaktor für Unternehmen im Zeitalter des Societing

Heft 02/2012: Romy Schmidt: Die Wahrnehmung der WinterDestination Tirol der Zielgruppe „junge Leute“ in Mecklenburg-Vorpommern

Heft 03/2012: Roland Zieseniß, Dominik Müller: Performancevergleiche zwischen Genossenschaften und anderen Rechtsformen anhand von Erfolgs-, Liquiditäts- und Wachstumskennzahlen

Heft 04/2012: Sebastian Kähler, Sönke Reise: Potenzialabschätzung der Regionalflughäfen Mecklenburg-Vorpommerns

Heft 05/2012: Barbara Bojack: Zum möglichen Zusammenhang von Psychotrauma und Operationsindikation bei Prostatahyperplasie

Heft 06/2012: Hans-Eggert Reimers: Early warning indicator model of financial developments using an ordered logit

Heft 07/2012: Günther Ringle: Werte der Genossenschaftsunternehmen – “Kultureller Kern” und neue Wertevorstellungen 94

Heft 08/2012: Harald Mumm: Optimale Lösungen von Tourenoptimierungsproblemen mit geteilter Belieferung, Zeitfenstern, Servicezeiten und vier LKW-Typen

Heft 01/2013: Dieter Gerdesmeier, Hans-Eggert Reimers, Barbara Roffia: Testing for the existence of a bubble in the stock market

Heft 02/2013: Angje Bernier, Katharina Kahrs, Anne-Sophie Woll:

Landesbaupresi für ALLE? 1. Fortsetzung – Analyse der Barrierefreiheit von Objekten des Landesbaupreises Mecklenburg-Vorpommern 2010/2012

Heft 03/2013: Günther Ringle: Auf der Suche nach der „richtigen“ Mitgliederförderung

Heft 04/2013: Frederik Schirdewahn: Analyse der Effizienz einzelner Maßnahmen zur Reduzierung des CO2 -Ausstoßes in der Transportlogistik

Heft 05/2013: Hans-Eggert Reimers: Remarks on the euro crisis

Heft 01/2014: Antje Bernier (Hrsg.): Na, altes Haus? – Stadt und Umland im Wandel. Planungs- und Entwick-lungsinstrumente mit demografischer Chance, Konferenz der Hochschule Wismar am 14. Okt. 2013 in Schwerin

Heft 02/2014: Stefan Voll/Daniel Alt: „Das große Ziel immer im Auge behalten“ Sportimmanente Indikatoren des Trainerstils von Jürgen Klopp – Transfermöglichkeiten für Führungskräfte in Genossenschaftsbanken

Heft 03/2014: Günther Ringle: Genossenschaftliche Solidarität auf dem Prüfstand

Heft 04/2014: Barbara Bojack: Alkoholmissbrauch, Alkoholabhängigkeit

Heft 01/2015: Dieter Gerdesmeier/ Hans-Eggert Reimers/ Barbara Roffia: Consumer and asset prices: some recent evidence

Heft 02/2015: Katrin Schmallowsky: Unternehmensbewertung mit Monte-Carlo-Simulationen

Heft 03/2015: Jan Bublitz/ Uwe Lämmel: Semantische Wiki und TopicMap-Visualisierung

Heft 04/2015: Herbert Müller: Der II. Hauptsatz der Thermody-na-

mik, die Philosophie und die gesellschaft-liche Praxis – eine Neubetrachtung

Heft 05/2015: Friederike Diaby-Pentzlin: Auslandsinvestitionsrecht und Entwicklungspolitik: Derzeitiges bloßes internationales Investitionsschutzrecht vertieft Armut

Heft 01/2016: Sonderheft: Jürgen Cleve, Erhard Alde (Hrsg.) WIWITA 2016. 10. Wismarer Wirtschaftsinformatiktage 9./10. Juni 2016. Proceedings

Heft 02/2016: Günther Ringle: Die soziale Funktion von Genossenschaften im Wandel

Heft 01/2017: Benjamin Reimers: Momentumeffekt: Eine empirische Analyse der DAXsector Indizes des deutschen Prime Standards

Heft 02/2017: Florian Knebel, Uwe Lämmel: Einsatz von Wiki-Systemen im Wissensmanagement

Heft 03/2017: Harald Mumm: Atlas optimaler Touren

Heft 01/2018: Günther Ringle: Verfremdung der Genossenschaften im Nationalsozialismus – Gemeinnutzvorrang und Führerprinzip

Heft 02/2018: Sonderheft: Jürgen Cleve, Erhard Alde, Matthias Wißotzki (Hrsg.) WIWITA 2018. 11. Wismarer Wirtschaftsinformatiktage 7. Juni 2018. Proceedings

ISBN 978-3-942100-60-1