

# **Automatische Dokumentenverarbeitung durch Künstliche Intelligenz**

**Universität Leipzig  
Fakultät Mathematik und Informatik**

**Interdisziplinäre Aspekte des digitalen Wandels  
Wintersemester 2017/18**

Arne Thamm  
Matrikelnummer: 3740451

17. April 2018

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
<b>2</b>	<b>Was ist Künstliche Intelligenz</b>	<b>2</b>
2.1	Intelligenz . . . . .	2
2.2	Künstliche Intelligenz - KI . . . . .	2
2.3	Künstliche Neuronale Netze - KNN . . . . .	3
<b>3</b>	<b>Technischer Hintergrund</b>	<b>4</b>
<b>4</b>	<b>Anwendungsbeispiele</b>	<b>6</b>
4.1	Weitere Anwendungsbeispiele . . . . .	6
4.1.1	Enterprise Search . . . . .	6
4.1.2	Swiss Post Solution . . . . .	6
4.1.3	Wikidata . . . . .	7
<b>5</b>	<b>Fazit</b>	<b>8</b>
	<b>Literaturverzeichnis</b>	<b>9</b>

# 1 Einleitung

In dieser Ausarbeitung wird die automatische Dokumentenverarbeitung mit Hilfe der Künstlichen Intelligenz etwas näher gebracht. Die heutige Technik entwickelt sich immer schneller und immer weiter und viele Tätigkeiten im Berufsleben sind sehr eintönig und gleichen einer Fließbandarbeit. Für solche Tätigkeiten könnte eine solche Künstliche Intelligenz eingesetzt werden.

Einige wichtige Fragen bezüglich einer Software mit automatischer Dokumentenverarbeitung durch Künstliche Intelligenz sind, wie und wo kann eine solche Software einen großen Nutzen für den Menschen bringen und was kann noch erwartet werden?

In einer Kanzlei beispielsweise, in der die jungen neuen Angestellten Berge über Berge von Dokumenten, Anträge, Notizen, Gesetzesbücher und vielen weiteren Papierkram über Tage, ja sogar Monate zusammentragen, um diese für einen spezifischen Fall aufbereiten und zusammenfassen müssen. Für diese Art von Software ist es ein leichtes, diese mühselige Arbeit zu verrichten und dies sogar in bedeutend weniger Zeit. [5]

Was genau man unter einer Künstlichen Intelligenz verstehen kann, wird im weiteren Verlauf erläutert und anhand des Bereits genannten Kanzlei-Beispiels und ein paar Technologien erklärt. Doppellungen sind dabei nicht auszuschließen.

Anschließend wird auf den technischen Hintergrund ein wenig näher eingegangen und kurz die Technologien einer solchen Software, für die Kanzlei beispielsweise, angeschnitten. Die wichtigsten Technologien werden danach aufgeführt und ein weiteres Mal erfasst.

Anwendungsbeispiele folgen für eine automatische Dokumentenverarbeitung durch Künstliche Intelligenz. Dafür wird ein kleiner Einblick in den Postbereich, sowie auf den etwas allgemeineren Bereich, mit Beispiel einer Versicherung, und in Wikipedia gewährt. Nachfolgend wird diese Ausarbeitung mit einem Fazit abgeschlossen.

## 2 Was ist Künstliche Intelligenz

### 2.1 Intelligenz

Zunächst muss kurz erläutert werden, was allgemein unter Intelligenz verstanden wird. Unter Intelligenz versteht man einen Sammelbegriff, welcher als Ausdruck für kognitiven Fähigkeiten des Menschen gilt. Hierzu gehört das Lernen, welches unter anderem das Erkennen von Mustern und Strukturen beinhaltet, und die Kreativität, welche gewisse bereits vorhandene Muster bzw. Strukturen voraussetzt nach denen gehandelt werden kann. Muster und Strukturen sind logische Denkweisen sowie Sprachmuster oder eben eine vorhandene und erlernte Textstruktur bzw. Satzstruktur.

### 2.2 Künstliche Intelligenz - KI

[10] Als KI wird ein technischer Nachbau der menschlichen Intelligenz verstanden, um beispielsweise, wie der Titel bereits ausdrückt, eine automatische Dokumentenverarbeitung zu erstellen. Eine einheitliche Definition für Intelligenz kann hier nicht vergeben werden, da es bereits an einer Solchen mangelt.

In der Künstlichen Intelligenz wird in schwache und starke KI unterschieden.

Schwache KI (keine Selbstständigkeit):

- Lösung komplexer Probleme dank einfacher Algorithmen
- Lerneffekte können erzielt werden, jedoch in einem vorgegebenen Rahmen

Starker KI (Selbstständig, unabhängiges System):

- Verfügt über ein eigenes Selbstbewusstsein (Metakognition)
- Selbstständiges Lernen und Schlussfolgern
- Uneinig darüber ob Gefühle für eine KI nötig sind
- Es ist umstritten, ob dies überhaupt möglich ist

Ein Beispiel hierfür ist die Künstliche Intelligenz von IBM, Ross. Diese wird bereits in der Kanzlei in San Francisco, „Baker & Hostetler“ eingesetzt. Ross wird auch als ein Roboter-Anwalt bezeichnet, denn dieser kann mit der Hilfe von Semantic web, Gesetzbüchern, Notizen und Anträgen alle passenden und relevanten Unterlagen zu einem Fall zusammenzutragen. Zudem lernt Ross mit jedem bearbeiteten Fall ein wenig dazu und kann die Antworten verfeinern. [1]

## 2.3 Künstliche Neuronale Netze - KNN

[11] Künstliche Neuronale Netze oder auch künstliche neuronale Netzwerke ist ein Teilgebiet der Künstlichen Intelligenz. Hierbei werden Themen aus den Neurowissenschaften und der Informatik vereint. Der Mensch ist abermals das biologische Vorbild, genauer gesagt die Vernetzung der Neuronen im Gehirn und im Rückenmark. Dabei ist zu erwähnen, dass es bei Künstlichen Neuronalen Netzen viel mehr um eine Abstraktion, also um eine Modellbildung, der Informationsverarbeitung, als um eine Nachbildung des menschlichen Gehirns.

Künstliche Neuronale Netze sind ein starkes Instrument, um zu versuchen die Denkprozesse eines Menschen nachzubilden. Ein sehr großer Nachteil bei dieser Methodik liegt eindeutig in den Ressourcen. Um ein starkes künstliches Neuronales Netz zu kreieren benötigt es nicht nur eine gute Struktur, sondern auch eine auf das Netz abgestimmte Chiparchitektur und eine große Menge von Prozessorkraft und Speicher.

Es existieren bereits eine Vielzahl an Algorithmen und Paradigmen, in denen Künstliche Neuronale Netze zum Einsatz kommen. Hier einige Beispiele:

- Deep Learning
- Machine Learning
- Data Mining

### 3 Technischer Hintergrund

Für die automatische Dokumentenverarbeitung durch die Künstliche Intelligenz kann eine Vielzahl von Methoden und Hilfsmitteln verwendet werden, um an das gewünschte Ziel zu gelangen. Als Beispiel wird ein weiteres Mal IBM Ross zu Rate gezogen.

Hierbei ist zu erwähnen, dass IBM Ross von dem bereits bestehenden System Watson, ebenfalls von IBM, abstammt. Dies bildet die Grundlage von Ross.

Watson nutzt hauptsächlich „Cognitive Computing“ [6] um durch Input, sei es Lokal oder auch durch das Semantic Web, in einem bestimmten Themenbereich zu lernen. In der Regel werden als Cognitive Computing [4], Technologien der Künstlichen Intelligenz, wie „Deep Learning“ und „Data Mining“, bezeichnet. Diese bestehen im Allgemeinen aus Künstlichen Neuronalen Netzen in Verbindung mit anderen Algorithmen.

Einige wichtige Voraussetzungen für das Cognitive Computing ist unter anderem die die Fähigkeit, aus den entstandenen Erfahrungen selbstständig lernen zu können und diese gleichzeitig reflektieren und hinterfragen zu können. Das heißt, dass bereits Gelerntes, wie auch bei einem Menschen, erneut genutzt werden kann um diese Erfahrungen zu bestätigen oder zu korrigieren. Jedoch sind weitere Voraussetzungen von Nöten, damit das System annähernd ähnlich wie ein menschliches Gehirn arbeiten kann.

Watson wie auch Ross verfügt außerdem über eine Spracheingabe und -ausgabe. Die Systeme erkennen die natürlich Sprache und können dadurch nach Dokumenten oder auch Artikeln suchen, diese analysieren und zusammenfassen und diese ausgeben.

Mit Cognitive Computing wird also versucht, menschliche Denkprozesse in einem Computermodell zu simulieren. Hinzukommend kann Ross durch das Semantic Web [2] sich, wie bereits erwähnt, Dokumente von ähnlichen Fällen oder gar gleichen Fällen aus dem „World Wide Web“ filtern und somit beispielsweise, wenn gewünscht, eine vorläufige Strategie zusammenbauen. Bezüglich des Semantic Web's ein kleines Beispiel, inwiefern die Informationen abgespeichert sein könnten.

Angenommen, ein Autodiebstahl mit schwerer Körperverletzung ist ein solcher Fall, Ross würde die lokalen Dokumente scannen und auf Key-Wörter wie „Autodiebstahl“ und „schwere Körperverletzung“ durchsuchen. Da im Semantic Web die Informationen mit eindeutigen Begriffen bzw. Informationen versehen sind, kann Ross, als ein Computersystem, diese mit Leichtigkeit „verstehen“, filtern und auch hierbei alle Dokumente, Gesetzesbuchauszüge und veröffentlichte Fälle nach „Autodiebstahl“ und „schwerer Körperverletzung“ durchsuchen und darauf zugreifen.

So könnten die Informationen dargestellt sein:

- Urteil im einem Fall mit Autodiebstahl<Tatbestand> und schwerer<Intensität> Körperverletzung<Tatbestand> in Leipzig<Stadt>, 5<Jahre> Jahre Haft<Maßnahme>
- Freispruch<Maßnahme> im Fall des Autodiebstahls<Tatbestand>

Durch diese Annotationen einer Auszeichnungssprache, meist XML, werden die vorhandenen Informationen mit Metainformationen, wie z.B. eben mit Tatbestand, versehen.

Somit kann gesagt werden das folgende Technologien zum Einsatz kommen können:

- Textverarbeitungsalgorithmen - Dokumente werden eingelesen und in Blöcke und Kategorien aufgeteilt
- Sprachverarbeitung - Wie bei Watson und Ross kann durch Algorithmen die natürliche Sprache erkannt und verarbeitet werden
- Cognitive Computing - Hierbei werden Algorithmen zur weiteren Kategorisierung von Informationen eingesetzt welche dann mittels maschinellen Lernens weiter verarbeitet werden zudem auch Lerneffekte erzielt werden können.
- Semantic Web - Informationen werden mit eindeutigen Begriffen mit einer vorher festgelegten Annotation abgespeichert, damit sind diese sowohl für den Menschen als auch für den Computer lesbar und verarbeitbar.

## 4 Anwendungsbeispiele

Wie bereits im vorherigen Kapitel ausgeführt ist IBM Ross ein System, welches automatisch Gesetzes Dokumente verarbeiten, analysieren und aufbereiten und sogar davon lernen kann. IBM Watson hingegen ist ein eher allgemeineres System.

### 4.1 Weitere Anwendungsbeispiele

IBM Ross ist nicht das einzige System welches eine automatische Dokumentenverarbeitung vornehmen kann.

#### 4.1.1 Enterprise Search

[7] So existiert auch „Enterprise Search“, welches für die Verarbeitung von Eingangspost oder auch elektronischen Dokumenten wie Fax und E-Mails inklusive Anhängen genutzt wird. Hierbei wird eine semantische Analyse der eingegangenen Post vorgenommen welche die Dokumente in Blöcke und Kategorien einteilt. Deep Learning Algorithmen erkennen dann in etwa automatisch den unterschied zwischen einem Antrag oder einer Bestellung. Auch dieses System ist Lernfähig und wird dadurch von Mal zu Mal besser.

Enterprise Search kann durch eine fortschrittliche Lösung für eine Erfolgsquote von 85 bis 95 Prozent sorgen. Außerdem spart dieses System dem Unternehmen viel Geld und sowohl auch dem Kunden viel Zeit.

#### 4.1.2 Swiss Post Solution

[9] Swiss Post Solution ist größter Anbieter von Lösungen für standardisierte Geschäftsprozesse in Deutschland und Zweitgrößter in Europa. Auch hier wird automatische Dokumentenverarbeitung betrieben. Ihre Lösungen digitalisieren eingegangene Dokumente und sind sogar in der Lage, unstrukturierte Informationen einzulesen, zu verarbeiten und somit zu ordnen und zu strukturieren. Diese werden dann automatisch in die richtigen Geschäftsprozesse eingespeist und können weiter verarbeitet werden.

Die Künstliche Intelligenz sorgt hierbei dafür, dass unstrukturierte Daten aus „frei formatierten“ Briefen oder E-Mails genutzt werden können. Falls der Fall eintreten sollte, dass die Software eingelesene Daten nicht eindeutig zuordnen kann, sind wieder Mitarbeiter gefragt. Diese Mitarbeiter bekommen dadurch nur noch die Fälle zugeteilt, in denen die Intelligenz eines Menschen wirklich von Nöten ist und den Beruf tatsächlich bereichert. Die Software lernt selbst bei diesen Fällen immer weiter dazu und analysiert auch die Eingaben und Bewertungen des Mitarbeiters.



Hierzu ein kleines Beispiel, ein Kunde lässt seiner Versicherung eine Nachricht zukommen, in dieser möchte er den Rückkaufswert seiner Kapitallebensversicherung erfragen. In der Nachricht werden von diesem Kunden keine einschlägigen Begriffe verwendet, somit kann ein einfacher Algorithmus, wie eine „Keyword-Suche“, nicht weiterhelfen.

Hier kommt die Künstliche Intelligenz ins Spiel. Diese erkennt sofort worum es sich handelt, dank ausreichendem Training, und beginnt die Versicherungsnummer und die Personendaten mit der Versicherungsdatenbank abzugleichen. Anschließend berechnet diese den aktuellen Rückkaufswert. Da die Daten verbindlich und vertraulich sind, wird ein entsprechender Mitarbeiter benachrichtigt, welcher das Antwortschreiben und den Kundenbrief nun nur noch mal kontrollieren und freigeben muss.

### **4.1.3 Wikidata**

[8] [3] Sogar für Wikipedia ist ein automatisches Dokumentenverarbeitungssystem zu finden. Die Universität von Southampton entwickelt eine KI - hier ein Künstliches Neuronales Netz - welches Wikipedia-Artikel ausliest diese klassifiziert und von ihnen lernt, wie man einen Artikel verfasst, kurz ein menschlicher Autor wird nachgeahmt.

Allerdings soll diese Software vorerst dazu genutzt werden, Artikel, welche zum Beispiel nicht in der Sprache des jeweiligen Benutzers existieren jedoch in anderen Ausgaben sehr wohl verfügbar sind, in einen tabellarischen Informationsstand gebracht werden, damit der Nutzer die relevanten Informationen erhält.

## 5 Fazit

Über die automatische Dokumentenverarbeitung durch Künstliche Intelligenz wurde in den vorherigen Kapiteln nun einiges herangetragen und auch erklärt was für Aufgaben übernommen werden und welchen Mehrwert es für sowohl Unternehmen als auch Mitarbeiter haben kann.

Wie schon genannt ist der Einsatzbereich sehr vielfältig, dieser kann eine Kanzlei sein, sogar Wikipedia lässt es sich nicht nehmen ein solches System zu nutzen, um damit die Nutzung zu verbessern und auszuweiten. Selbst Versicherungen und Postunternehmen können eine solche Software nutzen.

Durch die zusätzliche Nutzung des Semantic Web ist einem solchen System fast keine Grenze gesetzt, um sich systematisch in eine vorgegebene Richtung gewisse Kniffe und Paradigmen anzulernen, um so dem Menschen viel Arbeit zu ersparen und abzunehmen.

Das „wie“ wurde hierbei nicht ganz ausgelassen. Es wird derzeit rein unterstützend verwendet, um den Mitarbeitern die Arbeit zu erleichtern oder auch interessanter zu gestalten. Selbstverständlich sparen die Unternehmen dadurch Geld und Zeit ein. Leider kommt es mit Sicherheit auch vor das durch ein solches System Mitarbeiter ihre Arbeit verlieren. Daraus folgt, dass die Software definitiv einen sehr großen Nutzen für Unternehmen und einen im Vergleich eher geringen Nutzen für die Mitarbeiter aufweist.

Es ist anzunehmen, dass die Dokumentenverarbeitung durch Künstliche Intelligenz weitere Fortschritte machen wird und vielleicht sogar ganz die Arbeit von Menschen übernehmen könnte. In naher Zukunft bleibt jedoch die Frage, wie weit kann es noch gehen und welche Bereiche schließen sich allgemein auch dem Strom der KI an?

Allgemein kann gesagt werden, dass eine Software wie IBM Ross, im Hinblick auf eine Kanzlei, eine sehr große Bereicherung ist und auch die Fehlerquoten deutlich reduzieren wird und für viele Arbeitergruppen ein Fluch oder Segen sein wird.

## Literaturverzeichnis

- [1] DAILYMAIL.COM, Stacy Liberatore F.: *IBM ROSS und IBM Watson*. <http://www.dailymail.co.uk/sciencetech/article-3589795/Your-AI-lawyer-IBM-s-ROSS-world-s-artificially-intelligent-attorney.html>. Mai 2016
- [2] ISRAEL, Marcus: *Ontologien in der Philosophie und Informatik*. <http://www.dorfwiki.org/wiki.cgi?HansGertGraebe/SeminarWissen/2012-10-30>. Oktober 2012
- [3] KLEINZ, Torsten: *KI WikiData*. <https://algorithmenethik.de/2017/11/10/wolgorithmen-wikipedia-autoren-ersetzen-ergaenzen-und-motivieren/>. November 2017
- [4] LITZEL, Tutanch / N.: *Was ist Cognitive Computing?* <https://www.bigdata-insider.de/was-ist-cognitive-computing-a-641356/>. September 2017
- [5] POSTINETT, Axel: *KI Anwälte*. <http://www.handelsblatt.com/unternehmen/beruf-und-buero/buero-special/kuenstliche-intelligenz-die-robo-anwaelte-kommen/13601888.html>. Mai 2016
- [6] ROUSE, Margaret: *Cognitive Computing*. <https://www.searchenterprisesoftware.de/definition/Cognitive-Computing>. Februar 2017
- [7] STADLER, Kathrin: *Enterprise Search - Business-Anwendungen-KI*. <https://www.mindbreeze.com/de/blog/sechs-business-anwendungen-der-kuenstlichen-intelligenz.html>. Juni 2017
- [8] TIEDEMANN, Michaela: *Die 10 Ki-Trends*. <https://www.alexanderthamm.com/de/artikel/die-10-ki-trends-fuer-2018-teil-1-einzelanwendungen>. 2017
- [9] TRENDREPORT: *Trend der KI*. <https://trendreport.de/ki-fuer-das-dokumentenmanagement>. 2016
- [10] WAGNER, Jacob: *Von Mustern zur Künstlichen Intelligenz*. <http://www.dorfwiki.org/wiki.cgi?HansGertGraebe/SeminarWissen/2015-12-08>. 2015
- [11] WIKIPEDIA: *Künstliche Neuronale Netze*. [https://de.wikipedia.org/wiki/K%C3%BCnstliches\\_neuronales\\_Netz](https://de.wikipedia.org/wiki/K%C3%BCnstliches_neuronales_Netz). März 2018