

Entwurf und Implementierung einer Werkzeugunterstützung zur sprachlichen Analyse und automatisierten Transformation von Projektlastenheften im Kontext der Automobilindustrie

An der Fachhochschule Dortmund

im Fachbereich Informatik

Studiengang Informatik

Vertiefung Praktische Informatik

erstellte Thesis

zur Erlangung des akademischen Grades

Bachelor of Science

B. Sc.

von Aaron Schul,

geboren am 24.06.1997

und Felix Ritter

geboren am 31.08.1997

Betreuung durch:

Prof. Dr. Sebastian Bab und Prof. Dr. Steffen Helke

Dortmund, 28.02.2019

Zusammenfassung

Ein weitläufiges Problem in der Autoindustrie ist die effiziente Verarbeitung von Pflichtenheften. Einer der Gründe dafür ist, dass viele verschiedene Bereiche bei der Entwicklung der Fahrzeuge und sogar kleinster Einzelteile beteiligt sind. So müssen zu Beginn einer Produktentwicklung etwa Betriebswirtschaftler, Designer, Techniker und Ingenieure zusammen ein Dokument entwerfen, das die Produktmerkmale widerspiegelt. Darin müssen die Anforderungen an das gewünschte Produkt so genau beschrieben sein, sodass es anhand dieses Dokuments entwickelt werden kann. Als Experten ihrer jeweiligen Domäne weiß dabei jeder genau, was dazu nötig ist. Sobald jedoch Auswirkungen über die eigene Domäne hinausgehen, kann es schnell passieren, dass Widersprüche oder Abhängigkeiten entstehen. Diese werden später schnell übersehen, da solche Zuständigkeiten nicht eindeutig geklärt sind. Das Resultat ist dann ein Fehler in der Entwicklung, welcher von zeitlicher Verzögerung über zusätzliche Kosten bis hin zum Abbruch des Projekts führen kann. Methoden aus dem Natural Language Processing (NLP) können dabei helfen, die Überprüfung der Pflichtenhefte auf Konsistenz stark zu vereinfachen oder sogar teilweise zu automatisieren. Dafür kann beispielsweise der Text in den Lastenheften analysiert und dessen Inhalt innerhalb einer Wissensbasis, sog. Ontologie, abgespeichert werden. In dieser Arbeit wird daher die Entwicklung zweier NLP-basierter Werkzeuge zur Vereinfachung bzw. Lösung dieser Probleme vorgestellt werden. Zum einen ein *Requirements-to-Boilerplate-Converter*-Werkzeug (R2BC), welches dem Nutzer helfen soll, das Pflichtenheft eines Auftraggebers in die betriebsinternen Richtlinien und Standards zu überführen. Zum anderen wird der *Delta-Analyser* vorgestellt, welcher auf dem R2BC aufbaut, indem er automatisch zwei homogene Lastenheft vergleicht und dadurch im Kontext des gesamten Pflichtenhefts Widersprüche und Abhängigkeiten herausstellt. Ziel ist es dabei, jeweils einen Prototypen als „*proof of concept*“, also als Machbarkeitsstudie, zu implementieren. Nach dieser Vorstellung wird zusätzlich das weitere Potential der Programme erläutert, welches sich bei der Entwicklung der Prototypen gezeigt hat. Diese bietet Vorschläge zur weiteren bzw. vollständigen Implementierung der Programme.

Abstract

A widespread problem in the auto industry is the efficient processing of specifications. One of the reasons for this is that many different areas are involved in the development of vehicles and even the smallest of parts. For example, at the beginning of product development, managers, designers, technicians, and engineers all need to design a single document together. The requirements for the desired product must be described in such detail that it can be developed using this document. As experts in their domain, everyone knows exactly what is needed. However, as soon as effects go beyond the own domain, it can quickly happen that contradictions or dependencies arise. These are quickly overlooked later, as such responsibilities are not clearly clarified. The result is then a mistake in the development, which can lead from time delay over additional costs up to the demolition of the project. Methods from Natural Language Processing (NLP) can help to simplify or even partially automate the review of functional specifications for consistency. For example, the text in the specifications can be analyzed and its contents stored within a knowledge base, so-called ontology. In this work, therefore, the development of two NLP-based tools to simplify or solve these problems will be presented. On the one hand, there is a program called Requirements-to-Boilerplate Converter (R2BC), which should help the user to transfer the specifications of a client into the company's internal guidelines and standards. On the other hand, the so-called delta analyzer is shown, which is based on the R2BC, in that it automatically compares two homogeneous specification sheets and thereby highlights contradictions and dependencies in the context of the entire specification. The goal here is to implement a prototype as a proof of concept of the program concepts from ZIC19. Due to the limited resources of this work, a list with further potential of the programs, which has been shown during the development of the prototypes, is additionally explained. This offers suggestions for further or complete implementation of the programs.

Inhaltsverzeichnis

1	Einführung	7
1.1	Motivation	7
1.2	Unternehmenskontext	10
1.2.1	Einsatz bei HELLA	10
1.2.2	Anforderungsmanagement im Unternehmen	10
1.3	Hypothese	11
1.4	Methodik	12
1.5	Aufbau der Arbeit	12
1.6	Autorenverzeichnis	13
2	Stand der Technik	14
2.1	Rückblick auf die PA	14
2.1.1	Grundlagen zu Natural Language Processing	14
2.1.2	Umsetzung der Verarbeitungsschritte natürlicher Sprache	16
2.1.3	Syntaktische Analyse von Wörtern und Sätzen	19
2.2	Ontologien	21
2.3	Verwandte Arbeiten	21
3	Betriebliches Umfeld - Hella Use-Case	22
3.1	Re-Prozesse bei Hella und allgemein in Firmen	24
3.2	Betriebliche Anforderungen	24
3.3	Ansatz und Konzept unserer Werkzeuge	24
4	R2B-Converter	25
4.1	Architektur Klassen und Verteilung der Ressourcen	25
4.1.1	Umsetzung der NLP-Architektur in unserem Werkzeug	25

4.2	Implementierung (bisschen Code, GUI, Listenarchitektur, Work-flow für User	26
4.3	mögliche Erweiterungen	26
4.4	Test	26
4.4.1	Methodik	26
4.4.2	Durchführung	26
4.4.3	Ergebnisse	26
5	Delta-Analyse	27
5.1	Architektur	27
5.2	Implementierung	27
5.3	mögliche Erweiterungen	27
5.4	Test	27
5.4.1	Methodik	27
5.4.2	Durchführung	27
5.4.3	Ergebnisse	27
6	Evaluation	28
6.1	Auswertung der Testresultate	28
6.2	Mehrwert?	28
6.3	Ziel erreicht? Hypothese reviewen und schwafeln	28
7	Fazit	29
7.1	Zusammenfassung	29
7.2	Ausblick	29

Tabellenverzeichnis

Abbildungsverzeichnis

2.1	Nicht-deterministischer Zustandswandler der able-Regel [?]	. . . 20
-----	--	----------

Abkürzungsverzeichnis

R2BC	Requirements-To-Boilerplate-Converter
DA	Delta-Analyser
CRS	Customer Requirement Specification
OEM	Original Equipment Manufacturer
E-AE	Electronics-Advanced Engineering
PMT	Projects, Methods and Tools

Danksagung

An dieser Stelle möchten wir uns bei allen Personen bedanken, die uns bei unserer Bachelorarbeit und während des Studiums begleitet und unterstützt haben.

Der Dank gebührt unseren Familien und Freunden aus der Heimat sowie all jenen tollen Menschen, die wir durch das Studium an der FH Dortmund erst kennenlernen durften.

Wir danken auch allen Kollegen aus der Abteilung E-AE bei der Firma Hella in Lippstadt für ihr Interesse und die großartige Zusammenarbeit. Die einzigartigen Erfahrungen, die wir während unseres Praktikums und später während der Bachelorarbeit in der Firma sammeln konnten, haben uns wirklich motiviert. Besonderer Dank geht dabei an Konstantin Zichler, der weit mehr als nur Betreuer in der Firma, sondern auch Mentor und Ideengeber unserer Arbeit war.

Schlussendlich danken wir unseren Betreuern Prof. Dr. Sebastian Bab und Prof. Dr. Steffen Helke für die zahlreichen Tipps und Anregungen während der Bachelorarbeit.

Kapitel 1

Einführung

1.1 Motivation und thematische Grundlagen

In einer Vielzahl von Firmen stellt die Erhebung von Anforderungen den ersten Schritt bei der Entwicklung von neuen Produkten dar. Diese Anforderungen manifestieren sich später in der zu entwickelnden Hard- und Software. Spätere Produktmerkmale müssen dabei weit vor der tatsächlichen Entwicklung berücksichtigt werden und fließen in das Anforderungsdokument ein. Die Erhebung und Identifikation dieser Anforderungen wird als *Requirements-Engineering* bezeichnet [BAL10].

Dieser Prozess wird dabei sowohl bei kompletten Neuentwicklungen, als auch bei der Adaption eines bereits vorhandenen Produktes auf neue Projekte durchlaufen. Zur Rolle von Requirements-Engineering in Projekten siehe etwa [MW02] und [HL01].

Dem Betrieb liegen anschließend die Aufgaben und Ziele für die Entwicklung als Dokumente vor, auf deren Basis die Produktentwicklung beginnen kann. Die Entwicklung unterliegt dabei bestimmten Kriterien und Faktoren, die den unternehmerischen Erfolg beeinflussen. Neben betriebswirtschaftlichen Einflüssen wie der Einordnung des Produktes in der Wertschöpfungskette sind es dabei besonders technische Anforderungen an das Produkt, die definiert und während der Produktentwicklung eingehalten werden müssen.

Anforderungen im betrieblichen Entwicklungsprozess

Das Resultat wird als sogenanntes Pflichtenheft, in Unternehmen häufig als CRS (*Customer Requirement Specification*) bezeichnet, mithilfe dessen nun

die Entwicklung eines Produktes begonnen werden kann [DGE19]. Ein später hergestelltes Produkt sollte möglichst allen beschriebenen Anforderungen entsprechen. Da Betriebe jedoch nur selten alle einzelnen Komponenten des späteren Produktes selbst herstellen können, sind sie auf Zulieferer angewiesen, die einzelne Komponenten entwickeln und produzieren. Dieser Hersteller fertiger Komponenten aus Einzelteilen wird als *OEM* (*Original Equipment Manufacturer*).

In diesem Fall wird dem Zulieferer ein Pflichtenheft übergeben, das allen Anforderungen des Kunden genügt. Das bedeutet jedoch, dass auch die Produktentwicklung anhand des Pflichtenheftes an den Zulieferer ausgelagert wird. Dem Zulieferer werden damit Möglichkeiten geboten und Verantwortung übertragen, das Produkt weiter zu definieren. Einerseits können etwa interne Richtlinien und Erfahrungen auf dem Gebiet der Produktentwicklung aus früheren Projekten integriert werden, andererseits aber auch weitere Optimierungen sowie Merkmale, die noch nicht im Pflichtenheft dokumentiert sind. Somit agiert der Zulieferer nicht nur als produzierende Fabrik, sondern nimmt am Entwicklungsprozess teil.

Die Anforderungen an ein Produkt, etwa technische Rahmenbedingungen und Qualitätsanforderungen, setzen sich dabei aus einer Auflistung von funktionalen und nicht-funktionalen Anforderungen zusammen [BAL10]. Häufig enthält das Pflichtenheft auch weitere Informationen und abseits von Text Abbildungen, die einzelne Anforderungen ergänzen.

Umstände der Anforderungsdefinition

Die Anforderungen werden dabei von vielen verschiedenen Domänenexperten beim OEM formuliert und in das Pflichtenheft für die Entwicklung eingetragen. Verschiedenste Akteure aus einem Unternehmen sind dabei an der Festlegung der Anforderungen an ein Entwicklungsprojekt bzw. Produkt beteiligt. Beteiligte sind etwa Produktdesigner, Ingenieure und Systemtechniker, die an verschiedenen Stellen im Lastenheft Anforderungen an eine Komponente festlegen und an entsprechenden Stellen vermerken. Diese Beteiligten sind in der Regel auf ihren Bereich spezialisiert und nicht interdisziplinär, da häufig an verschiedenen Orten und an spezifischen Aspekten eines Produktes in mehreren Projektteams entwickelt wird.

Projektteams aus einem Bereich tauschen sich häufig nicht untereinander aus und legen unabhängig von anderen Beteiligten Anforderungen fest. Demzufolge sammeln sich im Lastenheft schon bei der Entstehung beim OEM

verschiedenste Merkmale einer Komponente, die aber nicht im Bezug zueinander stehen und sich im schlimmsten Fall gegenseitig ausschließen. Ferner gibt es sprachliche Eigenheiten der Autoren und sprachliche Fehler, die innerhalb des Teams nicht auffallen und später nicht mehr korrigiert werden. Auch gibt es unternehmensinterne Richtlinien für die Formulierung von Anforderungen beim OEM, die das Verständnis auf Seite des Zulieferers erschweren können. Christian Allmann, Lydia Winkler und Thorsten Kölzow haben in [AWK06] solche Herausforderungen für das Requirements-Engineering zwischen OEMs und Zulieferern identifiziert.

Durch die entstehende große fachliche Breite und Tiefe der Spezifikationen im Pflichtenheft, können mitunter Dokumente mit einem Umfang von von mehreren tausend Seiten entstehen. Dies bedeutet einen hohen Aufwand bei der Transformation und Bündelung der Anforderungen in ein tatsächliches Produkt beim Zulieferer.

Prozessoptimierung durch Werkzeuge

Bei der Untersuchung von Projektlastenheften muss also nach Zusammenhängen und Bezügen zwischen mehreren Anforderungen gesucht werden, damit die Korrektheit des späteren Produktes gewährleistet ist. Die Analyse von Zusammenhängen zwischen Anforderungen stellt dabei aus Gründen der Effizienz ein Problem dar, wenn jeder Projektbeteiligte von Hand die für ihn relevanten Anforderungen aus dem Lastenheft extrahieren muss. Auch müssen die Lastenhefte an die Formulierungen und Ausdrucksweisen für Requirements-Management im Unternehmen angepasst werden.

Bislang gibt es jedoch kaum Werkzeugunterstützung, die effiziente Möglichkeiten zur automatisierten Überarbeitung und Anpassung einzelner Anforderungen aus dem Dokument bietet. Ansätze aus dem *Natural-Language-Processing*, kurz NLP, stellen gleichzeitig vielversprechende Forschungsfelder in der Informatik dar, die eine solche automatisierte Verarbeitung auf Basis von Sprachanalyse ermöglichen. Diese werden in Kap. 2 ausführlich erläutert. Syntax und Semantik der einzelnen Sätze und Zusammenhänge in Texten können auf Basis aktueller Trends wie Machine-Learning und dynamischer Programmierung zunehmend besser abgebildet werden.

1.2 Unternehmenskontext der Arbeit

Besonders betroffen von dem geteilten Entwicklungsprozess zwischen OEMs und Zulieferern ist die Automobilindustrie. Eine Vielzahl von Zulieferern beliefert die Automobilhersteller mit Komponenten für ihre Fahrzeuge. Einzelne Komponenten bestehen dabei häufig auch aus mehreren Einzelteilen, die wiederum von mehreren verschiedenen Zulieferern geliefert werden. [AWK06]

1.2.1 Einsatz bei HELLA

Einer dieser Zulieferer ist die HELLA GmbH & Co. KGaA (im folgenden HELLA genannt), die sich auf die Entwicklung und Produktion von Licht- und Elektronikkomponenten in Hard- und Software spezialisiert hat. HELLA ist ein international operierender deutscher Automobilzulieferer mit Hauptsitz in Lippstadt und zählt zu den Top 100 Automobilzulieferern weltweit [LV10]. Für weitere Information siehe [HE19] sowie die Firmenwebsite.

Zum Verfassen dieser Arbeit waren die Autoren fünf Monate bei der Firma beschäftigt. Dies beinhaltete zwei Monate Arbeit als Praktikanten und die verbleibenden drei Monate als Bacheloranden. Die Hauptaufgabe dort war die Mitarbeit im Bereich Requirements-Engineering für die Abteilung der Elektronik-Vorentwicklung *E-AE* (*Electronics-Advances Engineering*).

Die Abteilung beschäftigt sich mit dem Projektmanagement der Vorentwicklungsprojekte und beschäftigt sich, wie zuvor beschrieben, mit dem Austausch von Informationen der Firma HELLA mit diversen OEMs bezüglich gemeinsamer Projekte. Die genaue Darstellung dieser Prozesse erfolgt in Kap. 3.1. . Somit kommen dort auch die Lastenhefte der OEMs an, die dann an die Projektteams weitergeleitet werden.

1.2.2 Anforderungsmanagement im Unternehmen

Die Gewichtung einzelner Anforderungen in einem größeren Systemkontext fällt dort schwer, da nun Projektteams beim Zulieferer, die an der Entstehung des Lastenheftes nicht beteiligt waren, dieses verstehen sollen und auf ihre Teilkomponente anwenden müssen. Zusätzlich zu den teilweise widersprüchlichen Anforderungen aus dem Pflichtenheft kommen weitere Anforderungen des Zulieferers, wie etwa die Gewinnmarge und unternehmensinterne Richtlinien hinzu. Schlussendlich soll jedoch ein Produkt entwickelt werden, dass möglichst alle Anforderungen berücksichtigt. [MW02]

Zu diesem Zweck können bei Zulieferern spezielle Abteilungen existieren, die Projektteams bei der Identifikation und Gewichtung von Anforderungen unterstützen. Diese Abteilungen sind auf die Auswertung der Projektlastenhefte spezialisiert und erstellen Modelle zu Merkmalen des späteren Produktes. Die Experten stellen dabei die Eindeutigkeit und Verständlichkeit sicher und identifizieren relevante Anforderungen für die Mitglieder des Projektteams. Somit kann ein eigenes Anforderungsdokument für das Projektteam beim Zulieferer erstellt werden. Falls nicht vorhanden, muss das Team selbst über die Bewertung von Anforderungen entscheiden und Rücksprache mit dem OEM halten.

Bei HELLA existiert eine dementsprechende Abteilung, genannt *Projects, Methods and Tools*, kurz *PMT*. Im Rahmen unserer Beschäftigung konnte in enger Zusammenarbeit mit PMT genau ermittelt werden, wie Lastenhefte momentan verarbeitet werden, welche Verbesserungspotentiale dabei bestehen und wie diese genutzt werden können.

1.3 Hypothese

Hypothese dieser Arbeit ist, dass sich mithilfe von NLP Lastenhefte effizient automatisiert verarbeiten lassen. Damit kann die Arbeit von Requirements-Engineers, aber auch von Beteiligten an der Entwicklung beim Verständnis der Anforderungen, erleichtert werden. Dabei ist die Auswertung der Syntax und Semantik für ein tieferes Verständnis von Textzusammenhängen, also von verschiedenen Anforderungen, relevant.

Die sprachliche Analyse kann Requirements erkennen und kategorisieren, sowie in Beziehung zueinander setzen. Insbesondere lassen sich die beschriebenen Probleme bei der Überprüfung der Pflichtenhefte auf Konsistenz lösen oder zumindest vereinfachen. Dies kann mithilfe der in dieser Arbeit beschriebenen Programme geschehen, welche konzeptuell dargestellt und prototypisch implementiert werden.

Weiterhin soll geprüft werden, wie stark ein Mehrwert bei der Nutzung solcher Programme für ein Firma ausfallen kann. Dies beinhaltet ersparte Arbeitszeit, Kosten und vermiedene Fehler.

1.4 Methodik

Verschiedene Methoden aus dem Bereich Natural Language-Processing werden zunächst anhand des Rückblicks auf die Projektarbeit theoretisch und anhand praktischer Beispiele erläutert. Der betriebliche Kontext der Firma Hella wird durch genaue Darstellung der dortigen Verfahren beleuchtet.

Mit diesen Grundlagen wird dann die Unterstützung der dortigen Prozesse im Bereich Requirements-Engineering auf Basis von natürlichsprachlicher Textanalyse evaluiert. Die gewonnenen Erkenntnisse aus der Zusammenarbeit mit der Abteilung PMT wurden in Anforderungen an die zu entwickelnden Werkzeuge zusammengefasst, die als Basis für unsere Softwareentwicklung dienten. Dadurch entstand eine genaue Vorstellung der zu entwickelnden Software, welche mit Hilfe der Programmkonzepte von Konstantin Zichler ZIC19 entworfen und implementiert werden konnte.

Das Ergebnis stellten zwei Software-Prototypen auf *java*-Basis dar, deren Struktur und Funktionsweisen wir anhand von Modellen und Codebeispielen darstellen. Die Prototypen wurden mit originalen Pflichtenheften von verschiedenen OEMs und der HELLA getestet und evaluiert.

1.5 Aufbau der Arbeit

In der folgenden Ausarbeitung wird zunächst in Kapitel 2 ein Überblick über den Stand der Technik gegeben. Dies beinhaltet einerseits einen Rückblick auf die zuvor verfasste Projektarbeit zu NLP und einer Erläuterung der Grundlagen von Ontologien und andererseits einen Überblick von Arbeiten, die sich der gleichen oder ähnlichen Problemstellungen gewidmet haben.

Kapitel 3 beschäftigt sich für eine bessere Einordnung der beiden Programme mit dem betrieblichen Umfeld bei der HELLA. Hierzu wird zunächst, vor dem Hintergrund des momentanen Vorgehens, der Anwendungsfall der Programme genauer beschrieben. Somit lassen sich die betrieblichen Anforderungen an die Werkzeuge bestimmen, welche dann von den vorgestellten Konzepten abgedeckt werden sollen.

In Kapitel 4 wird das erste Werkzeug - der R2B-Converter - ausführlich beschrieben. Dies beinhaltet die Darstellung dessen Architektur, eine Vorstellung der Implementierung, eine Auflistung möglicher Erweiterungen und eine genaue Beschreibung des Testverfahrens der Software. Zu den Tests werden dabei präzise die Methodik, die Durchführung und die Ergebnisse

vorgestellt, welche in dem späteren Evaluations-Kapitel aufgegriffen und eingeordnet werden.

Das zweite Werkzeug - der Delta-Analyser - wird in Kapitel 5 behandelt. Dabei werden ähnlich zu Kapitel 4 die Architektur, die Implementierung, mögliche Erweiterungen und durchgeführte Tests beschrieben. Es wird genau dargelegt wie die Testergebnisse erzielt wurden, die später zusammen mit den Ergebnissen von Kapitel 4 evaluiert werden, um das Resultat der beiden Werkzeuge bewerten zu können.

Kapitel 6 stellt das zuvor erwähnte Evaluations-Kapitel dar. Hier findet zunächst die Auswertung der erhobenen Testergebnisse statt. Anschließend wird anhand dessen dargestellt, ob und inwiefern das Nutzen der Werkzeuge einen Mehrwert für die Firma erwirtschaften kann und somit im betrieblichen Umfeld eingesetzt werden sollte. Somit lässt sich die Korrektheit der in 1.2 gestellten Hypothese überprüfen.

Die Arbeit wird von Kapitel 7 abgeschlossen. Dieses beinhaltet eine Zusammenfassung der geleisteten Arbeit und der Ergebnisse. Außerdem wird zuletzt noch ein Ausblick für die Zukunft der Werkzeuge und die Lösung des Problems erläutert.

1.6 Autorenverzeichnis

1.: Ritter, Schul

1.1: Ritter

1.2: Ritter

1.3: Ritter, Schul

1.4: Schul

1.5: Schul

2.: Ritter, Schul

2.1: Schul

Kapitel 2

Stand der Technik

In diesem Kapitel wird zunächst ein Rückblick auf die zuvor von den Autoren verfasste Projektarbeit und die darin enthaltenen Grundlagen von NLP gegeben. Aufgrund ihrer besonderen Relevanz für die Programme werden anschließend zusätzliche Ontologien als Teil von NLP genauer beleuchtet. Außerdem wird mit Hilfe von verwandten Arbeiten ein Überblick über die Problematik und verschiedene Lösungsansätze geboten.

2.1 Rückblick auf die Projektarbeit

In [?] sind die Grundlagen von NLP im Rahmen einer Projektarbeit zusammengefasst. In diesem Abschnitt werden die für die Arbeit relevanten Aspekte kurz wiederholt.

2.1.1 Grundlagen zu Natural Language Processing

Definition und Ziel

Natural Language Processing ist zu verstehen als die automatisierte oder halb-automatisierte Verarbeitung von natürlicher Sprache mit Hilfe eines Computers. In NLP sind verschiedenste wissenschaftliche Bereiche verknüpft. Hauptsächlich handelt es sich dabei um Linguistik und Informatik, jedoch gibt es auch viele Verbindungen zur Psychologie, Philosophie Mathematik und Logik[COP04].

Ziel von NLP ist üblicherweise die Extraktion von Informationen aus einem Text durch kleinschrittige Textanalyse. Diese Informationen werden dem

Text direkt über sogenannte Annotationen angehängt oder anders abgespeichert, etwa in Oestantologien, welche in einem späteren Abschnitt genauer erläutert werden.

Beispiele von Forschung und Entwicklung

Die wissenschaftliche Erforschung von NLP entstand Mitte des 20. Jahrhunderts, als der Linguist Noam Chomsky zeigte, dass sich der englischen Sprache formalisieren und somit automatisieren lassen [CHO57].

Frühe Projekte wie der Sprachcomputer ELIZA aus den 1960er Jahren waren somit bereits in der Lage eine natürlichsprachige Frage als Texteingabe entgegenzunehmen und üblicherweise eine plausible Antwort zu liefern.

Heutzutage gibt es ein Vielzahl von Spracherkennungssoftware auf Smartphones oder Smart-Devices wie ALEXA, welche in der Lage sind gesprochene Kommandos zu erkennen und über das Internet an Server zu leiten, welche in der Lage sind diese Kommandos schnell und effizient zu verarbeiten [HAO14].

Herausforderungen von NLP

Durch die Verbindung von Linguistik und Informatik stellt sich NLP generell der Herausforderung sowohl inhärente Probleme der Sprache, als auch Probleme der Automatisierung lösen zu müssen.

Zu den sprachlichen Problemen gehören Aspekte wie Mehrdeutigkeit, Sarkasmus und Umgangssprache bzw. Redewendungen. Diese lassen sich oft nicht einzig und allein durch den gegebenen Text erklären, sondern beruhen auf Kontext und Situation.

Bei der Implementierung ergeben sich dann Probleme, trotz der sprachlichen Besonderheiten konsistente Regelmäßigkeiten als Grundlage zur Automatisierung zu finden.

Linguistische Analyse

Um einen Text möglichst Fehlerfrei verarbeiten zu können bedarf es also einem kleinschrittigen und präzisen Verfahren. Dieses bezeichnet man als linguistische Analyse.

Sie ist aufgeteilt in vier wichtige Schritte, welche aufeinander aufbauen und an Komplexität zunehmen. Im folgenden sind diese Schritte aufgeführt und kurz erläutert [?]:

1. Morphologische Analyse - Die Zerlegung von Wörtern bezüglich ihrer Struktur. Die Komposition aus Präfix, Wortstamm und Suffix von Wörtern wird erkannt, um Fall, Tempus, Numerus etc. des Wortes zu bestimmen. Im Englischen zeigt so die Endung -ed bei Verben die Vergangenheitsform an. Dabei ist zu beachten, dass Mehrdeutigkeiten entstehen können.
2. Syntax - Aufbau von Sätzen durch einzelne Wörter. Jede Sprache besitzt syntaktische Regelungen bezüglich der auftauchenden Wörter; im Deutschen folgt etwa auf einen Artikel irgendwann ein Substantiv oder bestimmte Signalwörter geben Satztyp und Tempus an. Auf Basis dieses Wissens können formale und strukturelle Analysen der Textbestandteile erfolgen.
3. Semantiken - Die Identifikation der Bedeutung von einzelnen Sätzen und Wörtern. Die Semantik eines Textabschnittes wird häufig als „Logik“ bezeichnet und fragt nach dessen Bedeutung bzw. Thema. Semantische Deutungen können anhand von Kompositionen einzelner Wörter und Sätze auf Basis der semantischen Analyse erkannt werden.
4. Kontext - Darstellung von satzübergreifenden Zusammenhängen syntaktischer und semantischer Natur. Mehrere Sätze können über Konstrukte wie etwa Pronomen verbunden sein, wenn sich das Pronomen des einen Satzes auf ein Subjekt des anderen bezieht.

Diese Aufteilung lässt sich jedoch in noch weitere Teilschritte für die Spracherkennung und -generierung als NLP-Architektur unterteilen. Diese wird im folgenden Abschnitt genauer erläutert.

2.1.2 Umsetzung der Verarbeitungsschritte natürlicher Sprache

Dieser Abschnitt beschäftigt sich primär mit den verschiedenen Schritten der sprachlichen Analyse. Diese Differenzierung kann dabei, wie zuvor beschrieben, anhand der linguistischen Analyse unterschiedlicher Bestandteile von Sprache vorgenommen werden. Diese sukzessive Verarbeitung wird im folgenden anhand verschiedener Algorithmen und Verfahren erläutert, die im wesentlichen in [?] zu finden sind.

Das Ziel jeder Analyse ist es, Wissen über einen Teilbereich von Sprache zu generieren. Das Wissen über den Zusammenhang von Wörtern eines Textes kann eindeutig sein, jedoch können besonders bei der morphologischen Analyse sprachliche Mehrdeutigkeiten auftreten. Dies wird in 2.1.3-Morphologie näher erläutert.

Die Präzision dieser Aussagen fällt dementsprechend abhängig von der Treffergenauigkeit und dem Abdeckungsgrad der linguistischen Analyse durch die verwendeten Verfahren ab. Es ist dabei zu beachten, dass die Analyse typischerweise als Architektur aufgebaut ist, da die verschiedenen Schritte aufeinander aufbauen [COP04]. Somit können, basierend auf den Ergebnissen, inhaltliche Aussagen und Zusammenhänge über den zugrundeliegenden Text getroffen werden.

[COP04] identifiziert dabei eine allgemeine NLP-Architektur, die basierend auf der linguistischen Analyse diese Teilbereiche voneinander abgrenzt. Für ein besseres Verständnis der Verfahren werden die für diese Arbeit relevanten Schritte in [?] wie folgt kurz erläutert:

1. Input Processing - Erkennung der Dokumentensprache und Normalisierung des Textes. Im ersten Schritt geht es um das korrekte Format des Eingabedokumentes für die Verarbeitung. Dieser Vorgang stellt jedoch noch keine Analyse dar, sondern dient lediglich der Vorbereitung.
2. Morphologische Analyse - Die meisten Sprachen sind im Bezug auf ihre Grammatik und syntaktische Struktur der Wörter in Systemen abgeschlossen, etwa durch Modellierung mittels Automaten zur Erzeugung der Worte. Auch können Vokabeln in Wörterbüchern/Lexika gesammelt und Regeln auf ihnen formuliert und gespeichert werden.
3. Part-of-Speech-Tagging - Die einzelnen Wörter eines Satzes werden im Hinblick auf den Sach- oder Satzzusammenhang, wie auch auf die Stellung im Satz hin analysiert. Basierend auf Kontext und/oder Erfahrungswerten bzw. Heuristiken können die Wörter korrekt erfasst und deren Fall getaggt werden. Subjekte, Prädikate usw. werden identifiziert. Dazu kann eine Wissensbasis mit Trainingsdaten und die Einordnung des Kontextes darin verwendet werden.
4. Parsing - Die Ergebnisse der vorherigen Schritte werden weiter verarbeitet und in ein standardisiertes Format gebracht. Syntaktische Zusammenhänge, wie etwa das Zusammenfassen von Wörtern zu einer

gemeinsamen Bedeutungsphrase und das Zuordnen von Verben zu einem Nomen können nach dem Parsing dargestellt werden.

5. Disambiguation - Die Entfernung von Mehrdeutigkeiten auf Basis der Ergebnisse des Parsings stellt einen entscheidenden Schritt für die semantischen Analysen dar. Nur wenn die Aussage und der Kontext eines Satzes klar erkennbar sind, kann dessen Semantik gezielt abgeleitet werden.
6. Context Module - Textbausteine, deren Interpretation semantisch auf dem Kontext anderer Sätze oder Wörtern beruhen, können erfasst werden. Bei Anaphern ist etwa das Verständnis des aktuellen Kontexts abhängig von einer vorherigen Deutung.
7. Text Planning - Die Sachzusammenhänge und Semantiken, die aus dem zugrundeliegenden Text extrahiert wurden, werden für die Darstellung im fertigen Text definiert. Es wird festgelegt, welche der Bedeutungen qualitativ übertragen und vermittelt werden sollen. Die Reihenfolge und Umfang der behandelten Themen wird geschätzt und definiert.
8. Tactical Generation - Die Bedeutungen werden in Form konkreter Zeichenketten generiert, die die gewünschte Bedeutung enthalten. Diese Textbausteine basieren häufig direkt auf den Ergebnissen des vorherigen Parsings, da dort schon Bedeutungen zusammengefasst werden können. Der quantitative Teil des Textes wird erzeugt.
9. Morphological Generation - Die erzeugten Sätze werden morphologisch an die Rahmenbedingungen der verwendeten Sprache angepasst. Grammatiken und Regeln der Satzkonstruktion werden auf die erzeugten Wörter angewendet, sodass ein lesbarer und nachvollziehbarer Text entsteht.
10. Output Processing - Der Text wird nun, nachdem dieser inhaltlich komplett erzeugt vorliegt, an das Design und Format des jeweiligen Einsatzzwecks angepasst. Sollte Text, anders als in dieser Arbeit der Einfachheit halber angenommen, nicht als geschriebenes Wort ohne besondere Formatierung vorliegen, kann dieser einem *Formatter* zur Designanpassung übergeben werden. Ferner kann etwa mithilfe von Text-to-Speech eine Audioausgabe erfolgen.

Wie später gezeigt wird, kann die in dieser Arbeit beschriebene Entwicklung der Software-Prototypen als eine Umsetzung eben dieser NLP-Architektur verstanden werden.

2.1.3 Syntaktische Analyse von Wörtern und Sätzen

In diesem Abschnitt wird die Zerlegung der Syntax anhand logischer Modelle von Sprache beschrieben. Begonnen wird mit der einfachen Struktur und Zusammensetzung einzelner Wörter aus der Grammatik. Die Kombination aus Wortstämmen mit Prä- und Suffixen kann etwa in einem Lexikon gesucht werden, dass alle Wörter einer Sprache enthält. Basierend auf dieser Morphologie eines Wortes kann dann in der Analyse mehrerer Wörter auf die Rolle von Wörtern im Satzzusammenhang geschlossen werden.

Morphologie

Dieser Abschnitt befasst sich genauer mit der Analyse der Struktur und Zusammensetzung von Wörtern. Dies wird als morphologische Analyse bezeichnet. Wie im folgenden beschrieben, können auf dieser Basis logische Modelle wie etwa Automaten erstellt werden, die einen Teilbereich der Sprache effektiv durch Regeln abdecken. Für ein besseres Verständnis der im Rahmen dieser Arbeit vorgestellten Ergebnisse beschränken wir uns in dieser Arbeit auf die vergleichsweise einfach strukturierte englische Sprache. Aufgrund der großen Fortschritte und Erfahrungen in der Sprachforschung, werden die folgenden Beispiele auf Englisch behandelt.

In der englischen Sprache besteht jedes Wort aus einem Wortstamm, sowie den sogenannten Affixen. In einem Satz könnte etwa das Wort „*believable*“ auftauchen. Das Wort stammt von der Grundform „(to) *believe*“ (zu Deutsch: „glauben“) ab. Das *-able* stellt ein Suffix dar, also eine Wortendung. Zusammen mit den Präfixen, die dem Wortstamm vorangestellt sind, gehört *-able* somit zur Kategorie der Affixe.

Das Ziel der morphologischen Analyse ist nun, die Zusammensetzung des Wortes nachzuvollziehen und die Information für die weitere Verarbeitung bereitzustellen. Die Generierung des Wortes *believable* kann anhand von Buchstabierregeln erfolgen, die alle zulässigen Wörter auf englisch anhand der Wortstämme und Affixe erzeugen können. Es fällt auf, dass in unserem Beispiel *believable* vor dem Hinzufügen des Suffixes *-able* zunächst das *-e* von

der Grundform (*to*) *believe* entfernt werden muss. Gemäß [?] bedeutet dies formal ausgedrückt:

$$e \rightarrow \varepsilon^{\wedge}_{_} \text{able}$$

In der Logik kann man diese Regeln als *Zustandswandler* visualisieren. In Abb. 2.1 ist der Zustandswandler gezeigt, die das Suffix *-able* an entsprechende Wörter anhängen kann.

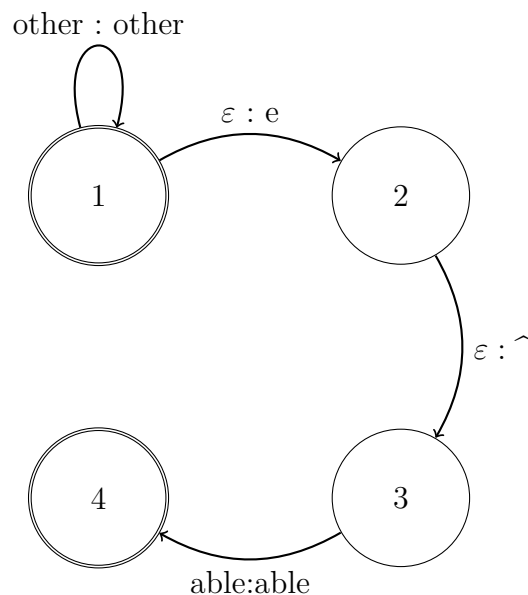


Abbildung 2.1: Nicht-deterministischer Zustandswandler der able-Regel [?]

Diese Bestandteile von Wörtern könnten innerhalb der morphologischen Analyse in verschiedenen Lexika enthalten sein, dass die entsprechenden Informationen bereitstellt. Gemäß [COP04] bedarf es dabei drei verschiedener Lexika:

1. Liste von *Affixen* (zusammen mit den Informationen, die damit verbunden sind, z.B. Negation „un-“)
2. Liste von *unregelmäßigen Wörtern* (zusammen mit den Informationen, die damit verbunden sind, z.B. „went“:simple past, „(to) go“)

3. Liste von *Wortstämmen* (zusammen mit syntaktischer Kategorie, z.B. „believe“:verb)

Es ist zu beachten, dass die in der morphologischen Analyse verwendeten Buchstabierregeln zwar zu jeweils eindeutigen Ergebnissen kommen können, die Analyse selbst jedoch nicht eindeutig sein muss. Dies kann, wie bereits zuvor beschreiben, durch sprachliche Mehrdeutigkeiten entstehen.

Das Wort „*even*“ könnte isoliert etwa einmal als *gleich*, also als Adjektiv kategorisiert, ins Deutsche übersetzt werden, jedoch auch sinngemäß *sogar wenn*, also das Adverb mit einem oder mehreren Bezugswörtern meinen. Die mehreren möglichen Bedeutungen des einzelnen Wortes können jedoch in der Regel durch die Position grammatikalische Funktion im Satz aufgelöst werden.

2.2 Ontologien

2.3 Verwandte Arbeiten

Kapitel 3

Betriebliches Umfeld - Hella Use-Case

Im betrieblichen Umfeld liegen zu Beginn jeden Entwicklungsprojektes für neue Produkte die Aufgaben und Ziele für die Entwicklung als Dokumente vor. Forschungsergebnisse finden Anwendung in der Vorentwicklungsphase, in der die Eignung der Erkenntnisse für neue Produkte eines Unternehmens evaluiert wird. Die Produktentwicklung unterliegt dabei bestimmten Kriterien und Faktoren, die den unternehmerischen Erfolg beeinflussen. Neben betriebswirtschaftlichen Einflüssen wie der Einordnung des Produktes in der Wertschöpfungskette sind es dabei besonders technische Anforderungen an das Produkt, die definiert und während der Produktentwicklung eingehalten werden müssen. Verschiedenste Akteure aus einem Unternehmen sind dabei an der Festlegung der Anforderungen an ein Entwicklungsprojekt bzw. Produkt beteiligt.

In der Automobilindustrie betrifft dieser Ablauf zumeist die Entwicklung neuer Fahrzeugkomponenten, heutzutage meist elektronische und mechanische Bausteine. Diese Bausteine werden dabei nicht sämtlich vom Fahrzeughersteller (OEM) selbst, sondern durch eine Vielzahl von Zulieferern produziert und entwickelt. Die Produktspezifikationen liegen meist digital als Texte, Tabellen und Grafiken vor und werden an den Zulieferer übermittelt. Nach dem Entwicklungsprozess steht dann die (Serien-)entwicklung und -fertigung des Produktes für das Ausrollen in großen Stückzahlen an den Hersteller, der das zugelieferte Produkt dann in seinen Produkten verwendet. Um dies zu erreichen, müssen während des gesamten Prozesses die Anforderungen, die das Systemumfeld des Fahrzeugherstellers hat, berücksichtigt

und eingehalten werden.

Die Anforderungen an das Produkt, etwa technische Rahmenbedingungen, werden dabei von vielen verschiedenen Domänenexperten beim OEM formuliert und in das sogenannte Pflichtenheft für die Entwicklung eingetragen. Beteiligte sind etwa Produktdesigner, Ingenieure und Systemtechniker, die an verschiedenen Stellen im Lastenheft Anforderungen an eine Komponente festlegen. Diese Beteiligten sind in der Regel auf ihren Bereich spezialisiert und nicht interdisziplinär, zudem gibt es sprachliche Eigenheiten der Autoren und unternehmensinterne Richtlinien für die Formulierung, die das Verständnis erschweren können. Demzufolge sammeln sich im Lastenheft verschiedenste Merkmale einer Komponente, die aber nicht im Bezug zueinander stehen und sich im schlimmsten Fall gegenseitig ausschließen.

Durch diese fachliche Breite und Tiefe der Spezifikationen im Pflichtenheft, aber auch durch den Umfang des Lastenheftes von mehreren tausend Seiten, kommt es häufig insbesondere zu Verständnisproblemen auf Seite des Zulieferers. Die Gewichtung einzelner Anforderungen in einem größeren Systemkontext fällt dort schwer, da nun Projektteammitglieder, die an der Entstehung des Lastenheftes nicht beteiligt waren, dieses verstehen und ein Produkt entwickeln sollen, dass möglichst alle Anforderungen berücksichtigt. In Texten muss also nach Zusammenhängen und Bezügen zwischen mehreren Anforderungen gesucht werden, damit die Korrektheit des späteren Produktes gewährleistet ist.

Die Analyse von Zusammenhängen zwischen Anforderungen stellt dabei aus Gründen der Effizienz ein Problem dar, wenn jeder Beteiligte von Hand die für ihn relevanten Anforderungen aus dem Lastenheft extrahieren muss. Auch müssen die Lastenhefte an die Formulierungen und Ausdrucksweisen für Requirements-Management im Unternehmen angepasst werden. Bislang gibt es jedoch kaum Werkzeugunterstützung, die effiziente Möglichkeiten zur automatisierten Überarbeitung und Anpassung einzelner Anforderungen aus dem Dokument bietet. Ansätze aus dem *Natural-Language-Processing* (NLP) stellen gleichzeitig vielversprechende Forschungsfelder in der Informatik dar, die eine solche automatisierte Verarbeitung auf Basis von Sprachanalyse ermöglichen. Syntax und Semantik der einzelnen Sätze und Zusammenhänge in Texten können auf Basis aktueller Trends wie Machine-Learning und dynamischer Programmierung zunehmend besser abgebildet werden.

3.1 Re-Prozesse bei Hella und allgemein in Firmen

3.2 Betriebliche Anforderungen

3.3 Ansatz und Konzept unserer Werkzeuge

Kapitel 4

R2B-Converter

4.1 Architektur Klassen und Verteilung der Ressourcen

4.1.1 Umsetzung der NLP-Architektur in unserem Werkzeug

In diesem Abschnitt wird die Einordnung unseres Prototypen als NLP-Werkzeug in die von [?] definierte Architektur behandelt. Die verschiedenen Schritte aus Kap. 2.2. haben wir bei der Analyse und Verarbeitung der Lastenhefte in Programmbestandteilen umgesetzt. Im Arbeitsprozess unserer Entwicklung können wir die Aufgaben unseres Programms dort wie folgt einordnen:

4.2 Implementierung (bisschen Code, GUI, Listenarchitektur, Workflow für User

4.3 mögliche Erweiterungen

4.4 Test

4.4.1 Methodik

4.4.2 Durchführung

4.4.3 Ergebnisse

Kapitel 5

Delta-Analyse

5.1 Architektur

5.2 Implementierung

5.3 mögliche Erweiterungen

5.4 Test

5.4.1 Methodik

5.4.2 Durchführung

5.4.3 Ergebnisse

Kapitel 6

Evaluation

6.1 Auswertung der Testresultate

6.2 Mehrwert?

6.3 Ziel erreicht? Hypothese reviewen und schwafeln

Kapitel 7

Fazit

7.1 Zusammenfassung

7.2 Ausblick

Literaturverzeichnis

- [LV10] F. Langenscheidt, B. Venohr. „Lexikon der deutschen Weltmarktführer: Die Königsklasse deutscher Unternehmen in Wort und Bild.“, Deutsche Standards–Gabal Verlag Google Scholar (2010).
- [HE19] Firma HELLA. „Unternehmensinformationen in Kürze.“, aus <https://www.hella.com/hella-com/de/HELLA-im-Ueberblick-723.html>, abgerufen am 09.01.19.
- [BAL10] Helmut Balzert. „Lehrbuch der softwaretechnik: Basiskonzepte und requirements engineering.“, Springer-Verlag, 2010.
- [MW02] Matthias Weber, Joachim Weisbrod. „Requirements engineering in automotive development-experiences and challenges.“, Requirements Engineering, 2002. Proceedings. IEEE Joint International Conference on. IEEE, 2002.
- [HL01] Hubert F. Hofmann, Franz Lehner. „Requirements engineering as a success factor in software projects.“IEEE software 4 (2001): 58-66.
- [AWK06] Christian Allmann, Lydia Winkler, Thorsten Kölzow. „The requirements engineering gap in the OEM-supplier relationship.“, Journal of Universal Knowledge Management 1.2 (2006): 103-111.
- [DGE19] Deutsche Gesellschaft für EMV-Technologie e.V. „CRS Customer Requirements Specification“, aus https://www.demvt.de/publish/viewfull.cfm?ObjectID=ba9a0878_e081_515d_74a3411df6771be8, abgerufen am 09.01.19.
- [COP04] Ann Copestake, University Of Cambridge, „Natural Language Processing“, 2004.