

KI-basierte Extrahierung von Anforderungen aus Regularien für die Automobilentwicklung

AI-based extraction of requirements from regulations for automotive engineering

Iris Gräßler¹, Deniz Özcan^{1,*}, Daniel Preuß¹

¹ Heinz Nixdorf Institute, Chair of Product Creation, Paderborn University

* *Korrespondierender Autor:*

Deniz Özcan
Universität Paderborn
Heinz Nixdorf Institut
Lehrstuhl für Produktentstehung
Warburger Str. 100
☎ 05251/606262
✉ deniz.oezcan@hni.upb.de

Abstract

Automotive engineering requires compliance with regulations for certification. In specifications, regulations are referenced, which need to be analyzed manually to elicit requirements. This process is time-consuming and leads to high costs. The aim of this research is to evaluate artificial intelligence (AI) models in terms of extracting requirements automatically from regulations. Relevant AI models are identified in a systematic literature analysis and evaluated using success criteria. The most promising AI models are implemented in a pipeline for requirements extraction. The performance of these AI models is assessed in a comparative study using automotive regulations. The results show which AI models are best suited for this task.

Keywords

Requirements Engineering, Natural Language Processing, Artificial Intelligence, Automotive Engineering

1. Motivation

In Projekten zur Automobilentwicklung zeigt sich, dass aktuell erheblicher Aufwand nicht in die Ermittlung funktionaler Anforderungen, sondern in die Extrahierung von Anforderungen aus Regularien investiert wird [1, 2]. Ein Beispiel ist das einheitliche Regelwerk mit Regularien zur Zertifizierung von Kraftfahrzeugen, das durch die UN/ECE Transport festgelegt wird [3]. In der Entwicklung von Subsystemen von Automobilen wird in Lastenheften auf diese Regularien der Kommission verwiesen [4]. Beispielsweise wird die Regularie UN/ECE R11 im Lastenheft referenziert, um festzulegen, welchen Kräften das Subsystem „Türverschlusssystem“ standhalten muss [5]. Aktuell müssen Entwickler die für die Zertifizierung relevanten Anforderungen an das Subsystem manuell in den Regularien identifizieren und entsprechend auslesen, um diese in die Anforderungsliste aufzunehmen [6]. Die UN/ECE Transport hat 160 Regularien mit einem Seitenumfang von jeweils bis zu 150 Seiten für die Automobilentwicklung verabschiedet; die beispielhaft genannte Regularie UN/ECE R11 umfasst 28 Seiten [5]. Aufgrund der hohen Anzahl von Regularien und deren Umfang verursacht dieser Prozess großen Aufwand. Änderungen von Regularien führen dazu, dass Anforderungen erneut extrahiert werden müssen. Außerdem werden Details von Anforderungen aufgrund des Umfangs der Regularien von Entwicklern häufig übersehen [7]. Aus Ergebnissen aus Interviews mit Partnern aus der Automobilindustrie wurde abgeleitet, dass ein Bedarf nach einer automatisierten Extrahierung von Anforderungen aus Regularien besteht, um den Aufwand zu verringern und alle relevanten Informationen auszulesen. Ziel dieses Beitrags ist es daher, geeignete Modelle zu identifizieren und vergleichend zu bewerten, um Anforderungen aus Regularien für die Automobilentwicklung effizient zu extrahieren.

Zu Beginn des Beitrags wird der Stand der Technik zur automatisierten Extrahierung von Anforderungen aus Regularien dargestellt (Kapitel 2). Das konkrete Forschungsproblem und die Forschungsziele werden aufgeführt (Kapitel 3). Darauf bezogen wird das Vorgehen zur Erreichung der Forschungsziele beschrieben (Kapitel 4) und die Anwendung zur Bewertung der identifizierten Modelle dargestellt (Kapitel 5). Außerdem werden die Ergebnisse der Anwendung der erfolgversprechendsten Modelle aufgezeigt und anhand einer Vergleichsstudie diskutiert. Abschließend wird die Zusammenfassung der Ergebnisse dargestellt und ein Ausblick zum weiteren Forschungsbedarf gegeben (Kapitel 6).

2. Stand der Technik

Lösungsansätze zur automatisierten Extrahierung von Anforderungen aus Regularien werden mittels systematischer Literaturanalyse nach MACHI und MCEVOY ermittelt [8]. Durch Scanning werden Beiträge in Online-Forschungsdatenbanken identifiziert. Hierzu werden Suchbegriffe definiert und mit Booleschen Operatoren zu einem Suchstring verknüpft: „A AND B AND (C1 OR C2)“ (siehe Tabelle 1). Die Suchbegriffe werden von dem Ziel abgeleitet, Anforderungen aus Regularien mit einem hohen Automatisierungsgrad zu extrahieren. Es werden zusätzlich zu den deutschen Suchbegriffen auch englische Suchbegriffe verwendet, um internationale Beiträge zu berücksichtigen. Als Quelle von Anforderungen werden neben Regularien (Suchbegriff C1) auch allgemein Dokumente (Suchbegriff C2) betrachtet. Diese Ergänzung des Suchbegriffs „document“ – als Abstrahierung von Regularien – wird aufgrund der Annahme vorgenommen, dass nur innerhalb einer geringen Anzahl von Beiträgen die automatisierte Extrahierung von Anforderungen aus Regularien untersucht wird. Die Relevanz der Beiträge wird anhand des Titels, der Zusammenfassung und des Hauptteils bewertet. Beiträge mit einem unpassenden Titel werden direkt aussortiert und nicht weiter betrachtet, aussortierte Titel beinhalten unter anderem das Wort „construction“. Bei passendem Titel und passender Kurzzusammenfassung wird der Hauptteil des Beitrags gesichtet. Auf dieser Grundlage werden inhaltlich nicht relevante Dokumente ausgeschlossen. Insgesamt werden acht relevante Beiträge in den Suchergebnissen identifiziert, welche die automatisierte

Extrahierung von Anforderungen aus Dokumenten bzw. Regularien untersuchen. Der Suchstring wird auf die Online-Forschungsdatenbank Google Scholar angewendet, anschließend werden durch Skimming relevante Beiträge identifiziert. Die Suchergebnisse werden gefiltert und kartiert, um verwandte Beiträge zu erkennen (Mapping). Beiträge, die älter als zehn Jahre sind, werden nicht berücksichtigt. 13 Duplikate werden aussortiert. Als Resultat verbleiben 232 Suchergebnisse. Diese Suchergebnisse werden hinsichtlich relevanter Modelle geprüft.

Tabelle 1: Suchbegriffe der systematischen Literaturanalyse

Kategorie	Thema	Suchbegriffe
A	Extrahierung von Anforderungen	A: requirement extraction OR Anforderungsextrahierung
B	Automatisierung	B: automated OR automatisiert
C	Quelle	C1: regulation OR Regularie
		C2: document OR Dokument

Für die automatisierte Extrahierung von Anforderungen aus natürlich-sprachlichen Dokumenten ist die Nutzung verschiedener Methoden aus dem Bereich Natural Language Processing (NLP) notwendig. Die identifizierten Modelle werden in zwei Gruppen eingeordnet: Extrahierung von Anforderungen aus Dokumenten (Gruppe 1) und Extrahierung von Anforderungen aus Regularien (Gruppe 2).

Gruppe 1: Eine Methode zur Klassifizierung von Sätzen einer Softwarespezifikation wird von KURTANOVIC und MAALEJ entwickelt [9]. In der Methode werden funktionale und nicht-funktionale Anforderungen klassifiziert. Für die Klassifikation werden Part-of-Speech-Tags oder N-Gramme von Wörtern als Merkmale verwendet. BOUTKOVA und HOUDEK identifizieren relevante Anforderungen für die Wiederverwendung durch die Extrahierung von Merkmalen aus natürlich-sprachlichen Anforderungen in Spezifikationsdokumenten [10]. In eigenen Vorarbeiten wurden Anforderungen in Lastenheften von Automobilkomponenten klassifiziert und künstliche Daten zum Training von Klassifikationsmodellen generiert [11].

Gruppe 2: JANPITAK ET AL. extrahieren Anforderungen aus Regularien, um die Einhaltung von Vorschriften einer Organisation zu überprüfen [7]. MICH und ZENI stellen die wichtigsten Erfahrungen, die bei der Entwicklung und Anwendung von KI-Modellen für die Extrahierung von Anforderungen aus Regularien im Kontext von Software-Projekten gemacht wurden, dar, um Probleme der Benutzerfreundlichkeit herauszuarbeiten [12]. HASSAN und LE extrahieren Anforderungen aus Regularien für Bauprojekte [13]. Sie klassifizieren Textsegmente mittels KI-Modellen nach „Anforderung“ und „keine Anforderung“.

In den Beiträgen aus Gruppe 1 und Gruppe 2 werden Modelle des maschinellen Lernens (siehe Grundlagen in [14]) zur automatisierten Extrahierung von Anforderungen aus Dokumenten bzw. Regularien verwendet. Besonders die Modelle Support Vector Machine (SVM), logistische Regression, multinomiale Naive Bayes, künstliche neuronale Netze (KNN) und Bidirectional Encoder Representations from Transformers (BERT) erzielen laut den Beiträgen erfolgversprechende Resultate, weshalb diese Modelle in diesem Beitrag weiter untersucht werden. Eine SVM ist ein maschineller Lernalgorithmus der Diskriminanzanalyse, der eine Trennung zwischen Datenpunkten durch maximale Abstände zwischen den Klassen herstellt. Die logistische Regression (LR) ist ein statistisches Modell, das verwendet wird, um die Wahrscheinlichkeit einer binären Ereignisvariable basierend auf den Werten von Prädiktorvariablen zu schätzen und ist somit ein binäres Klassifikatormodell. Die multinomiale Naive Bayes (MNB) ist ein probabilistisches Klassifikationsmodell, das auf dem Bayes-Theorem basiert und mehrere Klassen mit diskreten Merkmalsvariablen berücksichtigt und verarbeiten kann. Künstliche neuronale Netze sind computergestützte Modelle, die durch eine Netzwerkstruktur von künstlichen Neuronen und deren Verbindungen inspiriert sind und

komplexe Muster in Daten erkennen und verarbeiten können. Sie werden in Schichten eingeteilt. Bidirectional Encoder Representations from Transformers sind ein neuronales Netzwerkmodell, das auf der Transformer-Architektur basiert und zur Sprachverarbeitung verwendet wird, indem es kontextabhängige Wortrepräsentationen erzeugt. In den identifizierten Publikationen werden keine Regularien aus der Domäne Automobilentwicklung betrachtet, Regularien sind jedoch ein aufwandsintensiver Teil der Entwicklung. Verschiedene Aspekte erschweren die Übertragbarkeit der Anwendung der existierenden Lösungen auf andere Dokumententypen – Regularien zur Automobilentwicklung – und mindern dadurch die Leistungsfähigkeit. Dazu zählen das Dateiformat, Formatierungen im Text, Verweise, Qualitätsdefizite, Anforderungsbeschreibungen, Anforderungsarten und die Sprache [15]. Für diesen Beitrag sind Formatierungen im Text, Qualitätsdefizite und Anforderungsbeschreibungen relevant. Deshalb müssen die beschriebenen KI-Modelle für den Dokumententyp untersucht werden und die Leistungsfähigkeit bewertet werden.

3. Forschungsproblem und Forschungsziele

Die Analyse des Stands der Technik zeigt, dass bereits Lösungen zur automatisierten Extrahierung von Anforderungen aus Dokumenten existieren, die auf Künstlicher Intelligenz (KI) basieren. Die Leistungsfähigkeit dieser Ansätze, die anhand von Klassifikator-Kennzahlen wie dem F1 Score ausgedrückt wird, wurde noch nicht für die Anwendung auf Regularien zur Automobilentwicklung untersucht. Die Übertragbarkeit der Ansätze muss für diesen Dokumententyp untersucht werden. Das Ziel ist es, die KI-Modelle mit der höchsten Leistungsfähigkeit zur automatisierten Extrahierung von Anforderungen aus dem Dokumententyp „Regularien zur Automobilentwicklung“ zu identifizieren.

Zur Bewertung der Leistungsfähigkeit werden Klassifikator-Kennzahlen berechnet und in einer Vergleichsstudie gegenübergestellt. Es wird damit die folgende Forschungsfrage beantwortet: „Wie hoch ist die Leistungsfähigkeit existierender KI-Modelle zur Extrahierung von Anforderungen aus Regularien in der Automobilentwicklung?“

4. Verwendete Methoden und Vorgehensweisen

Das wissenschaftliche Vorgehen wird in Anlehnung an ULRICH [16] in sechs Schritten gestaltet. Ergebnisse der ersten drei Schritte sind in den Kapiteln 1 bis 3 dokumentiert. Diese Schritte dienen der Klärung des Forschungsziels und sind hellgrau in Bild 1 dargestellt. Im ersten Schritt werden praxisrelevante Probleme mit Partnern aus der Automobilindustrie abgeleitet und durch eine Literaturanalyse abgesichert. Aus der Zusammenarbeit mit Ingenieuren eines weltweit führenden Automobilherstellers werden Herausforderungen aus der Anforderungsentwicklung abgeleitet. Es wurde festgestellt, dass vor allem die manuelle Extrahierung von Anforderungen aus Regularien einen hohen Aufwand verursacht. Im zweiten Schritt werden Lösungsalternativen ermittelt. Es werden mittels systematischer Literaturanalyse nach MACHI und MCEVOY [8] Methoden zur automatisierten Extrahierung von Anforderungen aus Regularien ermittelt. Im dritten Schritt wird der Forschungsbedarf konkretisiert, indem die ermittelten Methoden analysiert werden und Handlungsbedarf identifiziert wird. Schritte vier bis sechs dienen zur Erreichung des Forschungsziels (dunkelgrau in Bild 2). In Schritt vier werden Erfolgsfaktoren für die automatisierte Extrahierung von Anforderungen aus Regularien abgeleitet und definiert. Die Lösungsalternativen werden anhand der fünf Erfolgsfaktoren bewertet und in einer Vergleichsmatrix gegenübergestellt. Die KI-Modelle mit der höchsten Bewertung werden selektiert. Im fünften Schritt werden die selektierten Modelle zur automatisierten Extrahierung von Anforderungen aus Regularien in einer „Data Pipeline“ implementiert und für einen Vergleich vorbereitet. Abschließend wird im sechsten Schritt die Leistungsfähigkeit der KI-Modelle verglichen, indem die Implementierung zur Extraktion von Anforderungen aus den Regularien der UN/ECE verwendet wird.

Maßnahme	Schritt	beschrieben in
<i>Literaturrecherche und Interviews</i>	Identifizieren praxisrelevanter Probleme	<i>Kapitel 1</i>
<i>Recherche geeigneter Methoden</i>	Spezifizieren problemrelevanter Methoden	<i>Kapitel 2</i>
<i>Workshops mit Industrievertretern</i>	Ableiten von Forschungsproblemen und Forschungszielen	<i>Kapitel 3</i>
<i>Erfolgsfaktoren und Vergleichsmatrix</i>	Ableiten von Erfolgsfaktoren, selektieren & bewerten der Methoden	<i>Abschnitt 5.1 & 5.2</i>
<i>Implementierung der Methoden in Python</i>	Erstellen einer Pipeline zur Extrahierung von Anforderungen	<i>Abschnitt 5.3</i>
<i>Berechnung von Kennzahlen</i>	Durchführen einer Vergleichsstudie	<i>Abschnitt 5.4</i>

Leistungsfähigkeit von KI-Modellen zur Extrahierung von Anforderungen aus Regularien

Bild 1: Methodik für anwendungsbezogene Wissenschaft in Anlehnung an ULRICH [16]

5. Ergebnisse und Diskussion

In diesem Abschnitt werden die definierten Erfolgsfaktoren, die vergleichende Bewertung der Modelle, die Implementierung der Modelle in einer Pipeline in Python und die Ergebnisse der Anwendung der Methoden zur Extrahierung von Anforderungen aus UN/ECE Regularien dargestellt und diskutiert.

5.1. Definition von Erfolgsfaktoren

Ergebnis der Workshops sind fünf Erfolgsfaktoren (EF) für die Methode zur Extrahierung von Anforderungen aus Regularien (siehe Tabelle 2). Die Erfolgsfaktoren werden in Anlehnung an HAMRAZ ET AL. in die Kategorien Eingabe (E) und Modellanwendung (M) eingeteilt [17].

Die Methode muss in der Lage sein, Anforderungen in natürlicher Sprache zu verarbeiten (EF 1). Die Anforderungsbeschreibungen müssen nicht konform zu einem festen Schema oder einer Schablone sein (z. B. Vorlagen von RUPP [18]). In Regularien wird vielfach kein festes Schema bei der Dokumentation eingehalten [19, 20]. Die Sprache der Anforderungen und der Dokumente muss Englisch sein (EF 2), weil Regularien üblicherweise in englischer Sprache publiziert sind. Begriffe müssen kontextspezifisch verarbeitet werden können. Die für die Anwendung der Methode erforderlichen Daten müssen in einer praxisüblichen Weise zur Verfügung stehen und die notwendigen Trainingsdaten für das KI-Modell müssen in der Praxis in ausreichender Anzahl vorhanden sein (EF 3). Der Aufwand zur Anwendung der Methode muss niedriger als eine manuelle Extrahierung der Anforderungen sein (EF 4). Die Leistungsfähigkeit der Methode muss bei der Extrahierung von Anforderungen aus Regularien hinreichend hoch sein, damit die Anforderungen korrekt extrahiert werden (EF 5).

Tabelle 2: Erfolgsfaktoren zur Bewertung der KI-Modelle

Nr.	Kategorie	Erfolgsfaktor (EF)	Textuelle Beschreibung der Erfolgsfaktoren
1	E	Anforderungen in natürlicher Sprache	Die Anwendung muss in der Lage sein, natürlich-sprachliche Anforderungen zu extrahieren.
2	E	Verarbeitung englischer Sprache	Die Anwendung muss in der Lage sein, englische natürlich-sprachliche Dokumente zu verarbeiten.
3	M	Verfügbarkeit von Daten	Die Anwendung muss durch das Training mit einer nach Industriestandard gängigen Anzahl an Trainingsdaten eine hinreichend hohe Leistungsfähigkeit erreichen.
4	M	Aufwand der Anwendung	Der Aufwand zur Nutzung der Methode muss für den Anwender hinreichend gering sein.
5	M	Leistungsfähigkeit der Methode	Die Leistungsfähigkeit der Methode muss hinreichend hoch sein.

Legende: E – Eingabe, M – Modellanwendung

5.2. Vergleichende Bewertung der Methoden

Die fünf identifizierten KI-Modelle (SVM, LR, MNB, KNN und BERT, siehe Kapitel 2) werden anhand der fünf Erfolgsfaktoren bewertet. Die Werte pro Erfolgsfaktor werden je KI-Modell summiert, um einen Gesamtwert der Bewertung zu berechnen. Hierzu wird das folgende Bewertungsschema genutzt, um den Erfüllungsgrad der Erfolgsfaktoren zu quantifizieren:

- 0: Erfolgsfaktor ist nicht erfüllt
- 0,5: Erfolgsfaktor ist teilweise erfüllt
- 1: Erfolgsfaktor ist vollständig erfüllt

Der Erfüllungsgrad der KI-Modelle wird auf Basis der Ergebnisse der analysierten Beiträge bewertet. Durch alle Modelle werden natürlich-sprachliche Anforderungen klassifiziert. Die Modelle verarbeiten englischsprachige Textsegmente [11]. BERT bezieht den Kontext von Wörtern ein und ermöglicht eine präzisere Verarbeitung der Sprache [14]. Die in den Publikationen dokumentierten Ergebnisse zeigen, dass BERT tendenziell eine höhere Anzahl an Daten im Vergleich zu anderen Modellen benötigen, um eine hohe Leistungsfähigkeit zu erreichen [14]. Die Modelle klassifizieren automatisiert Anforderungen, weshalb der Aufwand gering ist [11]. Für die Klassifizierung von Anforderungen zeigen besonders die SVM, KNN und BERT eine hohe Leistungsfähigkeit [9–11, 13]. Das Ergebnis der Bewertung wird in Tabelle 3 dargestellt.

Tabelle 3: Vergleichsmatrix zum Vergleich repräsentativer Ansätze

Erfolgsfaktoren \ KI-Modelle	SVM	LR	MNB	KNN	BERT
Anforderungen in natürlicher Sprache	●	●	●	●	●
Reguliersprache Englisch	◐	◐	◐	◐	●
Verfügbarkeit von Daten	●	●	●	●	◐
Aufwand der Anwendung	●	●	●	●	●
Leistungsfähigkeit der Methode	●	◐	◐	●	●
Summe	4,5	4	4	4,5	4,5

Legende: ○ nicht erfüllt ◐ teilweise erfüllt ● vollständig erfüllt

Drei der fünf Modelle erzielen eine Bewertung von 4,5 von 5 Punkten: SVM, KNN und BERT. Diese Modelle werden für die Implementierung in der Pipeline ausgewählt. Ausschlaggebend für die Auswahl dieser drei Modelle ist vor allem die höhere Bewertung der Leistungsfähigkeit zur Extrahierung von Anforderungen im Vergleich zu den anderen beiden KI-Modellen.

5.3. Implementierung der Methoden

Die Pipeline zur Extrahierung von Anforderungen aus Lastenheften aus [11] wird erweitert, um Anforderungen aus Regularien mittels der drei ausgewählten KI-Modelle zu extrahieren. Die Pipeline besteht aus den Schritten „Textsegmente extrahieren“, „Textsegmente vorverarbeiten“ und „Textsegmente klassifizieren“ (siehe Bild 2).

Textsegmente extrahieren: Die Regularien werden im ersten Schritt eingelesen und es werden Textsegmente mittels PDF-Parser extrahiert. Das Dateiformat von Regularien ist üblicherweise das PDF-Format und die Dokumentensprache ist Englisch. Der Inhalt der Regularien umfasst Titel, Tabellen, Bilder und Fließtext. Texte in Bildern werden in der Erstellung von Textsegmenten vernachlässigt. Innerhalb der erkannten Textsegmente wird ein Algorithmus angewendet, der Satzenden erkennt und die Sätze trennt. Das Ergebnis aus dem ersten Schritt sind extrahierte Textsegmente.

Textsegmente vorverarbeiten: Zur Vorverarbeitung der Daten werden verschiedene Operationen durchgeführt. Es werden für jeden Satz der extrahierten Textsegmente aus dem vorherigen Schritt Tokens aus den Wörtern gebildet. Für diese Operation wird der Tokenizer von spaCy verwendet. Durch den vortrainierten Universal Sentence Encoder von Tensorflow werden die Tokens als Vektoren dargestellt, damit sie von KI-Modellen verarbeitbar sind [21]. Die Ergebnisse aus diesem Schritt sind vorverarbeitete Textsegmente.

Textsegmente klassifizieren: Die Textsegmente werden durch die KI-Modelle hinsichtlich „Anforderung“ oder „keine Anforderung“ klassifiziert. Vor der Durchführung der Klassifizierung müssen Modelle zur Klassifizierung von Anforderungen trainiert werden. Die KI-Modelle, die in diesem Beitrag betrachtet werden, werden mittels überwachten Lernens trainiert. Trainingsdaten sind gekennzeichnete Textsegmente der Regularien. Das Ergebnis aus dem dritten Schritt sind klassifizierte Anforderungen.

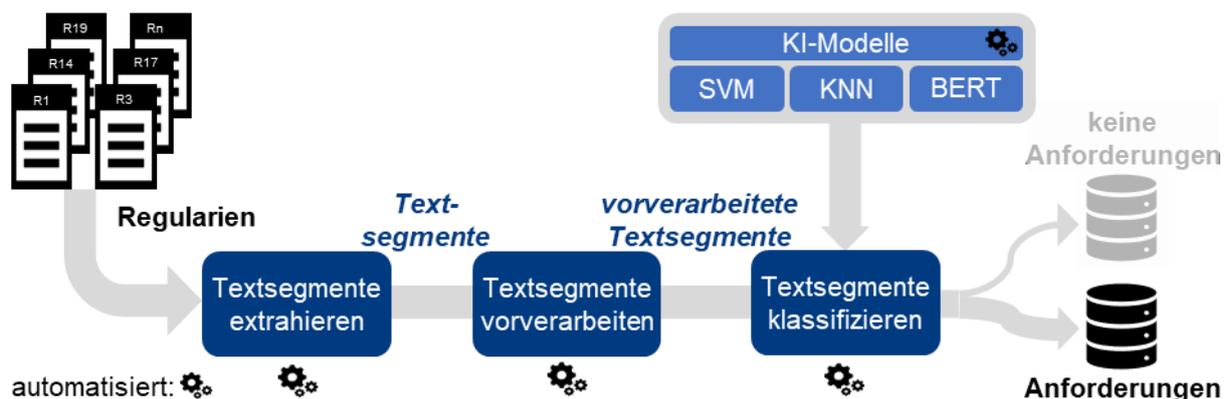


Bild 2: Pipeline mit integrierten KI-Modellen

5.4. Vergleichsstudie

Es wird eine Vergleichsstudie durchgeführt, um die KI-Modelle mit der höchsten Leistungsfähigkeit für die Extrahierung von Anforderungen aus Regularien zur Automobilentwicklung zu ermitteln. Hierzu wird die im vorherigen Abschnitt beschriebene Pipeline genutzt. Zur Generierung von Trainings- und Testdaten für die Vergleichsstudie werden 1032

Textsegmente aus den Regularien UN/ECE R11 und R12 gekennzeichnet [5, 22]. Diese Regularien werden in Lastenheften besonders häufig referenziert, wie die Ergebnisse aus den Workshops mit Ingenieuren des Automobilherstellers zeigen. Es werden die ausgewählten KI-Modelle SVM, KNN und BERT mit den gekennzeichneten Textsegmenten trainiert und getestet (Verhältnis 80/20). Die Klassen-Aufteilung der Daten ist in Tabelle 4 dargestellt.

Tabelle 4: Klassen-Aufteilung der Daten für die Vergleichsstudie

Klassen	Anzahl Daten
keine Anforderung	416
Anforderung	616
Summe	1032

Kriterien für den Vergleich sind die Korrektklassifikationsrate (engl. „Accuracy“), die Präzision (engl. „Precision“), die Sensitivität (engl. „Recall“) und der F1-Wert. Die Korrektklassifikationsrate gibt den Anteil aller Ergebnisse an, die korrekt klassifiziert werden. Die Präzision gibt die Anzahl der vorhergesagten true-positives im Vergleich zu den richtig und false-positives an. Eine Präzision nahe bei 100 % weist auf zuverlässige Modelle hin. Die Sensitivität gibt den Anteil der true-positive klassifizierten Ergebnisse an den gesamten positiven Ergebnissen an. Der F1-Wert ist das harmonische Mittel aus Präzision und Sensitivität und ermöglicht die initiale Bewertung der Gesamtqualität des Modells. [23]

Um Ausreißer in den Ergebnissen durch die Festlegung von Trainings- und Testsets zu vermeiden, werden fünf Trainings- und Testsets aus den Daten zufällig gebildet und die Modelle werden jeweils mit den Sets trainiert bzw. getestet und Kennzahlen berechnet. Es wird das arithmetische Mittel der Kennzahlen berechnet. Die Ergebnisse der Vergleichsstudie werden in Tabelle 5 dargestellt.

Tabelle 5: Ergebnisse der Vergleichsstudie

KI-Modelle	KK	Präzision	Sensitivität	F1-Wert
SVM	95,28 %	96,85 %	95,35 %	96,09 %
KNN	95,05 %	96,08 %	95,12 %	95,59 %
BERT	90,09 %	92,19 %	91,47 %	91,83 %

Legende: KK – Korrektklassifikationsrate, SVM – Support Vector Machine, KNN – Künstliches Neuronales Netz, BERT – Bidirectional Encoder Representations from Transformers

Die SVM erzielt die höchste Leistungsfähigkeit, da die Kennzahlen (siehe Tabelle 5) höher sind als die Kennzahlen der anderen KI-Modelle. Das KNN erzielt eine fast identisch hohe Leistungsfähigkeit mit seinen Kennzahlen, diese sind relativ um 0,45 % niedriger als die Kennzahlen der SVM. Die Kennzahlen zur Leistungsfähigkeit von BERT sind im Durchschnitt um 4,69 % niedriger als die Kennzahlen der Leistungsfähigkeit der SVM. Somit ist die Leistungsfähigkeit von BERT niedriger als die Leistungsfähigkeit des KNN. Eine Ursache für die niedrigere Leistungsfähigkeit von BERT ist, dass BERT im Vergleich zu den anderen beiden Modellen tendenziell eine höhere Anzahl von Trainingsdaten benötigt, um eine hohe Leistungsfähigkeit zu erreichen (siehe Abschnitt 5.2). Aufgrund der höheren Leistungsfähigkeit der SVM und des KNN zeigen die Ergebnisse, dass die beiden Modelle für die automatisierte Extrahierung von Anforderungen aus Regularien besonders geeignet sind.

Die Ergebnisse zeigen, dass die KI-Modelle Anforderungen automatisiert mit einer hohen Leistungsfähigkeit aus Regularien extrahieren. Die durch die Pipeline extrahierten

Anforderungen können in ein Tool zum Anforderungsmanagement überführt werden und verringern den Aufwand für die manuelle Extrahierung von Anforderungen.

6. Zusammenfassung und Ausblick

In diesem Beitrag werden Methoden zur automatisierten Extrahierung von Anforderungen aus Regularien identifiziert und verglichen. Zur Identifikation von Modellen wird eine systematische Literaturanalyse durchgeführt und verschiedene KI-Modelle ermittelt, die einen hohen Automatisierungsgrad zur Verarbeitung natürlich-sprachlicher Texte besitzen. Es werden Erfolgsfaktoren definiert, um die KI-Modelle für eine umfangreiche Vergleichsstudie initial zu selektieren. Erfolgsfaktoren sind z. B. der Aufwand zur Anwendung oder die Leistungsfähigkeit der Modelle. Die KI-Modelle SVM, KNN und BERT werden aufgrund der Bewertung für die Vergleichsstudie ausgewählt. Es wird eine Pipeline in Python implementiert, um die drei KI-Modelle anzuwenden und zu vergleichen. Innerhalb der Vergleichsstudie werden Textsegmente aus Regularien hinsichtlich „Anforderung“ oder „keine Anforderung“ klassifiziert. Es werden Kennzahlen berechnet, um die Leistungsfähigkeit der Modelle zu bestimmen. Die Ergebnisse zeigen, dass die SVM und das KNN die höchste Leistungsfähigkeit zur automatisierten Extrahierung von Anforderungen aus Regularien besitzen.

Die Anwendung der KI-Modelle ist ein „Enabler“ für die Reduzierung von Aufwand in der Extrahierung von Anforderungen. Aufbauend auf dieser Forschung muss untersucht werden, inwiefern Entwickler die klassifizierten Anforderungen in der Praxis nutzen können, um effizienter zu entwickeln, indem sie nicht mehr Querverweise in Lastenheften nachgehen müssen. Durch die ressourcenschonende Anforderungsextrahierung können sie Kapazitäten für die Ermittlung und die Sicherstellung der Erfüllung funktionaler Anforderungen aufbringen. Die praxisnahe Untersuchung zeigt den Handlungsbedarf zur Erweiterung der Methoden-anwendung für die industrielle Anwendung im Requirements Engineering für komplexe technische Systeme [24, 25] auf.

Danksagung

Die Autoren danken Florian Hölscher für die Unterstützung in der Auswertung der Anforderungsdaten für die Vergleichsstudie und Kennzeichnung von Regularien der Automobilentwicklung der UN/ECE Transport.

Literatur

- [1] Gräßler, I.; Wiechel, D.; Koch, A.-S.; Preuß, D.; Oleff, C.: “Model-based effect-chain analysis for complex systems”. In: Design Society (Hrsg.), *Proceedings of the Design Society*, Cambridge University Press, Cambridge, UK, 2022.
- [2] Gräßler, I.; Preuß, D.: „Automatisierte Abhängigkeitsanalyse von Anforderungen zur Wirkkettenmodellierung“. In: *Stuttgarter Symposium für Produktentwicklung 2023*, Fraunhofer-Institut für Arbeitswirtschaft und Organisation IAO, Stuttgart, 2023.
- [3] Gräßler, I.; Wiechel, D.; Koch, A.-S.; Sturm, T.; Markfelder, T.: “Methodology for Certification-Compliant Effect-Chain Modeling”. In: *Systems*, Jg. 11, Nr. 3, 2023, S. 154.
- [4] Gräßler, I.: “Competitive Engineering in the Age of Industry 4.0 and Beyond”. In: Horvath, I.; Suarez Rivero, J. P.; Hernandez Castellano, P. M. (Hrsg.), *Proceedings of the 12th International Symposium on Tools and Methods of Competitive Engineering (TMCE)*, 2018, S. 213-232.
- [5] UN/ECE: „Einheitliche Bedingungen für die Genehmigung der Kraftfahrzeuge hinsichtlich der Türverschlüsse und Türabhängungen, Nr. R11, 2010.
- [6] Fernandes, J. M.; Machado, R. J.: “Requirements in Engineering Projects”. Springer International Publishing; Springer, Cham, 2016.
- [7] Janpitak, N.; Sathitwiriwong, C.; Pipatthanaudomdee, P.: “Information Security Requirement Extraction from Regulatory Documents using GATE/ANNIC”. In: *2019 7th International Electrical Engineering Congress (iEECON)*, IEEE, 2019.
- [8] Machi, L. A.; McEvoy, B. T.: “The literature review”. Corwin, Thousand Oaks, California, 2016.

-
- [9] Kurtanovic, Z.; Maalej, W.: "Automatically Classifying Functional and Non-functional Requirements Using Supervised Machine Learning". In: *2017 IEEE 25th International Requirements Engineering Conference (RE)*, 2017, S. 490-495.
- [10] Boutkova, E.; Houdek, F.: "Semi-automatic identification of features in requirement specifications". In: *19th IEEE International Requirements Engineering Conference (RE), 2011*, IEEE, Piscataway, NJ, 2011, S. 313-318.
- [11] Gräßler, I.; Preuß, D.; Brandt, L.; Mohr, M.: "Efficient Extraction of Technical Requirements Applying Data Augmentation". In: *Proceedings of 8th IEEE International Symposium on Systems Engineering 2022*, Wien, 2022.
- [12] Zeni, N.; Mich, L.: "Usability issues for systems supporting requirements extraction from legal documents". In: *2014 IEEE 7th International Workshop on Requirements Engineering and Law (RELAW)*, IEEE, 2014, S. 35-38.
- [13] Hassan, F. u.; Le, T.: "Automated Requirements Identification from Construction Contract Documents Using Natural Language Processing". In: *Journal of Legal Affairs and Dispute Resolution in Engineering and Construction*, Jg. 12, Nr. 2". American Society of Civil Engineers, 2020, S. 4520009.
- [14] Matzka, S.: „Künstliche Intelligenz in den Ingenieurwissenschaften“. Springer Fachmedien Wiesbaden GmbH; Springer Vieweg, Wiesbaden, 2021.
- [15] Gräßler, I.; Preuß, D.; Pottebaum, J.: „Extrahierung von Anforderungen aus natürlich-sprachlichen Lastenheften: Was erschwert eine KI-basierte Extrahierung?“. In: *Softwaretechnik-Trends*, Jg. 42, Nr. 1, 2022.
- [16] Ulrich, H.: „Anwendungsorientierte Wissenschaft“. In: *Die Unternehmung*, Jg. 36, 1982, S. 1-10.
- [17] Hamraz, B.: "Engineering change modelling using a function-behaviour-structure scheme". Apollo - University of Cambridge Repository, 2013.
- [18] Rupp, C.: "MASTeR - Schablonen für alle Fälle". SOPHIST GmbH, Nürnberg, 2016.
- [19] Gräßler, I.; Hentze, J.; Yang, X.: "Eleven Potentials for Mechatronic V-Model". In: Villmer, F.-J.; Padoanao, E. (Hrsg.), *Production Engineering and Management*, 2016, S. 257-268.
- [20] Gräßler, I.; Dattner, M.; Bothen, M.: "Main Feature List as core success criteria of organizing Requirements Elicitation". In: RADMA (Hrsg.), *R&D Management Conference 2018*, 2018, S. 1-16.
- [21] Cer, D.; Yang, Y.; Kong, S.; Hua, N.; Limtiaco, N.; St. John, R.; Constant, N.; Guajardo-Cespedes, M.; Yuan, S.; Tar, C.; Sung, Y.-H.; Strope, B.; Kurzweil, R.: "Universal Sentence Encoder, 2018.
- [22] UN/ECE: „Einheitliche Bedingungen für die Genehmigung der Kraftfahrzeuge hinsichtlich des Schutzes des Fahrzeugführers vor der Lenkanlage bei einem Aufprall, Nr. R12, 2012.
- [23] Alpaydin, E.: "Maschinelles Lernen". De Gruyter Oldenbourg, München, 2022.
- [24] Gräßler, I.; Tusek, A. M.; Özcan, D.: "Entwicklung zukunftsfähiger Geschäftsmodelle in der Automatisierungstechnik anhand von Megatrends". In: 23. VDI Kongress AUTOMATION, 2023.
- [25] Gräßler, I.; Pottebaum, J.; Scholle, P.: "Integrated Process and Data Model for Agile Strategic Planning". In: Vajna, S. (Hrsg.), *11th International Workshop on Integrated Design Engineering*, 2017.