

Bachelorarbeit

...

von

Ricardo Valente de Matos

Matrikelnummer: 7203677

im Studiengang Wirtschaftsinformatik
der Fachhochschule Dortmund

Erstprüfer: Prof. Dr.-Ing. Guy Vollmer

Zweitprüfer: Stephan Schmeißer, M. Sc., Adessoplatz 1, 44269 Dortmund

Dortmund, den 5. Februar 2024

Motivation

Die allgemeine Bewusstheit von KI-gestützten Systemen ist in den letzten Jahren vor allem durch die Verbreitung von Large Language Models wie Chat-GPT gestiegen [4]. Sie können komplexe Fragen beantworten und kurze Texte zu Themen verfassen [4].

Das Thema Künstliche Intelligenz wird auch in der Europäischen Kommission diskutiert. Der AI Act behandelt Aspekte der Sicherheit und des Vertrauens bei der Nutzung von KI-Systemen [8].

„Um einen Beitrag zum Aufbau eines widerstandsfähigen Europas für die digitale Dekade zu leisten, sollten Menschen und Unternehmen in der Lage sein, die Vorteile von KI zu nutzen und sich gleichzeitig sicher und geschützt zu fühlen.“[10]

Das Gesetz behandelt einen Ansatz für vertrauenswürdige KI. Es werden Anforderungen gestellt, die neben Aspekten wie der Risikominderung und Qualität der Daten auch eine hohe Robustheit und Genauigkeit sicherstellen sollen [9].

Das Potenzial von KI wurde auch bei der Einbindung in Unternehmen entdeckt. Eine KI kann Arbeiten übernehmen, die für Beschäftigte eine Entlastung bedeuten können [14]. *adesso* hat auch das Potenzial erkannt und sucht nun nach und nach Wege, KI-gestützte Systeme in die eigenen Prozesse zu integrieren. Im internen Projekt *adesso Staffing Advisor* wird an einem Recommender-System zur Mitarbeiterempfehlung für ausgewählte Projekte gearbeitet. Die Umsetzung der Recommender Systems bedient sich verschiedener KI-basierten Ansätze. Ein ganz entscheidender Schritt im Prozess der Mitarbeiterempfehlung ist die Vorverarbeitung der Bedarfsmeldungen. Diese sind eine wertvolle Informationsquelle, die Fachkräften helfen kann, die Empfehlungen effizienter zu gestalten, um dadurch wettbewerbsfähig zu bleiben. Allerdings sind diese oft umfangreich und komplex, was ihre effektive Nutzung erschwert.

Deshalb ist es entscheidend, effiziente Methoden und Techniken des Information Retrieval anzuwenden, um so relevante Informationen schnell und präzise aus Bedarfsmeldungen zu extrahieren. Die Extraktion wichtiger Schlüsselwörter, Phrasen und Themen ermöglicht es einen besseren Einblick in die Ziele, Methoden und Ergebnisse der Projekte zu bekommen. Dadurch können fundierte Entscheidungen bezüglich der Personalbesetzung getroffen und Ressourcen effizient genutzt werden.

Problemstellung

In einer immer stärker vernetzten und informationsreichen Welt stehen Organisationen vor der Herausforderung, relevante Informationen effizient aus umfangreichen Bedarfsmeldungen zu extrahieren. Obwohl diese Beschreibungen wichtige Einblicke in Ziele, Methoden und Ergebnisse liefern, können sie aufgrund ihres Umfangs und ihrer Komplexität schwer durchsuchbar und analysierbar sein. Die manuelle Identifizierung und Extraktion relevanter Inhalte ist zeitaufwendig und fehleranfällig. Daher stellt sich die Problemstellung:

Wie können wir effektive Methoden und Techniken des Information Retrieval und Preprocessing nutzen, um automatisiert relevante Inhalte aus Bedarfsmeldungen im spezifischen Software Entwicklungs-Kontext zu extrahieren und somit die Effizienz, Genauigkeit und Geschwindigkeit der Informationsgewinnung für Führungskräfte zu verbessern.

Ziele und Ergebnisse der Arbeit

Diese Ausarbeitung präsentiert eine umfassende Untersuchung zur Entwicklung eines automatisierten Systems zur Extraktion relevanter Inhalte aus Bedarfsmeldungen im Software-Entwicklungs-Kontext.

- Die erste Phase dieser Ausarbeitung besteht darin, eine klare Erwartungshaltung hinsichtlich der Anforderungen und Bedürfnisse der Stakeholder zu entwickeln. Hierfür werden Interviews mit Führungskräften durchgeführt, um die Erwartungen bezüglich einer „perfekten“ Bedarfsmeldung herauszuarbeiten. Diese dient als Grundlage für die weiteren Entwicklungs- und Evaluierungsphasen.
- Im Anschluss erfolgt eine eingehende Analyse der verschiedenen Techniken des Information Retrieval wie *probabilistic retrieval* oder *vector space approach*, um die besten Ansätze zur Extraktion relevanter Inhalte zu identifizieren. Diese Analyse bildet die Grundlage für die Konzeptionierung eines Preprocessing-Modells, das eine Kombination der erforschten Ergebnisse darstellt. Die Implementierung dieses Modells erfolgt durch den Aufbau einer (RapidMiner/Python)-Pipeline, die eine effiziente Verarbeitung und Extraktion der Projektbeschreibungen ermöglicht.
- Zur Evaluierung der Leistungsfähigkeit des entwickelten Systems werden reale Bedarfsmeldungen aus JIRA verwendet. Dabei wird überprüft, inwiefern das Ergebnis der definierten Erwartungshaltung entspricht. Mithilfe von Metriken wie Precision, Recall und F1-Score werden Abweichungen, Ähnlichkeiten und verschiedene Parameter analysiert, um Erkenntnisse darüber zu gewinnen, wie das System inhaltlich abschnidet und verbessert werden kann.
- (Schließlich wird eine vergleichende Untersuchung mit einem auf Large Language Model basierenden Vorverarbeitungsansatz durchgeführt. Dabei werden die Performance, Zeit und Ergebnisqualität des entwickelten Systems mit diesem alternativen Ansatz verglichen. Dieser Vergleich dient dazu, die Stärken und Schwächen des entwickelten Systems zu identifizieren und gegebenenfalls weitere Verbesserungen vorzunehmen.)

Vorgehen und Zeitplan

Ziel ist es die Arbeit im Mai fertig zu stellen. Die einzelnen Monatsziele können aus der nachfolgenden Tabelle entnommen werden.

Februar	<ul style="list-style-type: none">• Durchführung der Interviews mit Fachkräften• Zusammentragung aller relevanter Information Retrieval- und Preprocessing-Ansätze
März	<ul style="list-style-type: none">• Durchführung der Interviews mit Fachkräften• Formulierung der Anforderungen für Bedarfsmeldungen
April	<ul style="list-style-type: none">• Entwicklung des Eigenen Preprocessing-Modells• Evaluierung der Ergebnisse
Mai	<ul style="list-style-type: none">• Schluss schreiben• Korrekturen

ToDo: Aufbau der Arbeit anpassen

<-hier