

**Thesis zur Erlangung des akademischen Grades
Bachelor of Science (B. Sc.)**

**AUTOMATISIERUNG DER
INFORMATIONSGEWINNUNG IN
BEDARFSMELDUNGEN**

von

Ricardo Valente de Matos

geboren am 30.10.1999

Matrikelnummer: 7203677

im Studiengang Wirtschaftsinformatik

der Fachhochschule Dortmund

im Fachbereich Informatik

Erstprüfer: Prof. Dr.-Ing. Guy Vollmer

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

Dortmund, den 7. Mai 2024

Abstract

ToDo: Abstract erstellen

<-hier

Lesehinweis

Aus Gründen der besseren Lesbarkeit werden Wörter und Wortgruppen, die hervorgehoben werden oder mehrfach auftauchen, durch *kursiven* Text kenntlich gemacht. Zudem wird in dieser Ausarbeitung die Sprachform des generischen Maskulinums angewandt. Sämtliche Ausführungen sind jedoch geschlechtsunabhängig und beziehen sich damit auf alle Geschlechter.

Inhaltsverzeichnis

1	Einleitung	1
1.1	Problemstellung	2
1.2	Ziele und Ergebnisse der Arbeit	2
1.3	Aufbau der Arbeit	3
2	Grundlagen	5
2.1	Kontext	5
2.2	keine Ahnung	5
3	Literaturüberblick	7
3.1	Recommender Systems Historie und aktueller Stand der Forschung . .	7
3.2	Verwandte Arbeiten	9
3.2.1	kp wie ich es nenne	9
3.2.2	Information Filtering	10
3.2.3	Vorverarbeitung	10
3.2.4	Hybride Ansätze	11
3.2.5	Pipeline	12
3.3	Definitionen und Konzepte	13
3.3.1	Information Retrieval und Information Filtering	13
3.3.2	Data-Mining	13
3.3.3	Bedarfsmeldungen	13
4	Entwicklung einer klaren Erwartungshaltung	15
4.1	Beschreibung der Interviews mit Führungskräften zur Identifizierung von Stakeholder-Erwartungen	15
4.2	Ablauf der Interviews	16
4.3	Übersicht der Experten	17
4.4	Analyse der Ergebnisse und Entwicklung einer klaren Erwartungshal- tung für die Bedarfsmeldungen	18
5	Analyse der Techniken des Information Retrieval und Data-Mining	22
5.1	Beschreibung der untersuchten Techniken und Ansätze	22
5.1.1	TF-IDF	22
5.1.2	Text-Ranking-Algorithmen	23
5.1.3	N-Gramm	23

5.1.4	POS-Tagging	23
5.1.5	Named Entity Recognition	24
5.1.6	Regelbasierte Ansätze	24
5.1.7	Hybride Ansätze	24
5.1.8	Data-Mining	25
5.1.9	preprocessing	25
5.1.10	data-fusion	25
5.2	Bewertung und Auswahl der besten Ansätze für die Extraktion relevanter Inhalte aus Bedarfsmeldungen	29
6	Konzeptionierung und Implementierung der Vorverarbeitung	31
6.1	Beschreibung des entwickelten Vorverarbeitungsmodells	31
6.2	Details zur Implementierung der Pipeline in Python	36
7	Evaluierung des entwickelten Systems	38
7.1	Beschreibung des verwendeten Datensatzes und der Evaluierungsmethodik	38
7.2	Präsentation und Diskussion der Ergebnisse	39
7.3	Vergleich des Systems mit einem Large Language Model-Ansatz	45
7.4	Analyse von Abweichungen, Ähnlichkeiten und Verbesserungspotenzialen des Systems	49
8	Zusammenfassung und Ausblick	51
9	Anhang	52
9.1	Interviewtranskripte	52
9.1.1	Vorabtest Fragen und Antworten	52
9.1.2	Interview 2	53
9.1.3	Interview 3	56
9.2	Erklärung zu eingesetzten Hilfsmitteln	65

Abbildungsverzeichnis

Listings

1 Einleitung

In einer globalisierten und dynamischen Wirtschaftswelt sind Unternehmen zunehmend auf Projekte angewiesen, um ihre Ziele zu erreichen und Wettbewerbsvorteile zu erlangen. Die Personalbeschaffung für solche Projekte erfordert oft spezialisiertes Fachwissen und vielfältige Fähigkeiten, um erfolgreich umgesetzt zu werden. Es ist entscheidend für den Projekterfolg, dass die Personalbeschaffung die passenden Mitarbeiter für ausgewählte Projekte findet. Hier setzt die Entwicklung eines Recommender Systems zur Mitarbeiterempfehlung an. Ein solches System kann Unternehmen dabei unterstützen, den Prozess der Mitarbeiterrekrutierung und -auswahl zu optimieren. Durch die Berücksichtigung verschiedener Kriterien wie Qualifikationen, Fähigkeiten und Erfahrungen kann das Recommender-System dazu beitragen, die Auswahl effektiv zu filtern und diejenigen herauszufiltern, die am besten zu einem Projekt im Unternehmen passen. Ein solches System bietet außerdem den Vorteil, den Prozess der Mitarbeiterempfehlung zu automatisieren und zu beschleunigen. Dies ermöglicht Unternehmen, schneller auf offene Stellen zu reagieren und potenzielle Kandidaten zeitnah zu identifizieren. Dadurch wird die Effizienz der Mitarbeitersuche verbessert und die Qualität der Einstellungsentscheidungen erhöht.

Das Potenzial von Recommender Systems wurde auch bei *adesso* entdeckt und nun wird nach und nach Wege gesucht, 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 Führungskräften helfen kann, die Empfehlungen effizienter zu gestalten, um dadurch wettbewerbsfähig zu bleiben. Allerdings sind diese oft umfangreich, unsortiert 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.

1.1 Problemstellung

Um das Entlastungspotenzial für Führungskräfte durch das Gesamtsystem eines Recommender Systems für Mitarbeiterempfehlungen zu realisieren, sind mehrere Schritte notwendig. Eine Informationsgewinnung aus den unstrukturierten Projekt- und Mitarbeiterdaten ist unerlässlich, um schließlich den Ähnlichkeitsvergleich für die Empfehlungen durchführen zu können. Diese Ausarbeitung befasst sich mit dem ersten Schritt der Strukturierung und Informationsextraktion der vorhandenen *Bedarfsmeldungen*. Somit steht *adesso* 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 Data-Mining 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.

In der Vergangenheit wurden bereits Methoden im Bereich des automatisierten Recruitings untersucht. Im Projektgeschäft sehen wir uns mit einem Problem konfrontiert, dessen Umfang jedoch präziser definiert werden kann, da die Kandidatenauswahl einem begrenzten Pool unterliegt. Besondere Relevanz hat hierbei die Erstellung einer Standardisierung der *Bedarfsmeldung*, da diese häufig unstrukturiert und mit fehlenden Informationen vorliegt.

1.2 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.

- In der Ausarbeitung wird zunächst ein Konzept einer standardisierten *Bedarfsmeldung* erarbeitet. Dazu wird eine klare Erwartungshaltung hinsichtlich der Anforderungen und Bedürfnisse der Stakeholder entwickeln. Hierfür werden Interviews mit Führungskräften durchgeführt, um die Erwartungen bezüglich

einer „perfekten“ *Bedarfmeldung* herauszuarbeiten. Dieses Konzept dient als Grundlage für die weiteren Entwicklungs- und Evaluierungsphasen.

- Es wird an einer ausführbaren prototypischen Software gearbeitet, die *Bedarfmeldungen* effizient verarbeitet und wichtige Informationen extrahiert. Hierfür wird eine Pipeline in Python aufgebaut und strukturell durch Use-Case- und UML-Diagramme dokumentiert. Es werden Modelle des Information Retrieval und Data-Mining implementiert. Dabei erfolgt zunächst eine eingehende Analyse der Techniken *TF-IDF*, *Text-Ranking-Algorithmen*, *N-Gramm-Analyse*, *POS-Tagging*, *Named Entity Recognition*, Regelbasierte Ansätze und Hybride Ansätze, um die besten Ansätze zur Extraktion relevanter Inhalte zu identifizieren. Diese Analyse bildet die Grundlage für die Konzeptionierung des Software-Prototypen, das eine Kombination der erforschten Ergebnisse darstellt.
- Um die Leistungsfähigkeit des entwickelten Systems zu evaluieren, werden Testfälle für reale *Bedarfmeldungen* definiert. Dabei wird überprüft, inwieweit das Ergebnis den Erwartungen entspricht. Mit Hilfe einer manuellen Überprüfung werden Abweichungen, Ähnlichkeiten und Anpassungen analysiert, um Erkenntnisse über die inhaltliche Leistung des Systems und die Techniken zu gewinnen, die allein oder in Kombination mit mehreren Ansätzen die wichtigsten Informationen herausfiltern. Da die Dauer eine entscheidende Rolle spielt, werden auch Zeit und Leistung gemessen. Diese Ergebnisse werden mit einem neuen Vorverarbeitungsansatz verglichen, der auf dem Large Language Model basiert. Die Performance, Zeit und Ergebnisqualität des entwickelten Systems soll im Vergleich mit diesem alternativen Ansatz die Stärken und Schwächen des entwickelten Systems aufzeigen, um daraus gegebenenfalls weitere Verbesserungsmöglichkeiten zu identifizieren.

1.3 Aufbau der Arbeit

ToDo: Aufbau der Arbeit erstellen

<-hier

gg

2 Grundlagen

2.1 Kontext

Als IT-Dienstleister wird adesso von Kunden unter anderem mit der Entwicklung individueller Softwarelösungen beauftragt. Derzeit verbringen Führungskräfte jedoch viel Zeit damit, interne Mitarbeiterinnen und Mitarbeiter manuell für Kundenprojekte zu suchen und diese dann aufgrund ihrer Erfahrungen und Fähigkeiten auszuwählen und entsprechend einzusetzen. Dieser Prozess soll durch eine KI-Lösung unterstützt werden. Da es sich bei der Personalsuche um einen geschäftskritischen Prozess handelt, ist der Spielraum für Fehler gering. Im internen Projekt adesso Staffing Advisor wird eine durch Large Language Model-gestützte Anwendung entwickelt, die Führungskräfte bei der Suche nach geeignetem Personal für ausgewählte Projekte unterstützt. Der Ansatz des Large Language Modeling ist jedoch nicht deterministisch. Es besteht die Gefahr, dass bei gleichem Input unterschiedliche Ergebnisse erzielt werden. Daher versucht adesso durch den Einsatz von Methoden und Technologien neben dem Large Language Model-Ansatz deterministische Ergebnisse zu erzielen, die auf einem ähnlichen Niveau liegen.

Die vorliegende Ausarbeitung befasst sich mit der Informationsgewinnung in Bedarfsmeldungen. Eine Bedarfsmeldung bezeichnet eine Projektbeschreibung eines Kunden, die Anforderungen an ein zu entwickelndes System enthält. Die Erstellung der Bedarfsmeldung erfolgt durch den Kunden, wobei eine gemeinsame Abstimmung mit adesso zur Finalisierung und Speicherung im JIRA erfolgt. Die Informationen sind insofern unstrukturiert, als dass sie ohne vordefinierte Struktur in Form eines Volltexts vorliegen. Infolgedessen kann es zu Abweichungen hinsichtlich der Ausgestaltung von Bedarfsmeldungen kommen.

2.2 keine Ahnung

Die Problemstellung umfasst eine Reihe von Punkten, die im Rahmen der Ausarbeitung zu behandeln sind. Aufgrund der unstrukturierten und mit fehlenden Informationen versehenen Bedarfsmeldungen ist eine Standardisierung dieser Bedarfsmeldungen von besonderer Relevanz. Dies würde die Extraktion relevanter Informa-

tionen erleichtern und somit die Effizienz des Systems verbessern. Zudem würde ein solcher Ansatz einen Einblick in die Relevanz von Informationen und Stichpunkten geben. Im Rahmen der weiteren Bearbeitung einer Standardisierung ist die Extraktion der erforderlichen Informationen aus dem Volltext erforderlich. In diesem Kontext existiert bereits eine Reihe an Methoden und Ansätzen, die sich in der Forschung bewährt haben. Der unstrukturierte Volltext muss in eine strukturierte inhaltliche Aufteilung in einzelne Sequenzen und Stichpunkte überführt werden.

In der Literatur finden sich verschiedene Ansätze zur Extraktion wichtiger Stichpunkte aus einem Volltext. Eine Methode zur Ermittlung wichtiger Stichpunkte in Texten stellt die *tf-idf-Methode* (Term Frequency-Inverse Document Frequency) dar. Innerhalb einer oder mehrerer Bedarfsmeldungen lassen sich mit dieser Methode häufig auftauchende Wörter ermitteln. Des Weiteren können graphenbasierte Methoden wie *TextRank* oder *YAKE* (Yet Another Keyword Extractor) zur Identifizierung von Schlüsselwörtern in Texten herangezogen werden. Ein weiterer Ansatz ist die Nutzung von *N-Grammen*, die häufig vorkommende Phrasen oder Begriffe identifizieren können. Auch grammatische Kategorien von Wörtern können Rückschlüsse auf potenzielle Schlüsselwörter zulassen. Die *POS-Tagging-Methode* (Part-of-Speech-Tagging) stellt eine Möglichkeit dar, um dieses Ziel zu erreichen. Des Weiteren kann *NER* (Named Entity Recognition) dazu beitragen, Personen, Firmennamen, Orte, Ereignisse oder Zeitangaben zu identifizieren. Auch wenn die Bedarfsmeldungen unstrukturiert sind, haben sich im Laufe der Zeit Konventionen entwickelt, die teilweise Strukturen eines *Patterns* aufweisen können.

3 Literaturüberblick

Das Ziel dieser Arbeit ist die Informationsgewinnung aus semistrukturierten *Bedarfmeldungen* für ein Recommender System, das Mitarbeiterempfehlungen innerhalb von *adesso* für ausgewählte Projekte generieren soll. In diesem Kapitel werden die für das Thema notwendigen Grundlagen und bereits erforschten Themengebiete im Kontext von Recommender Systemen und Informationsverarbeitung behandelt, die für das weitere Verständnis der Arbeit notwendig sind. Es wird ein Einblick in die Art und Weise gegeben, wie andere Autoren Information Retrieval und Filtering einsetzen und kombinieren.

3.1 Recommender Systems Historie und aktueller Stand der Forschung

Auch wenn die Erstellung eines Recommender Systems nicht Gegenstand der vorliegenden Ausarbeitung ist, stellt die Nutzung von Information Retrieval und Filtering ein entscheidender Schritt in Richtung eines funktionierenden Recommender Systems dar. Das Verständnis der Funktionsweise eines Recommender Systems sowie dessen Entwicklung in den vergangenen Jahren ist daher für das Verständnis des Teilbereichs dieser Thematik von Nutzen.

Recommender Systems existieren bereits seit vielen Jahren. Im Jahr 1992 führten Belkin und Croft eine Analyse und einen Vergleich des Information Retrievals und Filtering durch [10]. Das Information Retrieval behandelt dahingehend die grundlegende Technologie der Suchmaschine [10]. Das Recommender System basiert hauptsächlich auf der Technologie des Information Filtering. Im selben Jahr präsentierte Goldberg das Tapestry-System, welches das erste System zur Informationsfilterung darstellt, das auf kollaboratives Filtern durch menschliche Bewertung basiert. Die Mehrheit der frühen Empfehlungsmodelle basiert auf kollaborativer Empfehlungen, wobei K-Nearest-Neighbor (KNN)-Modelle eine besondere Rolle einnehmen. Diese Modelle prognostizieren die Nachbarn eines Zielnutzers, indem sie eine Ähnlichkeit zwischen den vorherigen Präferenzen und den Präferenzen der anderen Nutzer berechnen [10]. Die Studie von Goldberg inspirierte einige Forscher des Massachusetts Institute of Technology (MIT) und der University of Minnesota (UMN) dazu, einen Nachrichtenempfehlungsdienst mit dem Namen *GroupLens* zu entwickeln.

Die Hauptkomponente dieses Dienstes ist ein Modell zur kollaborativen Filterung zwischen Nutzern [10]. Das gleichnamige Forschungslabor kann somit als Pionier auf dem Gebiet der Recommender Systems bezeichnet werden. Die dort durchgeführten Forschungen bilden die Grundlage für nachfolgende Musik- und Video-Ähnlichkeitsempfehlungen [10].

Recommender Systeme haben in den letzten Jahren verschiedene Definitionen erhalten. Eine dieser Definitionen wird in dem Artikel von Resnick und Varian (1997) sinngemäß so beschrieben, dass ein typisches Recommender System Empfehlungen durch Personen als Eingabe erhält, die das System dann zusammenschließt und an geeignete Empfänger weiterleitet [5]. In einigen Fällen besteht die primäre Transformation in der Zusammenführung, in anderen Fällen liegt die Fähigkeit des Systems darin, gute Übereinstimmungen zwischen Empfehlungsgebern und Empfehlungsempfängern herzustellen [5]. Empfehlungssysteme stellen ein Instrument zur Interaktion mit umfangreichen und vielschichtigen Informationen dar. Sie ermöglichen eine personalisierte Sicht auf diese Informationen, indem sie die für den Nutzer wahrscheinlich relevanten Inhalte aufbereiten [5]. Besonders im Handelsverkehr im Internet sind Recommender Systeme ein häufiger Einsatzgebiet. Dabei werden Recommender Systeme als Werkzeuge zum Suchen und Filtern von Informationen verwendet, die dem Benutzer Vorschläge unterbreiten, die für ihn nützlich sein könnten. Sie sind in einer Vielzahl von Internetanwendungen weit verbreitet und helfen den Nutzern, bessere Entscheidungen bei der Suche nach Nachrichten, Musik, Urlaubsangeboten oder Geldanlagen zu treffen [34]. Ein spezifisches Recommender System konzentriert sich normalerweise auf eine Art von Themengebiet wie z. B. Filme oder Nachrichten [34]. Darüber hinaus sind sie zu einem entscheidenden Faktor in der Entscheidungsfindung von Organisationen geworden [6]. Unternehmen wie *adesso* bauen immer weiter auf Recommender System unterstützte System auf, um Prozesse zu beschleunigen oder zu vereinfachen.

Grundsätzlich können die Methoden in vier Typen unterteilt werden:

- collaborative Filtering-based (kollaborative Empfehlungssysteme)
- content-based (inhaltsbasierte Empfehlungssysteme)
- knowledge-based (wissensbasiert Empfehlungssysteme)
- hybrid (hybride Empfehlungssysteme)

Jede Empfehlungsmethode hat ihre Vorteile und Grenzen [22]. Insbesondere das inhaltsbasierte Empfehlungssystem bringt eine hohe Relevanz für das Mitarbeiterempfehlungssystem. Die Grundprinzipien inhaltsbasierter Empfehlungssysteme sind zum einen die Analyse der Beschreibung der von einem bestimmten Benutzer bevorzugten *Items*, um die gemeinsamen Hauptattribute (Präferenzen) zu identifizieren, die diese *Items* unterscheiden. Diese Präferenzen werden in einem *Benutzerprofil* gespeichert [22]. Zusätzlich werden die Eigenschaften jedes *Items* mit dem *Benutzerprofil* verglichen, so dass nur *Items* empfohlen werden, die eine hohe Ähnlichkeit mit dem *Benutzerprofil* aufweisen [22]. Bei der Idee der Mitarbeiterempfehlung kann also die *Bedarfsmeldung* mit den benötigten Projektskills und Anforderung als *Benutzerprofil* angesehen werden. Die Mitarbeiterprofile sind dabei die *Items*. Die Attribute werden verglichen (Skills der Mitarbeiter mit den Skills und Anforderungen der *Bedarfsmeldung*) und ähnliche *Items* werden vorgeschlagen. Mit Hilfe traditioneller Methoden des Information Retrievals, wie z.B. dem Kosinus-Ähnlichkeitsmaß, werden dann Empfehlungen generiert [22]. Darüber hinaus generieren sie Empfehlungen mit Hilfe von statistischen und maschinelle Lernverfahren, die in der Lage sind, Nutzerinteressen aus historischen Nutzerdaten zu lernen [22].

3.2 Verwandte Arbeiten

Es gibt eine Reihe an verwandten Arbeiten die sich mit unterschiedlichen Aspekten des Staffing Prozesses und der Nutzung von Information Retrieval und Filtering zur Informationsgewinnung beschäftigen. Dennoch beschäftigt sich keine Arbeit mit dem spezifischen Problem der Informationsgewinnung aus *Bedarfsmeldungen*.

3.2.1 kp wie ich es nenne

Im ersten Paper beschreiben die Autoren einen Ansatz zur Ableitung von Unternehmensdaten und digitalen Fußabdrücken von Mitarbeitern. Mit Hilfe eines Big-Data-Workflows, der die Komponenten Information Retrieval und Suche, Datenfusion, Matrixvervollständigung und ordinale Regression nutzt, können Informationen zur Expertise automatisch zusammengeführt und für die Nutzung durch Experten aufbereitet werden. Das System soll Fähigkeiten, Talente und Fachwissens der Mitarbeiter in einem breiten Bereich wie cloud computing oder cybersecurity einschätzen. Beim Ansatz des Information Retrieval und -fusion wird eine Liste von Suchbegriffen erstellt, die sich auf das breite Fachgebiet der Mitarbeiter beziehen. Die Suche wird nach jedem dieser Abfragebegriffe durchgeführt, um Zusammenhänge zwischen Mitarbeiter und Datenquellen zu finden. Die verschiedenen Zusammenhänge werden

miteinander verschmolzen, gewichtet und nach der Abfrage sortiert. Die Mitarbeiter werden nach Daten gewichtet und bewertet, um einen einzigen Wert (sehr niedrig, niedrig, moderat, etwas, begrenzt) für ihr Fachwissen in diesem breiten Bereich zu erhalten.[15]

3.2.2 Information Filtering

Diese Arbeit befasst sich unter anderem mit dem Aspekt des content based Information Filtering. Das Ziel dabei ist es Informationen auf die Interessengebiete der Benutzer zu reduzieren. Dazu werden nicht relevante Dokumente aus einem Strom von Informationen entfernt, sodass dem Anwendern nur relevante Dokumente präsentiert werden. Ein Teil der Arbeit beschäftigt sich mit der Informationsfilterung und mögliche Filterungsvarianten werden vorgestellt. Die Arbeit konzentriert sich auf die inhaltsbasierte Filterung von Textdokumenten und identifiziert Informationsfilterung als einen Spezialfall der Textklassifikation. Dazu wird ein Überblick über gängige Methoden des Information Filtering gegeben und ihre Leistung evaluiert. [20]

3.2.3 Vorverarbeitung

Diese Arbeit zeigt Wege und Schritte zur Aufbereitung von Datensätzen auf. Die Arbeit umfasst Data-Mining Vorverarbeitungsmethoden, um die Qualität der Daten zu verbessern. Diese weisen wichtiger Schritte auf, um die Effizienz in der Datensammlung zu verbessern [1]. (Nicht sicher ob ich das drin lassen soll)

In diesem Beitrag wird der Teil des Anforderungsspezifikationsprozesses diskutiert, der zwischen der textuellen Anforderungsdefinition und den dazugehörigen Diagrammen der Anforderungsspezifikation liegt. Es wird die These aufgestellt, dass die Erstellung einer textuellen Anforderungsbeschreibung, welche das Verständnis des Analysten für das Problem darstellt, die Effizienz der Anforderungvalidierung durch den Benutzer verbessert. Die vorliegende Idee ist aus dem Problem entstanden, dass Software-Entwickler nicht immer über die erforderlichen Kenntnisse in den fachlichen Abläufen der Themengebiete verfügen, die für die Erstellung der Software relevant sind. Im Rahmen der Anforderungsdefinition erfolgt eine textuelle Verfeinerung, welche als Anforderungsbeschreibung bezeichnet werden kann. Bei der Arbeit mit dem unterstützten Werkzeug *Tessi* ist der Analytiker durch die genannten Vorgaben gezwungen, Anforderungen zu vervollständigen und zu erklären sowie die

Rollen der Wörter im Text im Sinne der objektorientierten Analyse zu spezifizieren. Im Rahmen der Vorverarbeitung erfolgt eine Transformation der Requirements durch Templates.[18]

3.2.4 Hybride Ansätze

In der vorliegenden Untersuchung wird die Entwicklung von Kombinationen im Bereich des Information Retrievals analysiert. Dabei werden sowohl experimentelle Ergebnisse als auch die Retrieval-Modelle, die als formale Rahmen für die Kombination vorgeschlagen wurden, berücksichtigt. Es wird aufgezeigt, dass Kombinationsansätze für die Informationssuche als Kombination der Ergebnisse mehrerer Klassifikatoren auf der Grundlage einer oder mehrerer Darstellungen modelliert werden können. Zudem wird dargelegt, dass dieses einfache Modell Erklärungen für viele der experimentellen Ergebnisse liefern kann.[8]

Die vorliegende Arbeit kombiniert drei Ansätze des Information Retrievals mit dem Ziel, relevante Informationen aus Produktreviews zu extrahieren. Der Ansatz TF-IDF wird mit einem sogenannten CLASSIFIER Model kombiniert. Das Klassifikationsmodell verarbeitet drei Eingaben der Modelle LSTM, VADER und TF-IDF. Die Werte dieser Eingaben liegen im Bereich von $[0,1]$. Die Ausgabe des Klassifikationsmodells ist binär und gibt eine Vorhersage des vollständigen Textes der Modelleingabe aus (positiv oder negativ).[7]

Diese Arbeit befasst sich mit der Filterung von Fake news. In diesem Beitrag werden hybride Verfahren zur Gewinnung von Merkmalen untersucht, die in dem Gebiet noch nicht gründlich erforscht wurden. Die Anwendung von Hybridsystemen hat sich in einer Vielzahl von Anwendungsbereichen als nützlich erwiesen und zeigen eine Tendenz, die Fehlerquote zu reduzieren, indem sie Techniken wie TF-IDF und N-Grams verwenden.[36]

Im Rahmen dieser Studie wurde ein hybrider Algorithmus zur Extraktion von Schlüsselwörtern und Kosinusähnlichkeit zur Verbesserung der Satzkohäsion bei der Textzusammenfassung vorgeschlagen. Die vorgeschlagene Methode basiert auf einer Komprimierung von 50 %, 30 % und 20 %, um Kandidaten für die Zusammenfassung zu erstellen. Die Auswertung des Ergebnisses mittels t-Test zeigt, dass die vorgeschlagene Methode den Kohäsionsgrad signifikant erhöht. Der Ablauf umfasst die Analyse eines Dokuments mithilfe eines Extraktionsalgorithmus sowie die Berechnung der TF/IDF-Werte für jeden Begriff. Anschließend werden alle TF/IDF-Werte

für jeden Satz summiert. Im nächsten Schritt werden alle Sätze anhand der Summe von TF/IDF eingestuft. Das Kompressionsverhältnis bestimmt die Position des Satzrangs. In dieser Studie wird eine Kompression von 50 % verwendet, was bedeutet, dass die Satzzusammenfassung um 50 % des Originaltextes gekürzt wird. Nach der Auswahl des Satzes wird dessen Berechnung durchgeführt. Die Ähnlichkeit wird mit der Cosinus-Ähnlichkeitsmethode berechnet. Anschließend werden alle Sätze anhand ihrer Cosinus-Ähnlichkeit von der höchsten zur niedrigsten sortiert. Der resultierende Text mit neuer Satzanordnung stellt die finale Zusammenfassung dar.[9]

3.2.5 Pipeline

In dieser Arbeit wird eine Pipeline entwickelt, die die N-Gramm-Analyse verwendet, um Schlagwörter aus einem Text zu extrahieren und mit verschiedenen Ansätzen von Word-Clouds zu visualisieren.[32]

Die vorliegende Arbeit präsentiert eine Anleitung zur Erstellung einer Pipeline mit Python und TF-IDF. Darüber hinaus wird die Relevanz von TF-IDF als Vorverarbeitung beim maschinellen Lernen erörtert. Im Vergleich zur rohen Termhäufigkeit weist TF-IDF in der Regel einen höheren Vorhersagewert auf. Die Gewichtung von Themenwörtern wird erhöht, um die Bedeutung von Wörtern zu erhöhen, während die Gewichtung von hochfrequenten Funktionswörtern verringert wird. Es werden Verfahren zur Vorverarbeitung von Texten vorgestellt, die eine Umformung in die gewünschte Darstellungsform ermöglichen. Zudem werden Methoden zur Interpretation der Ergebnisse des TF-IDF-Verfahrens erörtert.[21]

Die Verarbeitung natürlicher Sprache wirft insbesondere bei der Analyse unüblicher Sprachen wie Griechisch Schwierigkeiten auf. In diesem Beitrag wird ein maschineller Lernansatz für die Bereiche Part-of-Speech-Tagging und Named-Entity-Recognition für die griechische Sprache unter Verwendung von spaCy erarbeitet und evaluiert. [29]

spam-filter (Empfinde das Thema eventuell als zu unpassend)

-Überblick über verfügbare Methoden, Herausforderungen und zukünftige Forschungsrichtungen im Bereich der Spam-Erkennung, Filterung und Eindämmung von SMS-Spam. Dabei werden auch Methodiken der keyword frequency ratio und Herunterbrechung auf keyword components behandelt [35]

3.3 Definitionen und Konzepte

ToDo: Lieber vor dem Kapitel Verwandte Arbeiten packen

<-hier

3.3.1 Information Retrieval und Information Filtering

Im Allgemeinen wird einem Informationssystem die Funktion zugeschrieben, den Benutzer zu den Dokumenten zu führen, die seinen Informationsbedarf am besten decken. Allgemeiner ausgedrückt ist das Ziel eines Informationssystems, dem Benutzer Informationen aus der Wissensressource zur Verfügung zu stellen, die ihm helfen, ein Problem zu lösen. Auf der anderen Seite ist unter Filtern das Entfernen von Daten aus einem eingehenden Datenstrom zu verstehen und nicht das Auffinden von Daten in diesem Datenstrom. Filtersysteme verarbeiten große Datenmengen. Typische Anwendungen betreffen Gigabytes von Text oder weitaus größere Mengen anderer Medien. Während es bei dem Information Retrieval typischerweise um die einmalige Nutzung des Systems durch eine Person mit einem einmaligen Ziel und einer einmaligen Anfrage geht, befasst sich die Informationsfilterung mit der wiederholten Nutzung des Systems durch eine oder mehrere Personen mit langfristigen Zielen oder Interessen.[3]

3.3.2 Data-Mining

Data Mining ist ein interdisziplinäres Teilgebiet der Informatik, das sich mit der rechnergestützten Entdeckung von Mustern in großen Datenbeständen befasst. Ziel dieses fortgeschrittenen Analyseverfahrens ist es, Informationen aus einem Datensatz zu extrahieren und in eine für die weitere Verwendung verständliche Struktur umzuwandeln. Die verwendeten Methoden liegen an der Schnittstelle zwischen künstlicher Intelligenz, maschinellem Lernen, Statistik, Datenbanksystemen und Business Intelligence. Beim Data Mining geht es um die Lösung von Problemen durch die Analyse von Daten, die bereits in Datenbanken vorhanden sind.[16]

3.3.3 Bedarfsmeldungen

gg

4 Entwicklung einer klaren Erwartungshaltung

Dieses Kapitel befasst sich mit der Methodologie und Durchführung von Experteninterviews mit dem Ziel ...

4.1 Beschreibung der Interviews mit Führungskräften zur Identifizierung von Stakeholder-Erwartungen

Im Rahmen der vorliegenden Ausarbeitung werden halbstrukturierte Interviews mit Experten aus dem Bereich <> durchgeführt. Sinn und Zweck von Experteninterviews ist die Rekonstruktion spezifischer Wissensbestände oder besonders exklusiver, detaillierter oder umfassender Kenntnisse über bestimmte Wissensbestände und Praktiken.

Der Begriff *Experte* bezeichnet eine Person, die über einen privilegierten Zugang zu Informationen verfügt [31]. Die Expertise eines Experten ist jedoch nicht allein durch die Informationen definiert, über die er exklusiv verfügt [31]. Auch die Verantwortung für Problemlösungsentscheidungen ist ein entscheidender Faktor [31]. Diesbezüglich ist Kompetenz erforderlich, die mit Verantwortung und mit Fähigkeiten sowie mit der Bereitschaft, Verantwortung zu übernehmen, verbunden ist [31]. Dabei ist zu beachten, dass Verantwortung, Fähigkeiten und Bereitschaft in der Regel zusammenfallen [31].

Die Experteninterviews in dieser Ausarbeitung zielen darauf ab, qualitative Daten zu erheben. Die Interviews werden als Einzelinterviews durchgeführt, wodurch der Fokus auf das spezifische Wissen jedes Befragten gerichtet werden kann. Jeder Interviewpartner reagiert individuell aufgrund seines eigenen Vorwissens auf die Interviewfragen und beeinflusst daher nicht die Aussagen anderer Interviewteilnehmer. Die Ergebnisse der Interviews bilden die Grundlage für die Formulierung der Anforderungen einer optimalen Bedarfsmeldung, welche als Basis für das zu entwickelnde System eingesetzt wird. Die Interviews wurden in Teams abgehalten und jedes Interview wird zu Dokumentationszwecken aufgezeichnet. Um die Interviews strukturiert für die qualitative Inhaltsanalyse vorzubereiten, ist es erforder-

lich, sie vorher in schriftliche Transkripte umzuwandeln. Zur ersten Umwandlung in Text wurde das Transkriptionstool von Teams verwendet. Ungenaue Umwandlungen wurden mit der Videoaufnahme nachgebessert. Grundsätzlich wird die einfache Transkribierung nach Dresing und Pehl angewandt [11]. Im Rahmen dieses Schritts wird der Text vom Umgangssprachlichem in einen gut Lesbaren Text ohne Lücken übersetzt. Der Interviewer wurde mit einem „I“ und die jeweils befragte Person mit „B“ gekennzeichnet. Im Rahmen des Transkriptionsprozesses werden die Aussagen der Interviews anonymisiert. Dies bedeutet, dass sämtliche personenbezogenen Informationen, wie Vor- und Nachname, durch neutrale Bezeichnungen ersetzt werden.

Im Vorfeld der Durchführung der Interviews wurde eine Überprüfung der inhaltlichen Verständlichkeit der Fragen sowie ihrer Beantwortbarkeit vorgenommen. Zudem wurde Feedback zur Reihenfolge der Fragen eingeholt. Zu diesem Zweck wurden die Fragen vorab an eine Führungsperson geschickt und schriftlich beantwortet. Die Fragen wurden den Experten vorab inklusive Kontext des Interviews geschickt, damit diese sich bei Bedarf Gedanken machen können. Zur zeitlichen Begrenzung wird das Interview auf zehn Fragen reduziert. Trotz der vorgegebenen Strukturierung des Interviews wird Raum für spontane Fragen gelassen, um eine natürliche Gesprächsführung zu ermöglichen. Die Fragen dienen als Orientierungshilfe und Leitfaden durch das Interview. Der Leitfaden ist dabei lediglich als inhaltliche Richtlinie zu verstehen, von der situativ abgewichen werden kann.

4.2 Ablauf der Interviews

Im Rahmen der Interviews erfolgt zunächst eine Einführung in die Problemstellung sowie das Ziel der Ausarbeitung. Auf diese Weise wird sichergestellt, dass die Interviewpartner den Sinn und Zweck des Interviews nachvollziehen können. Des Weiteren erfolgt eine definitorische Erläuterung des Begriffs „Bedarfsmeldung“. Auch wenn dies die fachliche Bezeichnung darstellt, ist sie nicht jedem einzelnen Experten geläufig. Im Rahmen der Einführung erfolgt zunächst eine Erörterung der Rolle des Experten bei adesso. Im Rahmen dessen erfolgt eine Klärung der genauen Aufgaben des Experten bei adesso. Im Anschluss erfolgt eine Erörterung der Rolle des Experten im Kontext des Staffing-Prozesses sowie der Bearbeitung der Bedarfsmeldung. Die Interviewfragen sind so konzipiert, dass sie zunächst allgemein gehalten sind und im Verlauf des Interviews zunehmend präziser werden. Im Anschluss an die Erörterung der Frage, welche Art von Projekten über Bedarfsmeldungen erfasst wird, wird die Frage aufgeworfen, auf welche Weise diese kommuniziert und dokumentiert

werden. Im Anschluss erfolgt eine Klärung der für die Erfassung von Bedarfsmeldungen relevanten Informationen. Dabei wird erörtert, welche Informationen von besonderer Bedeutung sind und folglich nicht fehlen dürfen. Die Herausforderungen und Bewertungskriterien geben Aufschluss über die bereits genutzten Ansätze zur Standardisierung der Bedarfsmeldungen. Des Weiteren können bereits unternommene Maßnahmen zur Qualitätssicherung als hilfreicher Ansatz zur Identifizierung von Anforderungen der Bedarfsmeldungen herangezogen werden. Die Beantwortung der Fragen erfordert insbesondere die Entwicklung eigener Ideen und Konzepte, die im weiteren Verlauf des Prozesses zur Erfassung von Bedarfsmeldung gegebenenfalls noch nicht zum Einsatz gekommen sind.

4.3 Übersicht der Experten

Nr. und Datum	Profil der Befragten	Durchführungsart	Dauer
1. ...	CC-Leiter	schriftlicher Vorabtest	-
2. 29.04.2024	und Projektleiter	Video-Interview	20min
3. 29.04.2024	CC-Leiter und Delivery Manager	Video-Interview	45min
4. 30.04.2024	CC-Leiter und Softwarearchitekt	Video-Interview	...min
5. 6.05.2024	CC-Leiter	Video-Interview	...min

Die erste Spalte der Tabelle zeigt die vergebene Nummerierung sowie das Datum der Durchführung. Dies dient der vereinfachten Referenzierung innerhalb der nachfolgenden Analyse. Die zweite Spalte der Tabelle enthält die Rolle bzw. Tätigkeit der Befragten bei adesso. Alle Befragten haben, oder hatten eine Leitende Rolle mit Erfahrungen in der Personaleinsatzplanung. Dementsprechend hat jeder Befragte in irgend einer Form Berührungspunkte mit Bedarfsmeldungen gehabt. Interview 2 und 3 haben zusätzlich noch die Perspektive zur Erstellung und Verwaltung von Bedarfsmeldungen, da ihre Hauptaufgaben genau in diesem Bereich fallen. Die 3. Spalte zeigt die Art der Befragung. Abgesehen von dem Vorabtest wurden alle Interviews in einem Video-Call abgehalten.

Die erste Befragte Person aus dem schriftlichen Vorabtest ist zum einen CC-Leiter. Dies ist die Bezeichnung für Führungspersonen mit Zuständigkeiten für Mitarbeiter bei adesso.

4.4 Analyse der Ergebnisse und Entwicklung einer klaren Erwartungshaltung für die Bedarfsmeldungen

1. Transkription der Interviews:

Falls du die Interviews aufgezeichnet hast, transkribiere sie vollständig und genau. Dadurch hast du eine schriftliche Version der Aussagen der Experten, die du leichter analysieren kannst.

2. Codierung der Daten:

Gehe durch die transkribierten Interviews und markiere oder kodiere relevante Themen, Aussagen oder Muster. Verwende dabei Codes oder Kategorien, die sich auf deine Forschungsfragen beziehen.

3. Thematische Analyse:

Führe eine thematische Analyse durch, indem du die kodierten Daten systematisch durchgehst und nach wiederkehrenden Themen oder Mustern suchst. Identifiziere Gemeinsamkeiten, Unterschiede oder interessante Einsichten, die sich aus den Aussagen der Experten ergeben.

4. Triangulation:

Vergleiche die Ergebnisse der Experteninterviews mit anderen Quellen, wie beispielsweise der Literatur, Fallstudien oder empirischen Daten. Durch die Triangulation kannst du die Glaubwürdigkeit und Validität deiner Ergebnisse erhöhen.

5. Interpretation der Ergebnisse:

Interpretiere die identifizierten Themen oder Muster im Kontext deiner Forschungsfragen und -ziele. Versuche zu verstehen, welche Bedeutung oder Implikationen die Aussagen der Experten für deine Forschung haben könnten.

6. Reflexion und Kritik:

Reflektiere kritisch über die Aussagen der Experten und die gewonnenen Erkenntnisse. Berücksichtige mögliche Einschränkungen oder Bias in den Interviews und betrachte die Ergebnisse aus verschiedenen Perspektiven.

7. Integration in die Gesamtanalyse:

Integriere die Ergebnisse der Experteninterviews in deine Gesamtanalyse deiner Ba-

chelorarbeit. Verknüpfe sie mit anderen Forschungsergebnissen, theoretischen Konzepten oder empirischen Daten, um ein umfassendes Verständnis deines Forschungsthemas zu entwickeln.

8. Darstellung der Ergebnisse:

Präsentiere die wichtigsten Ergebnisse und Erkenntnisse aus den Experteninterviews in deiner Bachelorarbeit. Verwende geeignete Zitate oder Beispiele, um die Aussagen der Experten zu veranschaulichen und deine Argumentation zu unterstützen. [23]

im anhang sind die transskripte wenn man nicht ne größere anzahl an infos hat gucken ob man das halb automatisch evaluieren. Vielleicht kategorisieren. Infos die wichtig sind gucken ob die dann auch nach dem preprocessing drin sind. Regressive tests schreiben.

transformation von bedarfsmeldung zu guter bedarfsmeldung, was ist der fokus von der bedarfsmeldung, wie gut machen die ansätze das, und muss man das dann noch weiter verarbeiten, haben wir alles was wir brauchen mit nur einem algorithmus, inferenz falls parameter fehlt, gibt es einen der alles löst

Fragen in das proposal aufnehmen, führungskraft vorher fragen ob die fragen nice sind.

g

1. Wer sind die typischen Stakeholder bei der Erstellung von Bedarfsmeldungen und welche Rolle spielen sie?
2. Welche Art von Projekten sind typischerweise in Ihrem Unternehmen an der Tagesordnung? Können Sie uns Beispiele für verschiedene Arten von Projekten geben, die adesso durchführt?
3. Wie werden Projektbedarfe und -anforderungen innerhalb von adesso typischerweise kommuniziert und dokumentiert?
4. Welche Informationen halten Sie in einer Bedarfsmeldung für besonders wichtig oder unverzichtbar?
5. Wie detailliert sollten Projektbeschreibungen Ihrer Meinung nach sein? Sind bestimmte Schlüsselaspekte oder -informationen in jeder Bedarfsmeldung enthalten?
6. Wie wird die Qualität von Bedarfsmeldungen bei adesso bewertet? Gibt es bestimmte Kriterien oder Standards, anhand derer Bedarfsmeldungen beurteilt werden?
7. Wie können Sie die Qualität und Klarheit von Bedarfsmeldungen verbessern?
8. Welche Herausforderungen oder Schwierigkeiten sind bei unklaren oder unvollständigen Bedarfsmeldungen aufgetreten?
9. Welche Auswirkungen haben unklare oder fehlende Informationen in Bedarfsmeldungen auf die Effizienz und den Erfolg von Projekten?
10. Wie können Sie sicherstellen, dass die Bedürfnisse und Anforderungen aller relevanten Stakeholder in einer Bedarfsmeldung angemessen berücksichtigt werden?

g

5 Analyse der Techniken des Information Retrieval und Data-Mining

5.1 Beschreibung der untersuchten Techniken und Ansätze

5.1.1 TF-IDF

Im Rahmen der Textanalyse wird die TF-IDF-Technik angewendet, welche die häufigsten Begriffe eliminiert und lediglich die relevantesten Begriffe aus einem Textkorpus extrahiert [2]. Die TF-IDF-Methode dient der Ermittlung der Häufigkeit von Wörtern in einem bestimmten Dokument im Vergleich zum Anteil dieses Wortes im gesamten Dokumenten [33]. Die Berechnung erlaubt eine Einschätzung der Relevanz eines bestimmten Wortes in einem bestimmten Dokument [33]. Die Grundidee des Ansatzes besteht darin, dass Wörter, die in einem einzigen Dokument oder in einer kleinen Gruppe von Dokumenten häufig vorkommen, tendenziell höhere TF-IDF-Werte aufweisen als häufig vorkommende Wörter wie Artikel und Präpositionen [33]. TF-IDF stellt ein effizientes Verfahren zum Abgleich von Wörtern in einer Anfrage mit Dokumenten dar [33]. Bei Eingabe einer Abfrage zu einem bestimmten Thema durch einen Benutzer kann TF-IDF relevante Informationen zu dieser Abfrage in Dokumenten finden [33]. In Bezug auf die Bedarfsmeldungen besteht somit die Möglichkeit, Wörter aus einer Bedarfsmeldung mit anderen Bedarfsmeldungen zu vergleichen und die Häufigkeit ihrer Verwendung zu ermitteln, um somit potenzielle Schlüsselwörter zu ermitteln.

Trotz der Stärken von TF-IDF, sind auch seine Grenzen zu berücksichtigen. In Bezug auf Synonyme ist zu beachten, dass TF-IDF nicht auf die Beziehung zwischen den Wörtern eingeht. Des Weiteren werden unterschiedliche Schreibweisen von Wörtern nicht berücksichtigt, was dazu führen kann, dass Wörter fälschlicherweise als nicht so häufig auftauchend deklariert werden, obwohl sie mit leicht abgewandelter Schreibweise häufiger vorkommen.

5.1.2 Text-Ranking-Algorithmen

Text-Ranking-Algorithmen: Text-Ranking-Algorithmen wie TextRank oder YAKE (Yet Another Keyword Extractor) verwenden graphenbasierte Methoden, um Schlüsselwörter in einem Text zu identifizieren. Die Algorithmen bewerten die Wichtigkeit von Wörtern basierend auf ihrer Verbindung zu anderen Wörtern im Text und extrahieren Schlüsselwörter entsprechend ihrer Rangfolge.

[26][37][30]

Ein graphenbasierter Rangordnungsalgorithmus stellt eine Möglichkeit dar, die Wichtigkeit eines Knotens innerhalb eines Graphen zu bestimmen. Dabei werden globale Informationen berücksichtigt, die rekursiv aus dem gesamten Graphen berechnet werden. Im Gegensatz zu anderen Methoden, die sich lediglich auf lokale, knotenspezifische Informationen stützen, ermöglicht dies eine objektivere Bewertung.[26]

5.1.3 N-Gramm

N-Gramme sind Folgen von Zeichen oder Wörtern, die aus einem Text extrahiert werden. N-Gramme lassen sich in zwei Kategorien unterteilen: i) zeichenbasiert und ii) wortbasiert. Ein Zeichen-N-Gramm bezeichnet eine Folge von n aufeinanderfolgenden Zeichen, die aus einem Wort extrahiert werden. Die Hauptmotivation hinter diesem Ansatz besteht darin, dass ähnliche Wörter einen hohen Anteil an N-Grammen gemeinsam haben werden. In der Regel umfasst ein N-Gramm lediglich die am häufigsten auftretenden Wortpaare und verwendet einen Backoff-Mechanismus, um die Wahrscheinlichkeit zu berechnen, die bei der Suche nach dem gewünschten Wortpaar nicht erfolgreich war. [24] Die Analyse von N-Grammen erlaubt die Identifikation häufig vorkommender Phrasen oder Begriffe, die als potenzielle Schlüsselwörter bezeichnet werden können.

5.1.4 POS-Tagging

Die Katalogisierung von Wortarten (POS) bezeichnet einen Prozess, bei dem jedem einzelnen Wort eines Satzes ein Wortart-Tag oder ein anderes philologisches Klassenzeichen zugeordnet wird. Die Vorverarbeitungsaufgabe des Taggings von Sprachbestandteilen stellt einen essenziellen Schritt in der Verarbeitung natürlicher Sprache dar. Die Zuordnung von Wortarten stellt eine grundlegende Aufgabe bei der Verarbeitung natürlicher Sprache dar. Die Erstellung erfolgt unter Zuhilfenahme linguistischer Theorien, zufälliger Muster sowie einer Kombination aus beidem. Ein Part-of-Speech-Tagger (POS-Tagger) ist definiert als ein Teil einer Software, der

jedem Wort einer Sprache, das er liest, eine Wortart zuordnet. Die Ansätze des POS-Tagging lassen sich in drei Kategorien unterteilen: regelbasiertes Tagging, statistisches Tagging und hybrides Tagging. Im Rahmen der Zuweisung von POS-Tags zu Wörtern im regelbasierten POS-System erfolgt die Verwendung einer Reihe von handgeschriebenen Regeln in Kombination mit Kontextinformationen. Der Nachteil dieses Systems besteht darin, dass es nicht funktioniert, wenn der Text nicht bekannt ist. Das Problem besteht darin, dass das System nicht in der Lage ist, den passenden Text vorherzusagen. Um eine höhere Effizienz und Genauigkeit in diesem System zu erreichen, ist es daher empfehlenswert, einen umfassenden Satz von handkodierten Regeln zu verwenden. Die Häufigkeit und Wahrscheinlichkeit sind in den statistischen Ansatz einbezogen. Der grundlegende statistische Ansatz basiert auf der am häufigsten verwendeten Markierung für ein bestimmtes Wort in den annotierten Trainingsdaten. Diese Information wird auch zur Markierung dieses Wortes im unannotierten Text verwendet.[19]

5.1.5 Named Entity Recognition

Named Entity Recognition (NER) [25] [27][29]

5.1.6 Regelbasierte Ansätze

Regelbasierte Ansätze: Regelbasierte Ansätze verwenden vordefinierte Regeln oder Muster, um Schlüsselwörter zu identifizieren. Dies kann beispielsweise das Extrahieren von Wörtern sein, die häufig im Text vorkommen oder bestimmten Mustern entsprechen.

5.1.7 Hybride Ansätze

Hybride Ansätze: Hybride Ansätze kombinieren verschiedene Methoden und Techniken, um eine genauere Extraktion von Schlüsselwörtern zu ermöglichen. (Z.B. Kombination aus TF-IDF-Gewichtung und Text-Ranking-Algorithmen verwendet).[9]

5.1.8 Data-Mining

Data-Mining: [17][16]

5.1.9 preprocessing

preprocessing: [14]

5.1.10 data-fusion

data-fusion: [12] [13] [4]

5.2 Bewertung und Auswahl der besten Ansätze für die Extraktion relevanter Inhalte aus Bedarfsmeldungen

6 Konzeptionierung und Implementierung der Vorverarbeitung

6.1 Beschreibung des entwickelten Vorverarbeitungsmodells

basierend auf den ausgewählten Techniken

-auch translate erwähnen. Wichtig damit die meisten ansätze gut funktionieren

g

g

g

g

6.2 Details zur Implementierung der Pipeline in Python

für die effiziente Verarbeitung von Bedarfsmeldungen

g

7 Evaluierung des entwickelten Systems

7.1 Beschreibung des verwendeten Datensatzes und der Evaluierungsmethodik

überlegung ob tfidf unterschied macht alle bedarfsmeldungen mit einer zu vergleichen
und daraus wichtige wörter identifizieren oder eine für sich alleine reicht.
gucken was tokenisierung wirklich macht

7.2 Präsentation und Diskussion der Ergebnisse

g

g

g

Zeit und Leistung Übersicht

g

7.3 Vergleich des Systems mit einem Large Language Model-Ansatz

g

g

g

7.4 Analyse von Abweichungen, Ähnlichkeiten und Verbesserungspotenzialen des Systems

g

8 Zusammenfassung und Ausblick

ergebnis der arbeit: diese modelle in der reihenfolge kommen am nächsten an die bedarfsmeldung

Ausblick

die keyword extraction auch für die profile nutzen

9 Anhang

9.1 Interviewtranskripte

9.1.1 Vorabtest Fragen und Antworten

1. Welche Art von Projekten sind typischerweise in Ihrem Unternehmen an der Tagesordnung? Können Sie uns Beispiele für verschiedene Arten von Projekten geben, die adesso durchführt?

Software-Entwicklungsprojekte, angefangen von Projekten in dem ein adessi in einem Kundenprojekt arbeitet über gemischte Teams aus adessi und Kunde bis hin zur kompletten Lieferung von Projektleitern, Testern, Requirements Engineer und Entwicklern

2. Wie werden Projektbedarfe und -anforderungen innerhalb von adesso typischerweise kommuniziert und dokumentiert?

Initial über den Maitre, der das Staffing übernimmt bzw. auch Vorschläge von Projektleitenden zum Staffing annimmt, teilweise auch über das eigene Netzwerk zwischen Führungskräften, im CC, Bereich oder der LoB. Am Ende über das Staffing Jira

3. Welche Informationen halten Sie in einer Bedarfsmeldung für besonders wichtig oder unverzichtbar?

Senioritätslevel, Tagessatz, Remote/on Site Einsatz, Dauer, Technischer Stack (Muss- und Kann Kriterien), Einarbeitungszeiträume (ist es verrechenbar oder nicht?), Lieferverpflichtung

4. Wie detailliert sollten Bedarfsmeldungen Ihrer Meinung nach sein? Sind bestimmte Schlüsselaspekte oder -informationen in jeder Bedarfsmeldung enthalten?

Es sollte aussagefähig sein zumindest welche technischen Kompetenzen wichtig sind und welche Tagessätze, ob Remote möglich ist und die Dauer mindestens in der Bedarfsmeldung vorhanden sein.

5. Welche Herausforderungen oder Schwierigkeiten sind bei unklaren oder unvollständigen Bedarfsmeldungen aufgetreten?

Der Anforderer muss ggf. Fragen mehrfach beantworten, der Kanal über den kommuniziert wird (Teams Chat, Anruf, im Ticket, ...). Dadurch verliert man ggf. den

Überblick

6. Wer sind die typischen Stakeholder bei der Erstellung von Bedarfsmeldungen und welche Rolle spielen sie?

Sales/PL: Anforderer mit den technischen Informationen, Maitre: kümmert sich um das Staffing bzw. die eigentliche Besetzung

7. Wie wird die Qualität von Bedarfsmeldungen bei *adesso* bewertet? Gibt es bestimmte Kriterien oder Standards, anhand derer Bedarfsmeldungen beurteilt werden?

Gar nicht meines Wissens nach.

8. Wie können Sie die Qualität und Klarheit von Bedarfsmeldungen verbessern?

Zukünftig: Durch klarere Vorgaben und weniger Freitext, aktuell: durch Nachfragen und Bitten um nachträgliche Pflege, ggf. durch Reviewprozesse bei eigenen Bedarfsmeldungen

9. Welche Auswirkungen haben unklare oder fehlende Informationen in Projektbeschreibungen auf die Effizienz und den Erfolg von Projekten?

Das Staffing dauert länger und ggf. werden die Stellen durch andere Dienstleister besetzt

10. Wie können Sie sicherstellen, dass die Bedürfnisse und Anforderungen aller relevanten Stakeholder in einer Bedarfsmeldung angemessen berücksichtigt werden?

Gute Abstimmungen bevor die Bedarfsmeldung erstellt wird, ggf. durch ein Quality-Gate (Review).

9.1.2 Interview 2

0:1:13.840 -> 0:1:27.60 Valente de Matos, Ricardo Du bist zum einen CC Leiter. Das heißt du hast auch Menschen unter deiner Leitung, die du auf Projekte zuweist.
0:1:33.180 -> 0:1:33.820 Bürger, Marco Genau ja. 0:1:29.780 -> 0:1:36.620 Valente de Matos, Ricardo Du weiß also zu Projekten zu, übernimmst auch die Projektleitung. 0:1:37.40 -> 0:1:37.630 Bürger, Marco Ja, genau. 0:1:38.800 -> 0:1:49.260 Valente de Matos, Ricardo Genau da würde ich dann erst mal gerne wissen: Was sind so die typischen Stakeholder bei der Erstellung von Bedarfsmeldungen und welche Rolle hast du dabei grundsätzlich? 0:1:50.110 -> 0:1:59.390 Bürger, Marco In den

Situationen nehme ich immer die Rolle des Beraters erstmal an. Die Stakeholder sind klassisch die Fachverantwortlichen beim Kunden, aber auch die Entscheider. Sprich also deren Vorgesetzte die quasi fachlich vielleicht das ganze nicht so bewerten können, aber das Budget dafür hergeben müssen und natürlich dann im Zweifelsfall auch CO, CEO oder sogar Geschäftsführer. 0:2:41.310 → 0:2:48.230 Valente de Matos, Ricardo Kannst du ein paar Beispiele nennen, welche Arten von Projekten adesso so erhält und durchführt. 0:2:48.870 → 0:3:21.890 Bürger, Marco Ja, also im Prinzip kannst du das in 2 Arten von Projekten teilen. Aus einer sind Teil Material Projekte, wo quasi der Kunde mit einer Idee kommt, wo wir gut unterstützen können. Beispielsweise bei Bestandsprojekten. Oder vielleicht weil ihnen selbst die Ressourcen dafür fehlen. Zum anderen hast du halt Festpreis Projekte, wo wir bestimmtes Gewerk für den Kunden abschätzen und das Ganze auch dann gänzlich liefern. In Festpreis Projekten haben wir eher die Staffing Hoheit. Das heißt, wir können entscheiden wen wir in das Projekt einsetzen? Wobei bei einigen Projekten durchaus auch der Kunde sich die Profile mit anschaut und dann auch entscheidet. Anhand von Interviews was macht Sinn, was passt bei mir ins Team vielleicht und denke ich, dass das am besten für mich wäre? 0:3:50.340 → 0:3:58.320 Valente de Matos, Ricardo Wie werden diese Bedarfsmeldungen und Anforderungen denn so typischer Weise kommuniziert und dokumentiert? Gibt es da eine Art Ablauf, oder wie wird das gemacht? 0:4:2.620 → 0:4:8.130 Bürger, Marco Hängt auch immer ehrlicherweise vom jeweiligen Kunden ab. Bei manchen reicht ein Interview, welches du führst und dann schreibst du es in ein Word Dokument als Anforderungsbeschreibung nieder. Dann lässt man das gegen Zeichnen und dann ist gut. Manchmal muss man aber auch ein paar mehr Integration fahren und dann nochmal genau abzustecken, was denn Bestandteil der Beauftragung ist und was nicht. Also was die Bedarfsmeldung. Da muss man auch mal durchaus eins tiefer bohren, weil da teilweise der Gedanke, was der Kunde möchte nicht mit dem übereinstimmt, was eigentlich gebraucht wird. Das ist so, weil du klassisch irgendein Anforderungsdokument dafür fertig machst, wenn das immer ein bisschen höher geht Richtung Management. Beim Management mit dem Kunden kann es auch durchaus mal eine Präsentation sein, wo das Ganze nochmal ein bisschen aufbereitet ist. Ein bisschen klassisch, ein bisschen bunter und mit weniger Infos und mit weniger Tiefe präsentiert. 0:5:16.340 → 0:5:23.480 Valente de Matos, Ricardo Welche Informationen sind denn für dich in Bedarfsmeldungen besonders wichtig oder auch unverzichtbar? Also zum Beispiel gibt es ja auch das Senioritätslevel. 0:5:38.820 → 0:5:45.70 Bürger, Marco Ja genau, also Senioritätslevel ist ehrlicherweise erstmal das zweitrangige. Das müssen wir im Nachgang einmal prüfen. Wichtig ist halt was der Text Deck und was vom

Kunden kommt. Und quasi anhand der Bedarfsmeldung an sich. Wie hoch ist der Aufwand, der dahinter steckt? Sowohl durch den Kunden als auch das, was wir teilweise schätzen. Dann kann durchaus sein, dass der Kunde sagt, was gebraucht wird, er aber eigentlich keine Ahnung hat. Es kann vorkommen, dass der Kunde gerne in 2 Monaten durch ist, aber es keinen Sinn macht und wir mindestens ein halbes Jahr benötigt. Dann trifft man sich irgendwo in der Mitte und muss aber noch abgrenzen was Sinn macht und was nicht. Wenn das alles klar ist, dann kannst du überlegen was du eher brauchst. Machst du z.B. nur was mit Senioren? Oder reicht ein Junior. Das hängt dann immer eher von der Gesamtsituation ab. 0:6:50.110 -> 0:6:52.600 Valente de Matos, Ricardo Wie detailliert sollte dann eine Bedarfsmeldung sein? Gibt es Aspekte oder Informationen, die eigentlich in jeder Bedarfsmeldung drin sein sollten und müssen? 0:7:5.130 -> 0:7:6.880 Bürger, Marco Ich glaube der Umfang ist immer ganz wichtig. Die Erwartungshaltung sollte immer detailliert sein. Und die Technologien ebenfalls. 0:7:24.970 -> 0:7:33.660 Valente de Matos, Ricardo Welche Herausforderungen oder Schwierigkeiten sind dann bei zum Beispiel unklaren oder unvollständigen Bedarfsmeldungen aufgetreten? Hattest du so etwas schon mal? 0:7:35.450 -> 0:7:42.600 Bürger, Marco Ja durchaus, das sind immer Lernprozesse sowohl beim Kunden als auch bei dem, der die Anforderungen aufnimmt. Hängt immer vom Reifegrad des jeweiligen Konterparts ab. Das heißt auch, dass umso mehr du dich in der Situation meldest und Sachen aufgenommen hast, umso genauer kannst du mal nachfragen und hörst auch die Unsicherheit auf der einen oder anderen Seite heraus. 0:8:8.110 -> 0:8:17.50 Valente de Matos, Ricardo Gibt es denn irgendwie ein Mechanismus, wie die Qualität von Bedarfs Meldungen bewertet wird, oder gibt es irgendwelche Kriterien oder Standards womit dann beurteilt wird, ob eine Bedarfsmeldung gut oder eher schlecht ist? 0:8:25.100 -> 0:8:32.870 Bürger, Marco Ich bin ein Fan davon auch Sachen mal querlesen zu lassen und dann vielleicht mal die ein oder andere Meinung einzuholen und auch noch mit dem Kunden, der die Bedarfsmeldungen stellt ganz nah dran zu bleiben und dann zu gucken, dass das immer funktioniert. Damit auch zum Schluss das rauskommt, was am Anfang vielleicht schon ne Idee gewesen ist. 0:8:58.820 -> 0:9:7.190 Valente de Matos, Ricardo Wie würdest du denn die Qualität von Bedarfsmeldungen verbessern? 0:9:11.70 -> 0:9:34.240 Bürger, Marco Auch wirklich mit dem Kunden das Ganze einmal durchexerzieren und festzustellen, ob das Verständnis auf allen Seiten das gleiche ist. Weil nur dann kann es auch gut und produktiv werden. Wenn direkt am Anfang schon irgendwie Unklarheit da ist, dann kannst du auch davon ausgehen, dass da Diskussionsbedarf entsteht. 0:9:41.370 -> 0:9:49.980 Valente de Matos, Ricardo Welche Auswirkungen haben denn unklare und fehlende Informatio-

nen in Bedarfsmeldungen in Bezug auf die Effizienz und den Erfolg des Projektes? Hast du da irgendwie Erfahrungen machen können? 0:9:52.830 -> 0:10:1.870 Bürger, Marco Ja, wenn wir uns unklar sind, wird der Aufwand immer um ein Vielfaches erhöhen, weil das irgendwie im Nachgang immer noch mal gerade gezogen werden muss. Wenn du Pech hast, kannst du alles, was du bisher gemacht hast wegschmeißen und nochmal neu anfangen. Bedeutet natürlich auch immer, dass ein großes Diskussionspotenzial zwischen sowohl den Projektbeteiligten als auch Kunde und Projekt existiert. 0:10:31.900 -> 0:10:44.900 Valente de Matos, Ricardo Wie könnte man theoretisch sicherstellen, dass die Bedürfnisse und Anforderungen aller Leute, die involviert sind, irgendwie angemessen berücksichtigt werden? 0:10:51.310 -> 0:10:56.730 Bürger, Marco Ich glaube wenn du da mit viel Erfahrung reingehst und dann auch vielleicht genau weißt, wo du drauf zu achten hast. Mir fehlt ein bisschen die Fantasie, aber das ist ja jetzt dann deine Aufgabe da so ein Automatismus zu erkennen. Das wird auf jeden Fall schwierig weil ich auch an der Stelle glaube, dass wenn du so ein System schaffst, die quasi den Match drauf machen können, es trotzdem viel lernen muss. Von daher glaube ich, Erfahrung ist das es ausmacht, um am Ende zu sagen was Sinn macht oder nicht.

9.1.3 Interview 3

0:0:14.550 -> 0:0:22.730 Valente de Matos, Ricardo Ich habe mir ein paar Infos geholt. Du warst zumindest mal richtig CC-Leiter oder? 0:0:21.980 -> 0:0:42.770 Kirchner, Lars Ja ich, ich war mal richtig CC Leiter genau. Ich kann zumindest sagen, dass ich sogar mal 2 Kompetenz Center, 1 in München und 1 in Dortmund geleitet habe und für 48 Menschen zuständig war. 0:0:45.580 -> 0:0:51.830 Valente de Matos, Ricardo Ok, das bedeutet aber du machst teilweise noch CC Leitung. Oder gar nicht mehr? 0:0:49.600 -> 0:1:21.730 Kirchner, Lars Nein, nein aus in der Tat persönlichen Gründen habe ich vor 2 Jahren die Entscheidung getroffen, dass ich diese Rolle verlassen muss. 0:1:39.350 -> 0:1:44.510 Valente de Matos, Ricardo Okay, das heißt aber auch du bist aktuell Projektleiter. 0:1:43.640 -> 0:1:47.210 Kirchner, Lars Ich bin aktuell von der Laufbahnstufe höher. Programm Manager das ist formal noch eine Führungslaufbahn oder Führungsrolle. Das ist etwas anderes zumindest in der adesso Welt als Projektleitung. Ich kümmerge mich in der Tat dann in einer Mischform von Projektleitung und Produktunterschiede um interne Projekte. Ich arbeite mit Studierenden zusammen, ich übernehme Angebotsmanagement für große Angebote. Ich vertrete einzelne Themen, bei dem man jemanden braucht, der entsprechend erfahren ist und dann auf der anderen Seite gewisse intellektuelle Fähigkeiten mit sich bringt. Das sind dann aber eher Sonderthemen. Also ein re-

lativ buntes Sammelsurium. Was ich nicht mehr habe ist Personal Verantwortung. Nicht weil ich nicht mit Menschen umgehen kann, sondern sind beispielsweise die sehr verwalterischen Aspekte nicht so angenehm, die nun mal in dieser Rolle mit drinstecken. Ich übernehme auch noch Delivery Management. Das heißt für große Kunden, gibt es eine spezielle Schnittstellen Rolle, die im Prinzip die adesso Organisation vor dem Kunden kapselt, weil der Kunde gar nicht im Detail wissen soll, wie wir aufgebaut sind. Der Kunde hat genau solche Bedarfsanfragen und sagt, ich brauche einen Senior Java Developer mit diesen Fähigkeiten und ich bin dann als Delivery Manager dafür zuständig. 0:6:1.670 → 0:6:7.780 Valente de Matos, Ricardo Dann wäre jetzt meine Frage was die typischen Stakeholder bei einer Erstellung von einer Bedarfsmeldung sind und welche Rolle hast du dabei? 0:6:12.710 → 0:6:31.180

Kirchner, Lars Also Bedarfsmeldungen sind ja ein sehr wichtiges Element in einem der adesso Kerngeschäftsprozesse. Wir sind ein IT-Dienstleister. Das bedeutet, wir entwickeln nicht selber etwas im Sinne von Produkten. Das ist etwas anderes und nennt sich Produkt Geschäft. Das heißt also, wenn wir etwas tun, Entwicklungstätigkeiten aufnehmen, beraten, brauchen wir und benötigen wir immer einen Auftrag. Also einen Kunden. Wenn wir Kundenaufträge haben, weil wir ja ein IT Dienstleister sind, heißt das unsere Kerngeschäftsobjekte sind die Mitarbeitenden, die mit unterschiedlichen Qualifizierungen daneben in genau diesen Kundenaufträgen tätig sind. Das können vom Kunden durchgeführte Projekte sein. Ein Kundenauftrag kann aber auch sein, dass adesso ein Projekt für den Kunden durchführt. Wir müssen wie gesagt diese Projektstellen gezielt besetzen. Das heißt, wir haben immer eine Projektorganisation, mal ist es auch eine Programmorganisation, wo diese Stellen mit entsprechenden rollen oder Stellenanforderungen beschrieben werden. Entweder kommen die vom Kunden direkt, oder wir formulieren die selbst. Diese Beschreibung der Anforderungen, also welche Fähigkeiten und welche Erfahrung und in welchem Umfang Menschen diese mitbringen müssen, damit sie genau diese Projektstelle besetzen können und damit eine ganz bestimmte Rolle in einem Projekt Kontext einnehmen oder in einem Programm Kontext... Diese Beschreibung ist es, was wir als Bedarfsmeldung bezeichnen. Die Quellen dafür sind entweder Kundenorganisationen, dass die dann durch große Accounts wie e-on, die Deutsche Bahn, die ihre ganz eigenen Systeme haben und dann in einer, nicht normierten, aber in einer semi strukturierten Form diese Anforderungen dokumentiert werden. Die gehen dann mehr oder weniger 1 zu 1 an uns über und wir müssen damit weiterarbeiten. Bis dahin, dass wir als Delivery Manager, als Projektleiter, als Programm Manager, als Account Manager oder Vertriebler mit Kunden sprechen, in Projekte reinschauen, Projektorganisationen definieren und selbst in diesen Rollen eine Bedarfsmeldung

erfassen. Es geht allerdings jedes Mal darum, innerhalb eines Projektes oder Programms eine bestimmte Stelle zu besetzen. 0:10:12.650 -> 0:10:20.560 Valente de Matos, Ricardo Welche Art von Projekten hat adesso typischerweise? Hast du da vielleicht ein paar Beispiele? 0:10:26.940 -> 0:10:35.600 Kirchner, Lars Die erstmal abstrakteste und wichtige oder grundsätzliche Unterscheidung ist: Es gibt Projekte, die ein Kunde eine Kundenorganisation aufsetzt und durchführt, an denen wir uns dann beteiligen, indem wir beispielsweise in bestimmte Rollen an bestimmte Stellen dort Menschen reinbringen. Das heißt, wir arbeiten dann aber in einem extern definierten und normalerweise auch gesteuerten und kontrollierten Projektkontext. Das andere ist, wenn wir im Kundenauftrag Projekte aufsetzen. Da könnte man dann auch nochmal differenzieren. Einmal Projekte unter unserer Kontrolle im Kundenauftrag und dann gibt es natürlich auch interne Projekte. Da kann man aber relativ schnell sagen, dann ist es halt eine interne Stakeholder Position. Zum Beispiel kann die HR-Abteilung auch sowas beauftragen. Das unterscheidet sich dann nicht wesentlich. Was allerdings da dann besser in unserer Kontrolle liegt, ist in der Regel, wenn wir selbst die Projekte in der Organisation in der Durchführung verantworten. Dann haben wir auch die Ausprägung der Projekt Organisation, das Vorgehensmodell, usw. in der Hand. Da haben wir sehr viel mehr Flexibilität, was eben die Definition von Rollen, Anforderungen usw. angeht. Das ist so der wesentliche Unterschied in Bezug auf externe Projekte und von uns durchgeführte Projekte. Es gibt dann aber nochmal einen wesentlichen Unterschied in Bezug auf die Vergütung. Wir unterscheiden da grundsätzlich zwischen Time und Material Aufträgen oder Beauftragung und Festpreis Beauftragung. Festpreis bedeutet, dass es von unserer Seite eine bestimmte, klar bemessene und abgegrenzte Leistung angeboten wird und die Auftraggeber Seite verhandelt mit uns dafür einen festgesetzten Preis. Damit muss man bei der Durchführung beachten, dass man nicht so lange vor sich hinarbeiten kann, bis dann zum Beispiel die Auftraggeberseite sagt OK, wir sind jetzt zufrieden. Sondern man hat halt ein beschränktes Budget und in der Regel auch beschränkte Zeit und trägt damit auch ein größeres Risiko. Dann gibt es da entsprechend die timen Material oder t und m Kontexte. Da trägt die Auftraggeberseite in der Regel das größere Risiko, weil wir abstrakt bezeichnet, erstmal nur dazu verpflichtet sind, nach durchschnittlicher Qualität und Güte solche einzelnen Leistungen über zum Beispiel Mitarbeitende beizusteuern, zu erbringen. Trotzdem hat auch da die Erfahrung gezeigt: Wir sind immer an unserer Kunden Zufriedenheit oder langfristigen Kundenbindung interessiert, dass es uns auch gar nicht hilft, wenn wir in einem t und m Kontext nur durchschnittlich oder vielleicht auch mal schlechte Arbeit leisten und dann hinterher sagen, das war jetzt aber gar nicht unsere Verantwor-

tung. Das ist euer Risiko gewesen. Damit kommen wir auch nicht durch. Und es ist in den seltensten Fällen auch so, dass ein Kunde unerschöpfliche Geldmittel hat. Selbst wenn dieser Kunde diese unerschöpflichen Geldmittel hätte, sagen sie aus rein wirtschaftlichen Aspekten natürlich auch zu irgendeinem Zeitpunkt das reicht jetzt, wir möchten nicht noch mehr Geld ausgeben. Das ist doch mal eine grundsätzliche Unterscheidung, was das Bezahlen angeht. Mit hinein spielt auch noch eine Unterscheidung in die Art der Projekte. Das ist nämlich einmal ein Gewerk, wo wir Gewährleistung übernehmen und das auch entsprechend kalkulieren müssen. Ganz häufig werden Gewerke in Kombination mit Festpreisen angeboten und durchgeführt. Interessanterweise müssen sie es aber gar nicht zwingend. Im Umkehrschluss meistens, wenn ich nach t und m arbeite, handelt es sich auch dann von unserer Leistung um Dienstleistung. Das sind aber teilweise, wenn man da wirklich genau draufschaut oder wenn man da juristisch drauf schauen würde feine Unterschiede. Wenn man beispielsweise bei der Beauftragung oder in der Kommunikation ganz bestimmte Begriffe verwendet und es würde irgendwann vor Gericht landen, egal ob man einen Dienstleistungsvertrag abgeschlossen hat und die ganze Zeit meinte auch als Dienstleistungen zu arbeiten, könnte zum Beispiel ein Gericht aufgrund von Formulierungen usw. hinterher feststellen, dass es sich doch um ein Gewerk gehandelt hat. Das sind dann teilweise eher juristische Unterschiede. Am Ende des Tages bedeutet Gewerk natürlich, dass wir an einem Stück Software gearbeitet haben und eben für die konsistente Fehler freie gesamte Funktionalität dann entsprechende Gewährleistung anbieten und übernehmen. Also damit der Kunde für das Geld, dass die Organisation gezahlt hat, eine einsetzbare Lösung bekommt. Wohingegen bei Dienstleistung eine gar nicht funktionale Lösung bei rauskommen muss. Beispiel typischer und klassischer Bereich für Dienstleistungsgeschäft ist ja Consulting. Consulting bedeutet qualifizierte, erfahrene Menschen zu bestimmten Themen kommunizieren, was ausarbeiten, etwas aufschreiben, können etwas spezifizieren können. Aber letztendlich ist da die menschliche Arbeit beziehungsweise der Erkenntnisgewinn im Zentrum und es wird keine Lösung in dem Sinne geschaffen und bereitgestellt. Das sind die für mich zumindest relevanten Unterscheidung. 0:18:3.520 -> 0:18:16.250 Valente de Matos, Ricardo Wie werden dann Bedarfsmeldungen und die Anforderungen typischerweise bei Adesso kommuniziert und auch dokumentiert. Gibt es einen Ablauf, wie das genau gehandhabt wird? 0:18:23.200 -> 0:18:36.560 Kirchner, Lars Für den Staffing-Prozess an sich gibt es einen definierten Ablauf, der aber in Großteilen aus manueller Arbeit und manuellem Arbeitseinsatz besteht. Es gibt keine normierte Form einer Bedarfsmeldung. Es hat sich allerdings zumindest eine grobe thematische Struktur etabliert. Das bedeutet in der Regel hat man so etwas wie eine Überschrift

und Bezeichnung. Dann hat man in der Regel einen Bereich, der den Einsatz Kontext ein wenig allgemeiner beschreibt. Dann hat man einen Bereich, der auf die individuell geforderten Fähigkeiten und Erfahrungen eingeht. Man hat normalerweise eine Gewichtung dieser Skills. Das bedeutet in Bezug auf Expertise, die da erwartet wird oder mal solche Dinge wie Primary in Secondary usw. Unterteilungen. Wir haben in diesen Bedarfsmeldungen dann aber auch wirtschaftlich relevante Informationen damit verknüpft. Das ist typisch für ein Dienstleistungsunternehmen. Das bedeutet uns interessiert dann die vertraglichen Konditionen im Sinne von Tagessatz, was für eine Organisation für adesso, im Sinne von Dienstleistung auch absolut relevant ist. Das sind dann immer diese Parameter. Ab wann ist der Einsatz gewünscht? Und für wie lange? Weil wir immer prüfen müssen, selbst wenn wir beispielsweise sehr gut fachlich passende Mitarbeitende finden, sind sie aber eventuell schon in anderen Projekten eingesetzt und können dementsprechend gar nicht dort leisten. Es mag dann je nach Kundenkontext noch weitere Informationsblöcke geben. Wir arbeiten mittlerweile auch mit einem Leveling-System, weil man sich als Organisation oder eben als Mensch innerhalb einer Organisation ein bisschen besser in Abfragen organisieren kann, wenn an solchen detaillierteren Bedarfsmeldungen gewisse Labels dran stehen wie beispielsweise Azure als Technologie oder oder Java. Dann kann ich eine potenziell größere Menge von Bedarfsmeldungen, die ich manuell durchsuche besser filtern und einschränken. Was in Bezug auf unserem Staffing-Prozess sehr relevant ist, ist beispielsweise dann der Status einer solchen Bedarfsmeldung, weil die Organisationen nicht alle Bedarfsmeldungen interessiert. Manche Bedarfsmeldungen sind schon verarbeitet und erfüllt. Manche sind nur dokumentiert und sollen aber gar nicht weiter beachtet werden und erst wenn sie beispielsweise bei uns eskaliert werden, sollte eigentlich die relevante Organisation darauf schauen. Wir haben eine vertikale Einteilung des Unternehmens in Branchen, bzw. technologisch getriebenen Organisationseinheiten. Das sind unsere sogenannten Line of Business. Mittlerweile sind das sogar Business Areas und es ist an der Stelle durchaus relevant, welche adesso Organisationseinheit für so eine Business Area oder Line of Business zuständig ist. Wenn wir an der Stelle mit einem gewissen Vorkaufsrecht feststellen, dass wir es nicht bedienen können, werden natürlich alle gefragt und es dürfen beispielsweise aus der Line of Business Motiv selbstverständlich auch mitarbeiten in Projekten der Line Cross Industries und umgekehrt getätigt werden. Aber umso mehr ist es wichtig, dass dann die jeweils geforderten Skills und Erfahrungen und wirtschaftlichen Konditionen möglichst präzise beschrieben werden, damit diese Fragen die aufkommen nicht immer wieder gestellt werden und immer wieder beantwortet werden müssen. Das ist aber, wie gesagt schon Kern Staffing-Prozess. 0:24:4.210 -> 0:24:10.830

Valente de Matos, Ricardo Du hast auch schon einige Punkte in Richtung Verfügbarkeit genannt, dass das sehr wichtige Aspekte sind. Gibt es besonders wichtiger oder auch unverzichtbare Punkte in einer Bedarfsmeldung, die in jeder drin sein sollte? 0:24:22.530 → 0:24:23.920 Kirchner, Lars Das kommt auf die Perspektive an. Wenn ich jetzt eine rein fachliche Perspektive einnehme, ist es da unverzichtbar, dass aufgelistet oder benannt wird, welche Fähigkeiten oder Skills konkret mitgebracht werden müssen und idealerweise auch in welcher Erfahrung und Güte das ist. Aus der Perspektive zum Beispiel des Dienstleistungsunternehmens adesso ist relevant, dass ein Beginn des Einsatzes, ein voraussichtlicher Einsatzzeitraum, ein Tagessatz dran steht. Dass da dran steht, ob Freelancer, ob Smartphone, smartshore Fähigkeit gegeben ist oder near Shore, oder wie die sprachliche Ausrichtung ist. Also ob das zum Beispiel deutschsprachig ist, ob Englisch. Entsprechend zulässige Kommunikationsmittel sind wichtige Zusatzinformationen, die genau darauf abzielen, dass alle Menschen, die potenzielle Kandidaten/Kandidatinnen auf so eine Bedarfsmeldung gefiltert werden. Wenn es nämlich auf der Ebene, dass die Verfügbarkeit aber auch darüberhinausgehend nicht passt oder wenn es zum Beispiel von einem Tagessatz, aus welchen Gründen auch immer sehr unattraktiv ist, es dazu führen kann, dass dann die Organisation bestimmte Personen nicht anbietet. Das ist aber wie gesagt eine Perspektive, die durch die Natur von adesso als Dienstleister, und wir sind ein auslastungsgetriebenes Unternehmen, was halt auch versucht wirtschaftlich zu arbeiten, darüber reinkommen. Idealerweise möchte zum gewünschten Einsatzbeginn auch dann wirklich eine Person nicht nur theoretisch benannt, sondern idealerweise interviewt, geprüft, eingeführt und wie auch immer wurde und dann wirklich los laufen kann. 0:28:46.70 → 0:28:51.420 Valente de Matos, Ricardo Gibt es Herausforderungen und Schwierigkeiten bei unklaren oder unvollständigen Bedarfsmeldungen? Hast du da Erfahrungen machen können? 0:28:51.490 → 0:29:23.70 Kirchner, Lars Ja. Bei einem Auftraggeber handelt es sich in der Regel um Branchen, die eben nicht als Kerngeschäft IT Software Entwicklung betreiben. Dementsprechend fällt es ihnen teilweise schwer, präzise Bedarfsmeldungen zu formulieren, die inhaltlich genau das transportieren oder beinhalten oder umfassen, was sie eigentlich an Fähigkeiten und Erfahrung benötigen. Das liegt einfach daran, dass sie in Teilen nicht für jeden Kunden gilt, aber in Teilen den Prinzip Menschen etwas beschreiben lassen, die davon nicht wirklich Ahnung haben. Das wiederum führt dazu oder kann dazu führen, dass der Kunde etwas anfordert, was der Kundenorganisation, dem Projekt oder wie auch immer im schlimmsten Falle sogar gar nicht hilft. Oder aber das ist eine andere Ausprägung, dass dann Kombinationen von Erfahrung und Fähigkeiten gesucht werden, die es in der realen Welt einfach so nicht gibt. Das ist die berühmte

eierlegende Wollmilchsau, wo man drauf schaut und sagt, diese Menschen hätten wir auch gerne als Mitarbeitende. Vielleicht gibt es auf diesem Planeten auch eine Handvoll davon, aber das ist unrealistisch. Und das kann wie gesagt darin begründet sein, dass auf der auftraggebenden Seite jetzt Menschen damit beauftragt werden, solche Bedarfsmeldungen zu formulieren, die das nicht wirklich können. Eigentlich haben wir an der Stelle schon immer einen idealerweise von uns moderierten Prozess. Deswegen macht zum Beispiel auch die Rolle Delivery Management Sinn. Das hängt wie gesagt auch sehr mit der insgesamten Qualifikation oder Qualität von der Auftraggeberseite in diesem Kontext zusammen. Bei manchen Kundenorganisationen ist das trotzdem sehr gut eingespielt und etabliert und funktioniert auch so. Bei manchen muss man da früh ansetzen und sagen wir sprechen miteinander, und ich arbeite dann beispielsweise in diesem Gespräch heraus, was der Kunde wirklich benötigt, was abgrenzbar eine sinnvolle Bedarfsmeldung ist, was es für eine Stelle beinhaltet und womit wir dann weiterarbeiten können. Das ist oder kann ein Problem sein, wenn wir nicht entsprechend damit umgehen. Nicht nur die fachliche Qualität oder auch die die Passgenauigkeit, die wir dort anbieten können ist da wichtig, sondern auch vor allem die Schnelligkeit vom Staffing-Prozess. Bedeutet, wenn ein Unternehmen sich so aufgestellt hat, dass eine Bedarfsmeldung, die rein kommt innerhalb von sehr kurzer Zeit bearbeitet wird, sagen wir mal 2 Stunden, dann habe ich einen klaren Vorteil gegenüber zum Beispiel einem Anbieter, der dafür 2 Tage braucht oder eine Woche oder 2 Wochen. Denn für die Auftraggebende Seite ist klar, wenn ich dann Rückmeldungen bekomme und ich schaue da rein und diese Rückmeldungen sind plausibel. . . Was hält mich davon ab, dann zu sagen ich beauftrage ihn jetzt. Es muss nicht perfekt sein, aber wenn es plausibel ist und die Konditionen sind gut, dann ist der Auftrag ausgesprochen und die anderen gehen logischerweise leer aus. Wenn wir eben entsprechend in unserem Matching entweder nicht gut arbeiten oder überfordert sind und beispielsweise einen Software Architekt für Java mit bestimmten weiteren Anforderungen gefordert ist und wir bieten da ein Profil drauf an, wo man darauf schaut und feststellt, dass es ein .Net Software Developer ist, ist es bei diesem extremen Beispiel so, dass es glücklicherweise auch sofort auffällt. Aber das würde natürlich dann zu Irritationen auf Kunden-seite führen. Bedeutet: Wir sollten nicht nur dafür sorgen, dass die Erwartungen oder Anforderungen möglichst gut an der Realität sind, sondern wenn wir dann wiederum auch da etwas matchen und einreichen, dass das dann auch diesen Anforderungen nach Möglichkeit entspricht und, dass innerhalb von möglichst kurzer Zeit so geschieht. Und die unglücklichste Variante ist, wenn man ein Prozess durchläuft und man bietet da jemanden an und der wird sogar genommen und wird eingear-

beitet und dann stellt man irgendwie fest der macht irgendwie Unsinn. Man hat nie eine Garantie. Letztendlich sind es Menschen die dort arbeiten. 0:39:18.140 -> 0:39:22.580 Valente de Matos, Ricardo Welche Auswirkungen haben unklare oder auch fehlende Informationen in Bedarfsmeldungen jetzt aber konkret in Bezug auf die Effizienz und den Erfolg von Projekten. 0:39:27.60 -> 0:39:36.170 Kirchner, Lars Im Idealfall wird, möglichst früh erkannt, dass eine Bedarfsmeldung lückenhaft, unpräzise wie auch immer formuliert ist. Dann muss nachgefragt werden. Ich schaue mir etwas an, versuche zu verstehen, was die andere Seite sucht und wenn das für mich dann nicht konsistent auf die zumindest für mich bekannten Rollenlösung ist, dann muss ich nachfragen. Alles andere ist eine Interpretation. Dann läuft man mit sehr großer Wahrscheinlichkeit in die von mir gerade beschriebenen Probleme rein. Die schlechteste Variante ist, dass ich sage ich nehme die Informationen, die ich jetzt da vorgelegt bekommen habe, interpretieren sie nach besten Wissen und Gewissen und dann ist es aber mehr oder weniger ein Glücksspiel. Das heißt, wenn ich dann Profile finde, könnten sie immer noch zufällig das sein, was der Kunde eigentlich wollte und gesucht hat. 0:40:49.150 -> 0:41:1.310 Valente de Matos, Ricardo Die letzte Frage hast du im Grunde auch schon mit beantwortet. Wie könnte man sicherstellen, dass Bedürfnisse und Anforderungen aller relevanten Stakeholder in einer Bedarfsmeldung Berücksichtigt werden? 0:41:4.190 -> 0:41:20.600 Kirchner, Lars Zwei Möglichkeiten. Man könnte, da glaube ich aber nicht dran, natürlich den Prozess standardisieren und stark formalisieren. Also das im Prinzip von einer öffentlichen Stelle aus gesagt wird: Alle Dienstleister und Auftraggeber dieser Welt wenn ihr in dieser Art Geschäft betreiben wollt, müsst ihr so ein Format einreichen. Also Bürokratie pur. Das würde uns nichts verbessern, aber das ist eine Möglichkeit. Die andere Möglichkeit ist meines Erachtens, dass dann für genau solche Prozesse entsprechend versierte Menschen diesen Prozess, das heißt die Anforderungserhebung, die Dokumentation, das erfüllen diese Anforderungen komplett begleiten und moderieren. Das ist Delivery Management. Das ist Business Development. Manchmal auch Account Management. 0:42:26.240 -> 0:42:30.90 Valente de Matos, Ricardo Dann sind wir eigentlich schon durch mit den Fragen. Hast du noch zu irgendeinem Punkt irgendwelche Fragen oder irgendwas, was vielleicht noch für mich in dem Themenbereich interessant sein könnte? 0:43:1.90 -> 0:43:6.710 Kirchner, Lars Die größte Schwierigkeit liegt darin, dass wir Informationen über eine natürliche Sprache transportieren. Das ist zwar flexibel, weil die Sprache an der Stelle ja eben nicht formalisiert ist. Was aber immer das Risiko mit sich bringt, dass etwas nicht präzise beschrieben, abgegrenzt oder interpretierbar wird und genau bei solchen Prozessen, wo wir eigentlich präzise arbeiten wollen, haben wir genau diese große Herausfor-

derung, dass die bisher benutzte Art, um diese Informationen zu ermitteln und die gerade wieder zu lesen zu interpretieren eben ein Stück weit ungenügende Mittel, nämlich den natürlichen sprachigen Raum verwendet. Da ist zwar mit gesundem Menschenverstand gearbeitet worden. Das heißt, es haben sich Semistrukturen gebildet. Aber es ist wirklich sehr individuell unterschiedlich in was für einer Qualität oder was für einer Realitätsnähe solche Bedarfsmeldungen formuliert werden. Wenn wir auf unserer Seite jemanden sitzen haben, der oder die eben auch in diesem Umfeld relativ wenig Ahnung und Erfahrung hat, dann haben wir auch nochmal ein Risiko, dass selbst wenn die Bedarfsmeldungen präzise und realitätsnah formuliert ist, bei uns in der Interpretation etwas schiefgeht. Das bedeutet, dass halt die Menschen, die auf unserer Seite Bedarfsmeldungen lesen und versuchen zu bedienen, leider auch hinreichend viel Erfahrung in der Projekt IT haben müssen.

Eigenständigkeitserklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit selbständig angefertigt und mich keiner fremden Hilfe bedient sowie keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Alle Stellen, die wörtlich oder sinngemäß veröffentlichten oder nicht veröffentlichten Schriften und anderen Quellen entnommen sind, habe ich als solche kenntlich gemacht.

Diese Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

9.2 Erklärung zu eingesetzten Hilfsmitteln

1. Korrekturservice der Fachhochschule bzw. des Fachbereichs genutzt:

☐ Ja
☒ Nein

2. Einsatz eines externen (kommerziellen) Korrekturservice:

☐ Ja
☒ Nein

3. Folgende Personen haben die Arbeit zusätzlich Korrektur gelesen:

- ...

4. Nutzung von Sprachmodellen für die Texterstellung (z.B. ChatGPT), wenn ja, welche und in welchen Abschnitten:

☐ Ja
☒ Nein

5. Sprachübersetzungstools (z.B. Google Übersetzer, DeepL), wenn ja, welche und in welchen Abschnitten:

☒ Ja
☐ Nein

- DeepL, Im Kapitel Literaturüberblick für das bessere Verständnis der Literatur
6. Einsatz von Software zur Sprachkorrektur (z.B. Grammarly), wenn ja, welche und in welchen Abschnitten:
- ☐ Ja
 - ☐ Nein
 - ...
7. Einsatz anderer Hilfsmittel:
-
8. Ich stimme dem möglichen Einsatz von Software zur Plagiatserkennung zu:
- ☒ Ja
 - ☐ Nein

Ich bestätige, dass obige Aussagen vollständig und nach bestem Wissen ausgefüllt wurden.

Dortmund, den 7. Mai 2024

Ricardo Valente de Matos

Literatur

- [1] S. A. Alasadi und W. S. Bhaya, “Review of data preprocessing techniques in data mining,” *Journal of Engineering and Applied Sciences*, Jg. 12, Nr. 16, S. 4102–4107, 2017.
- [2] P. Bafna, D. Pramod und A. Vaidya, “Document clustering: TF-IDF approach,” in *2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)*, IEEE, 2016, S. 61–66.
- [3] N. J. Belkin und W. B. Croft, “Information filtering and information retrieval: Two sides of the same coin?” *Communications of the ACM*, Jg. 35, Nr. 12, S. 29–38, 1992.
- [4] T. Bohne und U. M. Borghoff, “Data fusion: Boosting performance in keyword extraction,” in *2013 20th IEEE International Conference and Workshops on Engineering of Computer Based Systems (ECBS)*, IEEE, 2013, S. 166–173.
- [5] R. Burke, A. Felfernig und M. H. Göker, “Recommender systems: An overview,” *Ai Magazine*, Jg. 32, Nr. 3, S. 13–18, 2011.
- [6] G. Chartron und G. Kembellec, “General introduction to recommender systems,” *Recommender Systems*, S. 1–23, 2014.
- [7] M. Chiny, M. Chihab, O. Bencharef und Y. Chihab, “LSTM, VADER and TF-IDF based hybrid sentiment analysis model,” *International Journal of Advanced Computer Science and Applications*, Jg. 12, Nr. 7, 2021.
- [8] W. B. Croft, “Combining approaches to information retrieval,” in *Advances in Information Retrieval: Recent Research from the center for intelligent information retrieval*, Springer, 2000, S. 1–36.
- [9] R. Darmawan und R. S. Wahono, “Hybrid Keyword Extraction Algorithm and Cosine Similarity for Improving Sentences Cohesion in Text Summarization,” *Journal of Intelligent Systems*, Jg. 1, Nr. 2, S. 109–114, 2015.
- [10] Z. Dong, Z. Wang, J. Xu, R. Tang und J. Wen, “A brief history of recommender systems,” *arXiv preprint arXiv:2209.01860*, 2022.

-
- [11] T. Dresing und T. Pehl, *Praxisbuch Interview, Transkription & Analyse: Anleitungen und Regelsysteme für qualitativ Forschende*. dr dresing & pehl GmbH, 2015.
 - [12] A Famili, W.-M. Shen, R. Weber und E. Simoudis, “Data preprocessing and intelligent data analysis,” *Intelligent data analysis*, Jg. 1, Nr. 1, S. 3–23, 1997.
 - [13] D Frank Hsu und I. Taksa, “Comparing rank and score combination methods for data fusion in information retrieval,” *Information retrieval*, Jg. 8, Nr. 3, S. 449–480, 2005.
 - [14] S. García, S. Ramírez-Gallego, J. Luengo, J. M. Benítez und F. Herrera, “Big data preprocessing: methods and prospects,” *Big Data Analytics*, Jg. 1, Nr. 1, S. 1–22, 2016.
 - [15] R. Horesh, K. R. Varshney und J. Yi, “Information retrieval, fusion, completion, and clustering for employee expertise estimation,” in *2016 IEEE International Conference on Big Data (Big Data)*, IEEE, 2016, S. 1385–1393.
 - [16] N. Jain und V. Srivastava, “Data mining techniques: a survey paper,” *IJRET: International Journal of Research in Engineering and Technology*, Jg. 2, Nr. 11, S. 2319–1163, 2013.
 - [17] S. Jun Lee und K. Siau, “A review of data mining techniques,” *Industrial Management & Data Systems*, Jg. 101, Nr. 1, S. 41–46, 2001.
 - [18] P. Kroha, “Preprocessing of requirements specification,” in *Database and Expert Systems Applications: 11th International Conference, DEXA 2000 London, UK, September 4–8, 2000 Proceedings 11*, Springer, 2000, S. 675–684.
 - [19] D. Kumawat und V. Jain, “POS tagging approaches: A comparison,” *International Journal of Computer Applications*, Jg. 118, Nr. 6, 2015.
 - [20] C. Lanquillon, “Enhancing text classification to improve information filtering,” Diss., Otto-von-Guericke-Universität Magdeburg, Universitätsbibliothek, 2001.
 - [21] M. Lavin, “Analyzing documents with TF-IDF,” 2019.
 - [22] J. Lu, Q. Zhang und G. Zhang, *Recommender systems: advanced developments*. World Scientific, 2020.
 - [23] M. Maguire und N. Bevan, “User requirements analysis: a review of supporting methods,” in *IFIP World Computer Congress, TC 13*, Springer, 2002, S. 133–148.

-
- [24] P. Majumder, M. Mitra und B. Chaudhuri, “N-gram: a language independent approach to IR and NLP,” in *International conference on universal knowledge and language*, Bd. 2.
 - [25] A. Mansouri, L. S. Affendey und A. Mamat, “Named entity recognition approaches,” *International Journal of Computer Science and Network Security*, Jg. 8, Nr. 2, S. 339–344, 2008.
 - [26] R. Mihalcea und P. Tarau, “TextRank: Bringing order into text,” in *Proceedings of the 2004 conference on empirical methods in natural language processing*, 2004, S. 404–411.
 - [27] D. Nadeau und S. Sekine, “A survey of named entity recognition and classification,” *Linguisticae Investigationes*, Jg. 30, Nr. 1, S. 3–26, 2007.
 - [28] T. Nakagawa und K. Uchimoto, “A hybrid approach to word segmentation and pos tagging,” in *Proceedings of the 45th annual meeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions*, 2007, S. 217–220.
 - [29] E. Partalidou, E. Spyromitros-Xioufis, S. Doropoulos, S. Vologianidis und K. Diamantaras, “Design and implementation of an open source Greek POS Tagger and Entity Recognizer using spaCy,” in *IEEE/WIC/ACM International Conference on Web Intelligence*, 2019, S. 337–341.
 - [30] T. Pay, S. Lucci und J. L. Cox, “An ensemble of automatic keyword extractors: TextRank, RAKE and TAKE,” *Computación y Sistemas*, Jg. 23, Nr. 3, S. 703–710, 2019.
 - [31] M. Pfadenhauer, *At eye level: the expert interview—a talk between expert and quasi-expert*. Springer, 2009, S. 81–97.
 - [32] S. Pirk, “Implementierung und Visualisierung N-Gramm-basierter Word-Clouds,” B.S. thesis, 2019.
 - [33] J. Ramos u. a., “Using tf-idf to determine word relevance in document queries,” in *Proceedings of the first instructional conference on machine learning*, Citeseer, Bd. 242, 2003, S. 29–48.
 - [34] F. Ricci, *Recommender Systems: Models and Techniques*. 2014.
 - [35] M. A. Shafi’I, M. S. Abd Latiff, H. Chiroma u. a., “A review on mobile SMS spam filtering techniques,” *IEEE Access*, Jg. 5, S. 15 650–15 666, 2017.

- [36] V Suhasini und N Vimala, “A Hybrid TF-IDF and N-Grams Based Feature Extraction Approach for Accurate Detection of Fake News on Twitter Data,” *Turkish Journal of Computer and Mathematics Education*, Jg. 12, Nr. 6, S. 5710–5723, 2021.
- [37] M. Zhang, X. Li, S. Yue und L. Yang, “An empirical study of TextRank for keyword extraction,” *IEEE access*, Jg. 8, S. 178 849–178 858, 2020.