

Unterstützung von IT-Prozessen durch die Anwendung und Implementierung von Natural Language Processing

Stefan Kert



Information Engineering and Management
University of Applied Sciences Upper Austria

11th June, 2017

Inhaltsverzeichnis

1	Einleitung	1
1.1	Motivation	1
1.2	Ziel und Methodik	2
2	Natural Language Processing	4
2.1	Überblick	4
2.2	Verarbeitung und Interpretation von natürlicher Sprache	4
2.2.1	Eingabe von Sprache und Vorverarbeitung	4
2.2.2	Vorverarbeitung und Information Extraction	5
2.2.3	Interpretation	6
2.2.4	Ausgabe und Ergebnis	6
2.3	Anwendungen von Natural Language Processing	6
2.3.1	Chatbots	6
2.3.2	Spam-Filter	7
2.3.3	Sprachassistenten	7
3	Anwendung von NLP in IT-Prozessen	9
3.1	Allgemeines	9
3.2	Das Beispielunternehmen	9
3.3	Supportprozess	10
3.3.1	Ablauf	10
3.3.2	Verbesserung 1 - Kategorisierung der Anfrage	12
3.3.3	Verbesserung 2 - Hinzufügen von Informationen aus der Wissensbasis	13
3.3.4	Verbesserung 3 - Erkennen von duplizierten Tickets	15
3.4	Neukundenprozess	16
3.4.1	Ablauf	16
3.4.2	Verbesserung 1 - Automatisches Auslesen der Informationen aus der Email	17
3.4.3	Verbesserung 2 - Sammeln von Informationen	18
4	Fazit	20
	Quellenverzeichnis	21
	Literatur	21

Kapitel 1

Einleitung

1.1 Motivation

Natural Language Processing ist die Möglichkeit der maschinellen Verarbeitung und Interpretation von natürlicher Sprache. Dabei beschränken sich die Anwendungen von *NLP* nicht ausschließlich auf geschriebene Sprache in Form von Wörtern und Texten, sondern auch auf Lautsprache (gesprochene Sprache). Bereits seit mehreren Jahren wird im privaten Bereich auf sogenannte Sprachassistenten wie Cortana (Microsoft), Siri (Apple) und Amazon Echo (Amazon) gesetzt. Diese ermöglichen es, verschiedene Aufgaben mittels Lautsprache zu erledigen. Dabei sind Anwendungen wie z.B. das Ein und Ausschalten von Licht, das Abspielen von Playlists oder auch das Starten eines Anrufs nur einige der vielfältigen Anwendungsgebiete der Spracherkennung. Im Zusammenhang mit geschriebenen Wörtern und Texten wird *NLP* vor allem bei der Interpretation von Texten eingesetzt. Ein Beispiel für die Anwendung von *NLP* ist die Klassifizierung von Texten. Dies findet vor allem bei wissenschaftlichen Arbeiten Anwendung. Ein weiteres Anwendungsgebiet im Zusammenhang mit Texten ist das Zusammenfassen oder die semantische Überprüfung der Inhalte. Vor allem in den letzten Jahren hat das Thema Natural Language Processing im Zusammenhang mit sozialen Netzwerken an Bedeutung gewonnen. Eine Anwendung, welches in letzter Zeit sehr stark im Fokus von Behörden steht, ist die Erkennung von sogenannten *Hasspostings* [NTC16]. Durch die Masse an Postings die täglichen in den sozialen Medien getätigt werden ist es nahezu unmöglich alle strafrechtlich relevanten Beiträge herauszufinden und diese gegebenenfalls zu löschen oder auch weitere rechtliche Schritte einzuleiten. Ein weiteres Anwendungsgebiet im Zusammenhang mit Postings in sozialen Medien ist das Analysieren dieser Postings zu Werbe- oder Marketingzwecken (User A postet, dass er gerne wieder auf Urlaub fahren würde, bekommt gezielt Werbung zu Urlaubsangeboten). Vor allem auf Grund des sehr werbelastigen Geschäftsmodells ist diese Analyse ein Thema von hoher Relevanz für die Betreiber von Werbeplattformen bzw. sozialen Medien wie *Facebook* oder *Twitter*.

Ein Gebiet, in dem die Möglichkeiten der Texterkennung sowie der Spracherkennung sehr stark zu einer Verbesserung beitragen könnten, sind IT-Prozesse mit einem starken Fokus auf zwischenmenschliche Kommunikation. In diesen kommu-

nikationslastigen IT-Prozessen werden sehr viele Daten in Form von menschlicher Sprache erzeugt. Diese kann entweder in Form von Texten (E-Mail, Dokumentation, Defect-Beschreibung), oder in Form von Sprache (Telefonanruf, Mailboxnachricht) vorliegen. Ein klassisches Beispiel für einen IT-Prozess bei dem die Daten in Form von menschlicher Sprache vorliegen sind Support-Prozesse.

Ein konkretes Beispiel für die Vorselektierung bzw. Vorsortierung von Daten mit der Unterstützung von Natural Language Processing ist die Erkennung von Spam-Mails [Din14]. Beim Erkennen von Spam-Mails gibt es triviale Vorgehensweisen wie z.B. Blacklists mit Hilfe derer an Hand des Absenders festgestellt wird ob es sich um eine Spam-Mail handelt oder nicht. Komplexere Verfahren wären zum Beispiel die Analyse des Textes auf spezielle Schlagwörter oder auch den semantischen Inhalt des Textes mit Hilfe von Natural Language Processing.

Allgemein lässt sich sagen, dass der richtige und gezielte Einsatz von Natural Language Processing entscheidend zum Unternehmenserfolg beitragen kann.

Einige IT-Prozesse werden bereits rudimentär durch Natural Language Processing unterstützt, andere wiederum könnten von der Anwendung sehr profitieren. Ziel dieser Arbeit ist es, an Hand ausgewählter IT Prozesse (z.B. Support-Prozess, Anfrage eines Neukunden) beispielhaft zu erläutern, wie diese Prozesse durch die Anwendung von Natural Language Processing verbessert werden könnten. Vor allem in Prozessen, in denen die direkte Kommunikation zwischen Menschen eine wichtige Rolle spielt, kann durch die Verarbeitung und Auswertung der entstehenden Daten (textuell oder Sprache) eine Verbesserung erzielt werden. Der Fokus dieser Arbeit liegt dabei darauf, dass die Personen, die Teil dieses Prozesses sind, entlastet bzw. unterstützt werden. Dies kann die Mitarbeiter des Unternehmens betreffen, welche die Serviceleistung in Form eines Prozesses anbietet, aber auch die Kunden dieses Unternehmens durch eine möglicherweise schnellere Abwicklung dieses Prozesses (schneller bearbeitete Support-Anfragen, Chat-Bot statt FAQ) unterstützen und schlussendlich die Zufriedenheit auf beiden Seiten zu erhöhen.

1.2 Ziel und Methodik

Der Inhalt dieser Arbeit besteht in erster Linie aus der Analyse ausgewählter IT-Prozesse und dem Erläutern der Möglichkeiten der Anwendung von Natural Language Processing zur Unterstützung dieser.

Für die Analyse werden die IT-Prozesse eines klassischen Softwareentwicklungsbetriebes verwendet. Diese Prozesse werden kurz analysiert und stellen danach den IST Zustand dar. Es wird ein kurzer Überblick über den aktuellen Prozess gegeben und anschließend erläutert, welche Daten in diesem Prozess zwischen den einzelnen Schritten erzeugt bzw. verarbeitet werden. Danach wird gezeigt, wie mit Hilfe von Natural Language Processing eine Vorverarbeitung, Optimierung bzw. Erweiterung dieser Daten erfolgen kann. Am Ende sollte ein neuer Prozess skizziert werden, der durch die Unterstützung von Natural Language Processing verbessert wurde.

Ein wichtiger Punkt, der bei der Analyse miteinbezogen wird, ist die Machbarkeit der einzelnen Verbesserungen, sowie der Limitierungen die sich mit der Anwendung ergeben könnten. Diese Limitierungen bzw. die Machbarkeit sollte in eine abschließende Bemerkung mit einfließen und Aufschluss darüber geben, unter

welchen Voraussetzung NLP eingesetzt werden könnte.

Kapitel 2

Natural Language Processing

2.1 Überblick

In diesem Kapitel wird ein grober Überblick über den Ablauf beim *NLP* gegeben und außerdem einige Beispiele gezeigt, die bereits im IT-Umfeld eingesetzt werden.

2.2 Verarbeitung und Interpretation von natürlicher Sprache

Grundsätzlich ist Basis jeder *NLP-Engine* die Verarbeitung und Interpretierung von geschriebener Sprache. Dabei ist zu unterscheiden, aus welcher Quelle der Input für die Verarbeitung kommt. Meist wird hier zwischen zwei Teilgebieten unterschieden:

- Maschinenlesbare Texte
- Spracheingabe

Diese beiden Teilgebiete unterscheiden sich darin, dass bei der Spracheingabe noch eine Vorverarbeitung stattfinden muss, sodass der gesprochene Text interpretiert werden kann. Sobald der gesprochene Text in Form eines maschinenlesbaren Textes vorliegt, kann mit der Weiterverarbeitung begonnen werden. Dies bedeutet im Umkehrschluss, dass beide Teilgebiete auf der gleichen Engine basieren können, sich jedoch in ihrer Komplexität und der Möglichkeit der Eingabe unterscheiden.

Im Folgenden wird kurz erläutert, wie der Ablauf einer Analyse im *NLP* vom Input bis zum Output abläuft.

2.2.1 Eingabe von Sprache und Vorverarbeitung

Wie bereits erläutert, kann die Eingabe von Sprache für das *NLP* auf mehrere verschiedene Arten erfolgen. Es gibt die Möglichkeit, über ein Mikrofon Lautsprache aufzunehmen, oder verschiedene Texte als Input zu verwenden. Dabei könnten zum Beispiel Texte automatisch von Internetseiten ausgelesen, oder auch Dokumente am Rechner durchsucht werden. Eine weitere Möglichkeit für textuellen Input lässt sich in Form eines Chats finden. Hier findet eine textuelle Eingabe statt, die später von einer Engine verarbeitet werden kann.

Prinzipiell unterscheiden sich diese beiden Eingabemöglichkeiten sehr stark in ihrer Komplexität. Bei der Eingabe über Lautsprache muss beim Einlesen nicht nur auf das Erkennen der Wörter, sondern auf eine etwaige Fehlerkorrektur durch Unterbrechungen oder Nebengeräusche geachtet werden. Durch diese Probleme gestaltet sich die Spracherkennung als sehr schwierig, wodurch die Möglichkeiten der Anwendung sehr stark eingeschränkt werden. Nicht zuletzt durch die Tatsache, dass in den letzten Jahren sehr viel auf dem Gebiet der Spracherkennung geforscht wurde, konnte im vergangenen Jahr ein Erfolg auf diesem Gebiet erreicht werden. Einem Forschungsteam der Firma Microsoft ist es gelungen, eine *Word error rate* von 5,9% zu erreichen, was ungefähr der Fehlerrate entspricht, die bei der normalen zwischenmenschlichen Kommunikation auftritt [W X+16].

2.2.2 Vorverarbeitung und Information Extraction

Sobald Text in maschinenlesbarer Form vorliegt (z.B. ASCII-Code) kann auf den syntaktischen bzw. den semantischen Inhalt geprüft werden. Dazu gibt es verschiedene Möglichkeiten für die sogenannte *Information Extraction*. Ein Beispiel dafür ist die sogenannte *Named Entity Recognition (NER)*. *NER* ermöglicht das Verarbeiten der Texte und liefert Informationen über den Inhalt der Texte. Ergebnis der *NER* sind die sogenannten *Entities*. Es handelt sich dabei um Personen oder Objekte, welche in dem zu analysierenden Text enthalten sind. Es werden außerdem Attribute der *Entities*, wie zum Beispiel dem Alter, extrahiert. Für eine genauere Analyse können auch Beziehungen zwischen *Entities* oder Ereignisse in welche *Entities* einbezogen sind ausgelesen werden. Diese Informationen können für die weitere Verarbeitung und Interpretation verwendet werden und je nach Genauigkeit der ausgelesenen Daten kann die Qualität der Aussage die über den Inhalt getroffen werden kann variieren. In der untenstehenden Grafik ist der Ablauf eines *NER* Zyklus dargestellt.

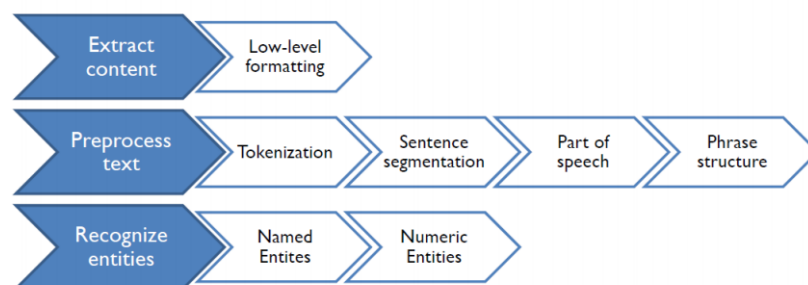


Abbildung 2.1: Named Entity Recognition Prozess

Der in Abbildung 2.1 dargestellte NER Prozesses zeigt sehr gut, wie der normale Ablauf bei der Texterkennung erfolgt. Zuerst wird der Text verarbeitet, dazu werden unnötige Zeichen entfernt; Bilder, Werbung und andere Objekte mit keiner semantischen Bedeutung werden ebenfalls entfernt. Danach wird der Text vorverarbeitet. Es werden dazu einzelne Tokens gebildet, wobei jeder Token ein Wort repräsentiert. Diese Tokens werden wiederum zu Sätzen zusammengefasst, sodass

die semantische Bedeutung nicht verloren geht. Sobald Sätze gebildet wurden, wird festgestellt, welche verschiedenen Wortarten die Tokens darstellen und die Tokens bekommen ein Label mit der jeweiligen Wortart. Dies dient vor allem der besseren semantischen Interpretation. Im letzten Schritt der Vorverarbeitung wird die grammatikalische Struktur des Satzes in Form eines Baumes abgebildet. Sobald alle Schritte abgeschlossen sind werden die einzelnen sogenannten Entities ausgelesen. Man unterscheidet hierbei zwischen *Named Entities*, es handelt sich dabei um Objekte wie z.B. Personen, oder um *Numeric Entities*, welche z.B. Jahreszahlen repräsentieren. Mit diesen Informationen kann schließlich eine weitere Interpretation auf den semantischen Inhalt dieses Textes erfolgen.

2.2.3 Interpretation

Der letzte Schritt bei der Verarbeitung von natürlicher Sprache ist die Interpretation des Inhaltes. Dabei werden die extrahierten Informationen mittels Algorithmus so vernetzt, dass Aussagen über den Inhalt getroffen werden können. Dieser Teilspekt des *NLP* ist sehr vielseitig und abhängig von der Komplexität der Anforderung. Es können hier von einfachen logischen Überprüfungen bis zu der Anwendung von künstlicher Intelligenz ein breites Spektrum an Algorithmen verwendet werden.

2.2.4 Ausgabe und Ergebnis

Am Ende sollte je nach Anforderung bzw. Anwendung ein Text ausgegeben, ein Ereignis getriggert, oder weitere Schritte eingeleitet werden. In den nächsten Abschnitten werden einige Möglichkeiten der Anwendung von *NLP* erläutert, die bereits in IT-Prozessen Anwendung finden oder eventuell verwendet werden könnten.

2.3 Anwendungen von Natural Language Processing

Nachdem im letzten Abschnitt grob skizziert wurde, wie die Verarbeitung und Interpretation von natürlicher Sprache stattfindet, wird im Folgenden auf verschiedene Beispiele für Anwendungen von *NLP* im IT-Bereich eingegangen.

2.3.1 Chatbots

Ein Bereich in dem die automatisierte Texterkennung verstärkt Anwendung findet, sind Bots. Die Bots könnten dabei z.B. den First-Level Support ersetzen und grundlegende Fragen klären, wie die Art des Problems, etwaige Kontaktdaten oder auch die Dringlichkeit des Problems. Es findet dabei eine Bidirektionale Kommunikation zwischen dem Kunden und einem Bot statt. Der Kunde schreibt eine Nachricht, der Bot versucht diese zu interpretieren und weitere Fragen zu erschließen. Moderne Bots bieten hier zusätzlich die Möglichkeit, dass der aktuelle Kontext persistent ist, was zur Folge hat, dass die Konversation durch die bereits gestellten Fragen und erhaltenen Antworten angepasst wird.

In 2.2 ist beispielhaft dargestellt, wie mit Hilfe der Texterkennung analysiert wird, welche Beiträge der Anwender gerne sehen möchte. Es werden danach die Artikel übermittelt, welche am besten mit den gesuchten Begriffen übereinstimmen.

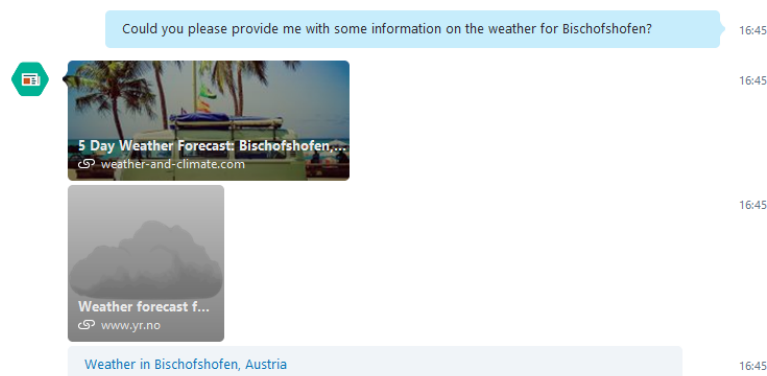


Abbildung 2.2: Bingnews Skype bot

Dabei kann man gut erkennen, dass die Frage in normaler menschlicher Sprache formuliert und nicht ausschließlich in Schlagwörtern gesucht wurde.

2.3.2 Spam-Filter

Eine äußerst bekannte und sehr weit verbreitete Anwendung von *NLP* sind Spam-Filter. In der Vergangenheit wurde bei Spam-Filtern lediglich ein Whitelisting bzw. Blacklisting für verschiedene Absender oder Wörter betrieben. Sobald eine Email von einem Absender kommt, der sich auf einer Blacklist befindet, wird diese als Spam markiert. Selbiges gilt für Emails mit gewissem Inhalt, der darauf schließen lässt, dass es sich bei der Email höchstwahrscheinlich um Spam handelt. In den letzten Jahren konnte bei der Verarbeitung von Emails durch die Anwendung von *NLP* eine Verbesserung bei der Erkennung und Verwaltung von Spam erreicht werden wie z.B. in [Roh14] erläutert.

2.3.3 Sprachassistenten

Bereits seit einigen Jahren erfreuen sich die zahlreichen Sprachassistenten immer größter Beliebtheit. Siri (Apple), Alexa (Amazon) und Cortana (Microsoft) sind dabei nur einige wenige Produkte, welche durch einen immer weiterwachsenden Funktionsumfang und eine immer tiefere Integration in vorhandene Systeme verstärkt auch im Unternehmensumfeld Anwendung finden. Dabei wird vor allem Cortana im Windows Umfeld sehr stark von Microsoft beworben und zahlreiche neue Möglichkeiten geschaffen, wie der Arbeitsalltag erleichtert werden sollte. Das Spektrum reicht dabei von der Suche nach Dokumenten, bis zum Planen von Terminen inkl. einer Rückmeldung, falls der gewünschte Zeitpunkt nicht möglich ist. Dabei ist eine sehr starke Verknüpfung mit den unterschiedlichen Systemen wie z.B. Exchange oder Office 365 erforderlich.

Im Folgenden wird ein Beispiel skizziert, welches mit Hilfe des von Microsoft entwickelten Sprachassistenten Cortana getestet wurde. Prinzipiell ist dieses Szenario auch mit anderen Sprachassistenten möglich.

Der Anwender möchte beispielsweise einen Termin für den 30. Oktober 2017

um 10:00 planen. Cortana reagiert dabei auf die Spracheingabe und nach einer erfolgreichen Verarbeitung der Eingabe, wird in Exchange überprüft, ob der Termin belegt ist. Für diesen Tag ist bereits ein Termin eingeplant und Cortana startet eine bi-direktionale Kommunikation, indem dem Anwender mitgeteilt wird, dass der Termin bereits belegt ist und einige verschiedene Optionen aufgezeigt werden. Der Anwender entscheidet sich dafür, dass der Termin der bereits vorhanden ist verschoben wird. In einer Rückmeldung informiert Cortana den Anwender, dass der Termin verschoben und der neue Termin eingetragen wurde. Cortana erkundigt sich, ob alle Teilnehmer des verschobenen Termins informiert werden sollten. Der Anwender antwortet mit Ja und Cortana leitet im Hintergrund die nötigen Schritte ein.

Dieser Fall zeigt nur eines der Beispiele, in denen ein Sprachassistent auch im Unternehmensumfeld Anwendung finden kann.

Kapitel 3

Anwendung von NLP in IT-Prozessen

3.1 Allgemeines

In diesem Kapitel wird erläutert, wie Natural Language Processing in ausgewählten IT-Prozessen Anwendung finden könnte, sowie die Verbesserung die dadurch erwartet wird. Es wird außerdem darauf eingegangen, welche technischen Limitierungen bzw. welche Probleme bei der Anwendung von Natural Language Processing noch entstehen, sowie ein Ausblick über die Machbarkeit gegeben.

Für die Analyse wird beschrieben, wie sich der Ablauf des Prozesses im Moment darstellt. Dazu wird der Ablauf mit Hilfe der *BPMN*-Notation skizziert. Es wird erläutert wie die einzelnen Schritte im Detail abgearbeitet. Danach wird vermittelt, wie durch die Anwendung von *NLP* eine Verbesserung erzielt werden kann. Dabei werden am Anfang eher einfachere Maßnahmen beschrieben, die leicht umzusetzen sind und keinen enormen Aufwand bedeuten, sowie einige eher komplexere Maßnahmen die teilweise weitere Voraussetzung, wie zum Beispiel, das Vorhandensein einer Wissensbasis, erfordern. Es wird ein Überblick gegeben, welche Schwierigkeiten sich ergeben und welche Vorteile durch die Anwendung von *NLP* für die Prozessteilnehmer und den Prozess erzielt werden können.

3.2 Das Beispielunternehmen

Das Unternehmen *XYZ Softwarehaus*, welches in den folgenden Abschnitten skizziert wird, ist ein klassisches Softwareentwicklungsunternehmen. Das *XYZ Softwarehaus* entwickelt und vertreibt dabei seine Software für die Planung von Meetings. Dabei wurden einige zentrale Geschäftsprozesse festgelegt, wovon die meisten klassische IT-Prozesse sind, die zwar mit Softwareunterstützung abgewickelt werden, jedoch keine Anwendung von *NLP* stattfindet. Es wird im Folgenden beispielhaft auf folgende Geschäftsprozesse eingegangen, in denen *NLP* möglicherweise eine Verbesserung bringen könnte:

1. Supportprozess
2. Neukundenprozess

3.3 Supportprozess

Für die Abwicklung von Probleme oder allgemeinen Abfragen wurde im Unternehmen ein Supportprozess definiert. Dieser gibt vor, welche Schritte zu unternehmen sind und wie der Kunde schlussendlich zu einer Lösung kommt. In Abbildung 3.1 ist der Prozess in *BPMN-Notation* dargestellt.

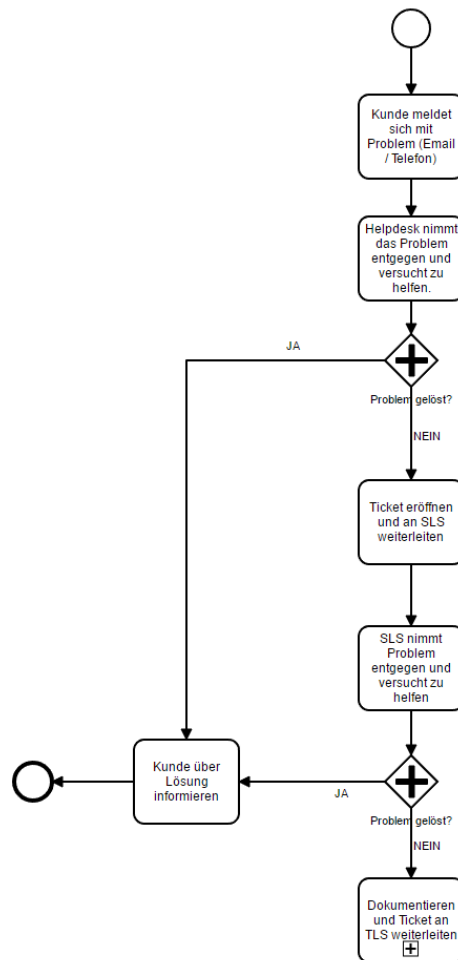


Abbildung 3.1: Supportprozess des XYZ Softwarehauses

In dieser Abbildung sind die einzelnen Schritte des Prozesses dargestellt und es wird beschrieben wie diese abgearbeitet werden. Im Folgenden wird der Ablauf skizziert und geschildert, wie die einzelnen Schritte abgearbeitet werden und welche Ergebnisse geliefert werden.

3.3.1 Ablauf

Der erste Schritt des Prozesses ist die Anfrage eines Kunden über Email oder per Telefon. Dabei entstehen abhängig vom gewählten Medium unterschiedliche

Daten. Diese stellen sich entweder in Form von Lautsprache, oder in Form von geschriebenem Text dar. Im Folgenden wird ausschließlich auf die Anfrage per Email eingegangen, da für die Anfrage per Telefon kein direkter Vorteil durch den Einsatz von *NLP* erzielt werden kann, da hier vor allem auch der menschliche Kontakt eine wichtige Rolle spielt.

Der Helpdesk nimmt die Anfragen des Kunden entgegen und versucht diese zu lösen. Dazu stehen dem Helpdesk mehrere Möglichkeiten zur Verfügung. In den meisten Fällen wird der Mitarbeiter am Helpdesk versuchen, Informationen zu dem gemeldeten Problem in der Wissensbasis zu finden. Dazu wird die Wissensbasis mit Schlagwörtern durchsucht, die das Anliegen des Kunden möglichst gut wiedergeben sollten. Kann keine Information aus der Wissensbasis gefunden werden kann versucht werden, das Problem ohne das vorhandene Wissen zu lösen. Falls eine Lösung auf einem anderen Weg gefunden werden kann, wird der Mitarbeiter die Informationen zur Lösung in der Wissensbasis hinzufügen und den Kunden über die Lösung des Problems informieren. Für jede Lösung wird dem Kunden zusätzlich ein Link zu dem Artikel mit der Lösung in der Wissensbasis übermittelt. Dies führt im Idealfall dazu, dass der Kunde in Zukunft ohne Hilfe des Helpdesks zu einer Lösung kommt.

Kann auf Grund der Komplexität des Problems im ersten Schritt vom Helpdesk keine Lösung gefunden werden, wird überprüft, ob bereits ein Ticket für dieses Problem geöffnet wurde. Möglicherweise ist das gemeldete Problem bereits bei anderen Kunden aufgetreten und daher auch schon im Ticketsystem vermerkt. Dieses manuelle auffinden der Tickets ist sehr häufig mit einigem an Zeitaufwand verbunden und häufig werden doppelt erstellte Tickets erst später entdeckt. Falls das Problem noch nicht in Form eines Tickets beschrieben wurde, wird ein Ticket erstellt, welches alle bereits im Helpdesk erhaltenen Informationen enthält und an den sogenannten *Second Level Support (SLS)* weitergeleitet. Wenn im SLS eine Lösung gefunden werden kann, folgen die gleichen Schritte wie bei erfolgreicher Abarbeitung im Helpdesk: Der Kunde wird informiert und die Wissensbasis ergänzt. Zusätzlich wird dem Helpdesk eine Information gegeben, dass die Wissensbasis um die nötigen Informationen erweitert wurde, sodass bei zukünftigen Anfragen dieser Art eine Lösung möglicherweise schon beim Helpdesk gelöst werden kann.

Wenn das Problem auch im SLS nicht gelöst werden kann, geht es schließlich zum sogenannten *Third Level Support (TLS)*. Dieser Teil des Prozesses ist in Abbildung 3.1 nur als Subprozess modelliert, da er sich sehr ähnlich gestaltet wie der Prozessschritt im SLS. Das Ergebnis dieses Prozessschrittes ist erneut, die Rückmeldung für den Kunden und die Erweiterung der Wissensbasis, sowie die Information des Helpdesks und des SLS über die Erweiterung.

Im Endeffekt sollte durch die kontinuierliche Erweiterung und Verbesserung der Wissensbasis ein System geschaffen werden, dass immer mehr Probleme direkt vom Kunden selbst gelöst werden können, oder bereits vom Helpdesk. Vor allem technische Probleme werden sehr häufig nur im SLS oder im TLS lösbar sein, da im Helpdesk schlichtweg keine Zeit für die Lösung dieser bleibt.

3.3.2 Verbesserung 1 - Kategorisierung der Anfrage

Die erste Verbesserung die im Zuge des Supportprozesses vorgenommen werden kann ist eine Kategorisierung der Anfrage. Die meisten Email Programme und Mail-server bieten bereits eine grundlegende Kategorisierung über sogenannte Regeln an. Es kann zum Beispiel festgelegt werden, dass alle Emails die von einem bestimmten Absender kommen, in ein bestimmtes Unterverzeichnis verschoben werden können. In Abbildung 3.2 ist ein Dialog für das Festlegen von Regeln in *Microsoft Office 2016* dargestellt. Diese regelbasierte bietet bereits eine sehr einfache Möglichkeit der Vorsortierung der Emails, leider ist der Konfigurationsaufwand für das Erstellen dieser Regeln oft sehr groß. Die Regeln sind sehr strikt und können nur sehr einfache Bedingungen enthalten. Es ist z.B. nicht möglich, alle Emails von einer gemeinsamen Domain (microsoft.com) in ein Unterverzeichnis weiterzuleiten. Weiters gibt es keine Möglichkeiten die Mails auf den Inhalt zu prüfen und Regeln anzugeben, die abhängig vom Inhalt Mails Kategorisieren.

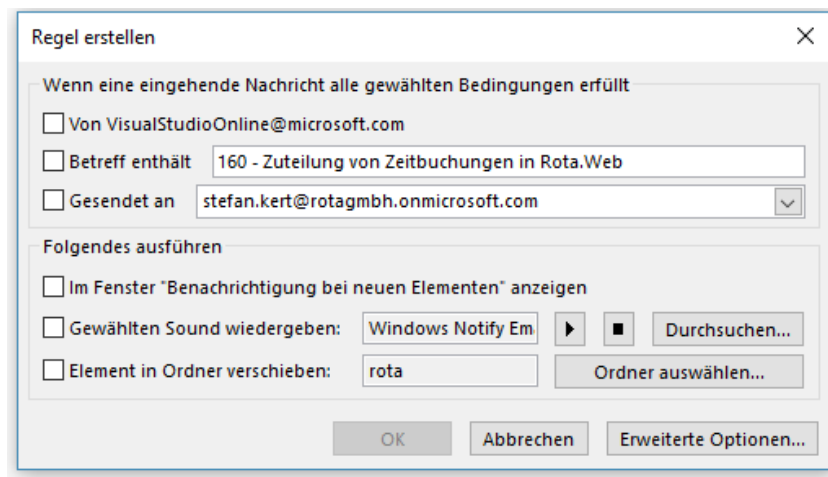


Abbildung 3.2: Dialog zum Festlegen von Regeln in MS Outlook

Hier kann eine Anwendung von *NLP* zum Beispiel dazu dienen, dass abhängig vom Inhalt der Mail Kategorien zugeordnet werden, oder komplexere Regeln für die Vorsortierung gewählt werden können. Mit Hilfe der *Named Entity Recognition* können bereits Wörter ausgewählt werden die häufig vorkommen und abhängig von diesen eine Weiterleitung an spezielle Stellen eingerichtet werden. Ein Beispiel hierfür wäre folgendes Szenario:

Kunde A stellt eine Anfrage über die Möglichkeiten der Lizenzerweiterung. Die Email geht direkt in den sogenannten *Office-Posteingang* der als gemeinsames Postfach für das Unternehmen verwendet wird. Dieser Posteingang wird vom Helpdesk verwaltet. Im Moment leitet der Helpdesk die Anfragen direkt an den Vertrieb weiter, da es sich um keine Supportanfragen handelt.

Da Entscheidung über eine Weiterleitung an den Vertrieb auf dem Inhalt der Email basiert kann hier eine Anwendung von *NLP* erfolgen. Der Text kann direkt beim Eingehen der Email vorab geprüft werden und sobald bestimmte Kriteri-

en erfüllt werden, erfolgt eine automatische Weiterleitung an den *Vertrieb*. Diese Kriterien reichen von einfachen Regeln, wie dem Inhalt von Schlagwörtern, bis zu komplexeren Kriterien wie der beschriebenen *Named Entity Recognition* die über die Anwendung einer *Taxonomie* auch eine Klassifizierung von Wörter ermöglicht.

Vorteile

Für diese Verbesserung ergeben sich folgende Vorteile:

- Bessere Vorverarbeitung von Emails
- Automatisches Zuordnen zu den richtigen Stellen für Emails
- Komplexere Regeln möglich

Machbarkeit

Wie bereits eingangs erwähnt sind diese Maßnahmen relativ leicht umzusetzen. Das Zuordnen von Stichwörtern entspricht nicht direkt dem *NLP* sondern sind eher fix definierte Regeln und es gibt bereits Tools die dies ermöglichen. Das Anwenden einer *Taxonomie* für die Klassifizierung und Zuordnung erfordert eine Integration in das System. Dies könnte beispielsweise über Plugins realisiert werden. Vor allem im englischsprachigen Bereich gibt es sehr umfangreiche *Taxonomien* die auch eine sehr komplexe Zuordnung erlauben. Für die Anwendung von *NER* gilt sehr ähnliches wie für die *Taxonomie*. Auch ihre Anwendung erfordert eine tiefere Integration ins System. Auch hier hat man mit englischsprachigem Inhalt mehr Möglichkeiten als für deutsche Texte. Grundsätzlich lässt sich für diese Verbesserung zusammenfassend sagen, dass ein Einstieg über die Zuordnung über Schlagwörter bereits einen Vorteil und eine Prozessoptimierung bringen würde, da diese Aufgabe, die vom Helpdesk nach vorgegeben Regeln abgearbeitet wird, über diese Zuordnung nach Schlagwörtern auch automatisiert passieren kann. Im ersten Schritt wäre es also möglich, diese Weiterleitung auf Grund von Schlagwörtern mit einer einfachen Liste zu verwalten, welche im Prinzip *m:1* Beziehungen darstellt, die ein Mapping von Wörtern zu Emails repräsentieren. Die Email wird schließlich der Email zugeordnet, bei der die meisten Regeln erfüllt sind.

3.3.3 Verbesserung 2 - Hinzufügen von Informationen aus der Wissensbasis

Ein sich im Prozess immer wiederholender Schritt ist die kontinuierliche Erweiterung der Wissensbasis, sowie die Verwendung der darin enthaltenen Informationen. Hier wird im Moment durch manuelles Durchsuchen dieser Wissensbasis versucht Informationen zu erhalten. Dies ist mit sehr viel Aufwand verbunden, da der Helpdeskmitarbeiter die Schlagwörter oder Inhalte die in der Email enthalten sind einzeln in der Wissensbasis eingeben muss und schließlich aus den Ergebnissen eine Antwort zusammenfassen muss. Dies ist häufig mit sehr viel Zeit und Aufwand verbunden und liefert aber häufig sehr gute Ergebnisse die genau auf das Problem des Kunden zugeschnitten sind.

Um die Effizienz dieser Suche zu steigern kann mit Hilfe von *NLP* nach einer Vorverarbeitung und Kategorisierung der Anfrage, ein automatisches Durchsuchen der Wissensbasis erfolgen. Dabei wird die Wissensbasis gleich wie beim manuel-

len Durchsuchen nach bestimmten Schlagwörtern oder Szenarien durchsucht, und die Ergebnisse ausgewertet. Wie bei vielen textbasierenden Suchmaschinen oder Algorithmen ist es auch hier möglich ein Ranking nach der Relevanz der einzelnen gefunden Informationen im Bezug auf die Suche anzugeben. Das Ergebnis ist eine Liste an Artikeln die nach der Relevanz sortiert sind.

Der Mitarbeiter des Helpdesks hat nun die Möglichkeit, dem Kunden einen Artikel aus der erhaltenen Liste zurückzusenden, wodurch noch eine weitere manuelle Überprüfung der erhaltenen Informationen auf Korrektheit möglich ist.

Hier wäre es möglich, durch die Rückgabe der Auswahl des Artikels durch den Kundenbetreuer, das Model für die Auswahl der Artikel weiter zu trainieren. Wenn beispielsweise keiner der Artikel zu der gestellten Anfrage passt, kann eine manuelle Suche durch den Helpdeskmitarbeiter erfolgen. Wenn bei dieser manuellen Suche ein Artikel gefunden werden konnte, wird mit dieser Information das Model trainiert, sodass in Zukunft ein besseres Ergebnis erzielt werden kann.

Dieses Trainieren des Models kann schließlich auch für die manuelle Suche in der Wissensbasis helfen, von dem schlussendlich auch Kunden und Helpdeskmitarbeiter profitieren, da die Ergebnisse, die zu Anfragen gefunden werden, durch das Trainieren des Models immer genauer werden.

Am Ende sollte der Helpdeskmitarbeiter die Möglichkeit haben, nach der manuellen Auswahl des Artikels, eine Email generieren zu lassen, die alle nötigen Informationen wie Links zum Artikel in der Wissensbasis, oder auch den Artikel im PDF Format als Anhang mitzusenden.

Ein weiterer möglicher Schritt wäre, die Rückmeldung des Kunden, ob der Artikel für die aufgetretene Problematik relevant ist, auch wieder ins System zurückfließen zu lassen und so das Model weiter trainiert wird. Hier, wie auch bei der Auswahl durch den Helpdeskmitarbeiter, besteht natürlich die Gefahr, dass eine falsche Auswahl getroffen wird und auf Grund dessen eine falsche Information ins Model gelangt. Um diese Problematik so gering wie möglich zu halten, erfordert es vor allem am Anfang, wenn das Model noch nicht sehr ausgeprägt trainiert ist, eine hohe Genauigkeit am Helpdesk.

Vorteile

Für diese Verbesserung ergeben sich folgende Vorteile:

- Schnelles und automatisches Erhalten von Informationen aus der Wissensbasis
- Geringer manueller Aufwand für Helpdesk
- Manuelle Suche wird durch Trainieren des Models verbessert

Machbarkeit

Wie auch schon bei Verbesserung 1 erfordert Verbesserung 2 zahlreiche Maßnahmen die im Vorhinein getroffen werden müssen. Wenn für das Unternehmen keine Wissensbasis vorhanden ist, oder diese nur unzureichend befüllt ist, wird diese Verbesserung keinen Vorteil für das Unternehmen bringen, da entweder gar keine oder nur schlechte Ergebnisse geliefert werden. Wichtig ist vor allem der manuelle Input. Das Model für die Suche der Daten sollte im Idealfall laufend trainiert werden,

sodass die Ergebnisse im Laufe der Zeit immer besser werden. Dies bedeutet vor allem am Anfang einen erhöhten manuellen Aufwand. Häufig gibt es bereits eine Auswertung wie hilfreich gewisse Informationen aus der Wissensbasis sind, dies kann natürlich für das Reihung der Informationen nach Relevanz verwendet werden. Auch hier gilt wieder, dass eine einfache Implementierung bereits einige Vorteile bringt, auch wenn die Zeitersparnis nur daraus besteht, dass der Mitarbeiter oder die Mitarbeiterin nicht die Wissensbasis durchsuchen muss, oder dies erst tun muss, wenn die gelieferten Ergebnisse unzufriedenstellend sind.

3.3.4 Verbesserung 3 - Erkennen von duplizierten Tickets

Wenn das Problem von einem Mitarbeiter oder Mitarbeiterin des Helpdesk nicht gelöst werden kann, wird ein sogenanntes Ticket erstellt. Dieses Ticket enthält Informationen über das Problem, den Kunden oder die Kundin, wo das Problem aufgetreten ist und weitere Metainformationen wie, Betriebssystem, Version der Applikation oder auch ob das Problem nur zu einer bestimmten Uhrzeit auftritt. Vor der Erstellung dieses Tickets, wird das Ticketsystem noch auf ähnliche Probleme durchsucht. Häufig ergibt diese Suche, dass einige Probleme bereits gemeldet wurden. Ist dies der Fall wird das neue Ticket mit dem alten Ticket verküpft, dass beide gemeinsam gelöst werden können. Dieser Schritt erfordert erneut eine manuelle Suche die durch einen Mitarbeiter oder eine Mitarbeiterin des Helpdesks erfolgen muss. Hier könnte eine Verbesserung durch die Anwendung von *NLP* dadurch erreicht werden, dass sobald ein Ticket erstellt wird, die Wissensbasis automatisch nach ähnlichen oder gleichen Tickets durchsucht wird. Es sollten also Duplikate von Tickets erkannt werden und dem Mitarbeiter oder der Mitarbeiterin eine Rückmeldung gegeben werden, dass dieses oder ein ähnliches Ticket bereits erstellt wurde. Auch hier erfordert es wieder eine manuelle Eingabe des Mitarbeiters oder der Mitarbeiterin. Es wird bei dieser Zuordnung überprüft, ob die ermittelten Ergebnisse wirklich übereinstimmen. Ist dies der Fall, wird das Ticket zugeordnet.

Vorteile

Für diese Verbesserung ergeben sich folgende Vorteile:

- Schnelles und automatisches Auffinden von duplizierten Tickets
- Weniger Tickets kommen zum SLS oder zum TLS, da duplizierte Tickets bereits zuvor zusammengefasst werden

Machbarkeit

Die Funktionalität, welche für diese Verbesserung benötigt wird, ist wie auch bei der Wissensbasis eine einfache textbasierte Suche. Die vorhandenen Tickets werden mit dem neuen Ticket verglichen und die Ähnlichkeit der beiden Tickets ist ausschlaggebend für die Reihung der Ergebnisse. Dies könnte beispielsweise über ein Plugin für das Ticketsystem gelöst werden.

3.4 Neukundenprozess

Die Anfrage und die Abarbeitung der Anfragen von Neukunden, stellt einen weiteren wichtigen Geschäftsprozess im Unternehmen dar. In Abbildung 3.3 ist der Prozess in *BPMN-Notation* dargestellt.

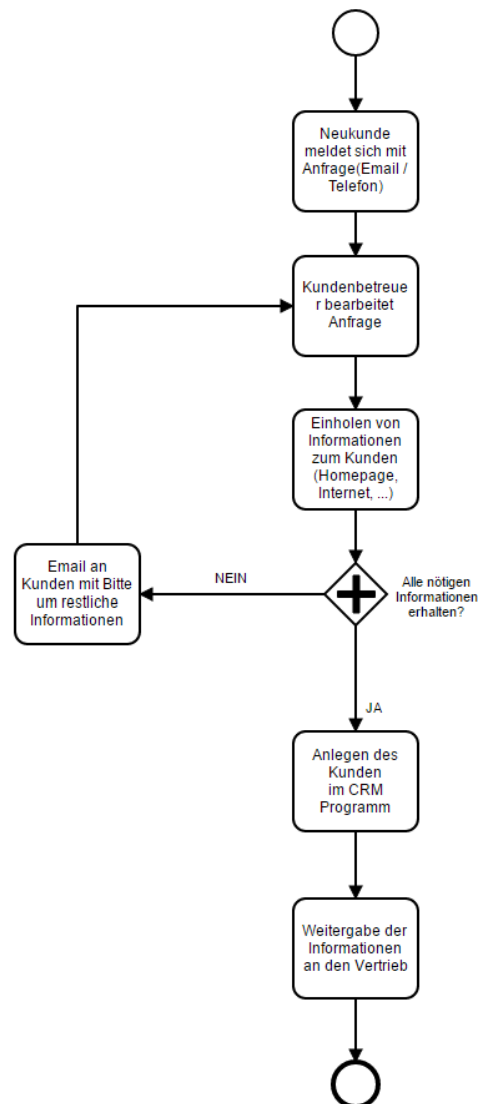


Abbildung 3.3: Neukundenprozess des XYZ Softwarehauses

3.4.1 Ablauf

Gleich wie beim Supportprozess startet der Prozess mit der Anfrage eines Kunden. Dabei ist der Unterschied zum Supportprozess, dass diese Anfrage von jemanden

kommt, der im CRM-System noch nicht vorhanden ist und für den noch keine Daten hinterlegt sind. Nach Erhalt der Anfrage, per Telefon oder per Email, wird diese von einem Kundenbetreuer bearbeitet. Dabei wird mit den in der Anfrage enthaltenen Informationen nach weiteren Informationen gesucht. Wenn die Anfrage beispielsweise einen Link zur Homepage enthält, wird diese Homepage konsultiert und die darin enthaltenen Informationen wie Standort, Branche oder auch die Anzahl der Mitarbeiter dokumentiert. Weiters wird mit Hilfe einer Onlinesuche noch nach weiteren Informationen gesucht, die für den Verkaufsprozess von Bedeutung sind. Falls noch Fragen offen sind, oder Informationen nicht oder nur teilweise erhalten werden konnten, wird dem Kunden eine Anfrage mit der Bitte um Zusendung der restlichen Informationen zugesendet. Die Rückmeldung auf diese Anfrage wird erneut vom Kundenbetreuer bearbeitet und falls alle wichtigen Informationen vorhanden sind, wird er Kunde im CRM angelegt und die Information an den Vertrieb weitergegeben. Dies ist zugleich der Auslöser für den Verkaufsprozess.

3.4.2 Verbesserung 1 - Automatisches Auslesen der Informationen aus der Email

Eine erste Möglichkeit für eine Verbesserung ist das automatische Auslesen der bereits in der Email enthaltenen Informationen. Häufig werden für Neukundenanfragen Formulare verwendet. Diese ermöglicht das automatische Auslesen ungeeignet, da es immer nach einem fixen Muster vorgenommen werden kann. Durch die Anwendung von *NLP* wird es auch möglich, Informationen, welche im Fließtext vorhanden sind auszulesen. Durch die Anwendung der *Named Entity Recognition* können bereits einige Informationen wie, Personen, Firmenname, Adressen etc. ausgelesen werden. Das Ergebnis ist eine Liste, die der Kundenbetreuer oder die Kundenbetreuerin überprüft, verbessert und schließlich im System hinterlegt.

Vorteile

Für diese Verbesserung ergeben sich folgende Vorteile:

- Automatisches Auslesen der Informationen aus der Email

Machbarkeit

Diese Verbesserung stellt sich als eher schwierig dar, da vor allem durch Rechtschreibfehler oder syntaktische Fehler im Text eine *Named Entity Recognition* nicht die gewünschten Ergebnisse liefern wird. Meist ist es schneller, die erhaltenen Informationen durch einen Mitarbeiter oder eine Mitarbeiterin der Kundenbetreuung manuell zu verarbeiten. Vorteile bringt diese Verbesserung nur bei sehr lange Emails, die sehr viele Informationen enthalten, oder auch bei einer sehr hohen Anzahl an Anfragen, dass bereits eine Vorverarbeitung dieser erfolgen kann. Bei kurzen Emails, oder wenn nur gelegentlich Anfragen kommen, ist der Aufwand für die Implementierung und der Aufwand für die manuelle Fehlerkorrektur nicht gerechtfertigt.

3.4.3 Verbesserung 2 - Sammeln von Informationen

In einem Schritt des Neukundenprozesses wird vom Kundenbetreuer oder der Kundenbetreuerin versucht, weitere Informationen über den Kunden im Internet zu finden. Meist wird dabei nach einem fixen Muster vorgegangen:

1. Durchführen einer Suche im Internet nach dem Neukunden (evtl. ist die Webseite auch in der Anfrage enthalten)
2. Durchsuchen der Webseite des Kunden nach relevanten Informationen
3. Aufrufen des Impressums und notieren der darin enthaltenen Information
4. Informationen notieren und später im CRM hinterlegen.

Diese Schritte wiederholen sich prinzipiell bei jedem Neukunden. Die Anwendung von *NLP* könnte hier auf zwei verschiedene Arten angestoßen werden:

Automatisch

Bei Erhalt einer Anfrage eines Neukunden wird die Anfrage automatisch nach Informationen zum Kunden durchsucht. Links die in der Email enthalten sind werden aufgerufen und überprüft, ob die Webseite weitere Informationen für den Kunden enthält. Dafür ist auch eine Überprüfung notwendig, ob die Webseite überhaupt dem Kunden zuzuordnen ist oder ob sich der Link aus einem anderen Grund in der Anfrage befunden hat. Ist keine Webseite und kein Link in der Anfrage enthalten wird mit Hilfe einer Suchmaschine nach dem Kunden gesucht. Hierfür werden die in der Anfrage enthaltenen Informationen verwendet.

Manuell

Wenn die Anfrage keine Webseite für den Kunden enthält und auch keine Informationen zu einer erfolgreichen Suche führen, muss manuell eingegriffen werden. Der Kundenbetreuer oder die Kundenbetreuerin hat dann die Möglichkeit, die Verarbeitung der Webseitendaten manuell anzustoßen. Eine Möglichkeit zum Erhalt der nötigen Informationen wäre zum Beispiel eine Rückfrage beim Kunden.

Nachdem die Verarbeitung angestoßen wurde und die Webseite vom Algorithmus durchsucht wurde wird im Idealfall eine Liste mit zusätzlichen Informationen zum Kunden ausgegeben. Diese Liste wird wiederum von einem Kundenbetreuer oder Kundenbetreuerin überprüft und eventuell erweitert oder ausgebessert. Die erhaltenen Informationen werden schließlich im CRM hinterlegt.

Vorteile

Für diese Verbesserung ergeben sich folgende Vorteile:

- Automatisches Sammeln weiterer Informationen aus dem Internet
- Umfangreichere Informationen zum Kunden

Machbarkeit

Wie auch schon beim automatischen Verarbeiten der Informationen, die in der Anfrage enthalten sind, erfordert die Verarbeitung der Daten auf der Webseite ma-

nuelle Arbeit. Die Überprüfung der Ermittelten Daten auf Korrektheit ist zwangsläufig notwendig, da viele Faktoren das automatische Auslesen der richtigen Daten schwierig gestalten. Ein Problem stellen falsch gepflegte Daten auf der Webseite dar, die durch einen Kundenbetreuer oder eine Kundenbetreuerin erkannt werden können, durch eine Verarbeitung mit Hilfe von *NLP* aber nur sehr selten. Grundsätzlich wäre eine Anwendung der automatischen Informationsbeschaffung im Internet aber hilfreich, da dem Kundenbetreuer oder der Kundenbetreuerin sehr schnell viele Informationen zum Unternehmen zur Verfügung stehen die ausgewertet und verarbeitet werden können.

Kapitel 4

Fazit

Grundsätzlich lässt sich sagen, dass die Anwendung von *Natural Language Processing* vor allem in Prozessen in denen viele Daten in Form von Text verarbeitet werden eine Verbesserung darstellen. Dies ist vor allem dort der Fall, wo eine Kommunikation mittels Chat, Email oder Onlineformular stattfindet. Für die meisten Anwendungen von *NLP* ist eine manuelle Überprüfung durch einen Prozessteilnehmer oder eine Prozessteilnehmerin notwendig, da die erhaltenen Ergebnisse noch überprüft werden müssen. Durch diese manuelle Überprüfung kann das System außerdem trainiert werden, sodass die gelieferten Ergebnisse in Zukunft möglicherweise besser dem erwarteten entsprechen. Es wäre für Unternehmen von großem Vorteil, sich mit der Thematik des *NLP* zu beschäftigen. Bereits rudimentäre Anwendungen, bringen Vorteile für die Prozesse und die Teilnehmer des Prozesses mit sich. Der Supportprozess ist hierfür ein gutes Beispiel. Bereits durch eine einfache Texterkennung kann eine Weiterleitung der Anfragen an die richtigen Stellen passieren. Der Aufwand für die Implementierung hält sich in Grenzen und erspart den Mitarbeitern die sich um die Anfragen kümmern häufig einiges an Arbeit. Im Moment ist *NLP* vor allem im englischsprachigen Raum bereits sehr stark fortgeschritten. Im deutschen ist die Unterstützung teilweise nur rudimentär und nur in sehr speziellen Szenarien verwendbar. Dies liegt unter anderem daran, dass die englische Sprache auf Grund ihrer geringeren Komplexität leichter zu verarbeiten ist. Ein weiterer Grund ist das wachsende Interesse von großen Unternehmen wie Microsoft, Apple oder Amazon an *NLP* und den Investitionen die vor allem für die Unterstützung von englischer Sprache getätigt werden.

Quellenverzeichnis

Literatur

- [Din14] Potolea Dinsoreanu Rădulescu. „Identification of spam comments using natural language processing techniques“. *Intelligent Computer Communication and Processing* (2014) (siehe S. 2).
- [NTC16] Tetreault Nobata, Mehdad Thomas und Chang. „Abusive Language Detection in Online User Content“ (2016) (siehe S. 1).
- [Roh14] Mukti Desai Rohit Giyanani. „Spam Detection using Natural Language Processing“ (2014) (siehe S. 7).
- [W X+16] J. Droppo W. Xiong u. a. „Achieving Human Parity in Conversational Speech Recognition“ (2016) (siehe S. 5).