

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

Expose

Stefan Kert

18. März 2017

## 1 Inhalt

Die Anwendungen rund um *Natural Language Processing* haben in den letzten Jahren sehr stark an Bedeutung gewonnen. Sei es durch die klassischen Sprachassistenten wie Cortana (Microsoft), Siri (Apple) und Amazon Echo (Amazon), oder durch die vielseitigen Möglichkeiten zur Analyse von Texten. Über Natural Language Processing ist es z.B. möglich einen Text zu klassifizieren. 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. Ein immer wichtiger werdendes Thema ist vor allem die Erkennung von sogenannten *Hasspostings* [5], aber auch das Analysieren von Postings zu Werbe- oder Marketingzwecken (User A postet, dass er gerne wieder auf Urlaub fahren würde, bekommt gezielt Werbung zu Urlaubsangeboten) ist ein Thema von höchster Wichtigkeit.

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 kommunikationslastigen 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-Prozessen 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 [1]. 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, oder komplexere Verfahren wie zum Beispiel die Analyse

des Textes auf speziell 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.

## 2 Motivation

Wie bereits erwähnt gibt es in den verschiedenen IT-Prozessen zahlreiche Anwendungsmöglichkeiten für Natural Language Processing. Einige IT-Prozessen 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, welches 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).

## 3 Forschungsfragen

Die folgenden Forschungsfragen sollten im Zuge dieser Arbeit beantwortet werden:

1. Natural Language Processing [3] [4]
  - a) Was ist Natural Language Processing?
  - b) Welche Anwendungen für Natural Language Processing gibt es, die für Prozesse in der IT von Relevanz sind?
  - c) Wann sollte Natural Language Processing eingesetzt werden?
2. Können IT-Prozesse durch die Unterstützung von Natural Language Processing verbessert werden? [2] [6]
  - a) In welchen IT-Prozessen kann Natural Language Processing Anwendung finden?
  - b) Wie können diese IT-Prozesse verbessert werden?

## 4 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.

## 5 Inhaltsverzeichnis

1. Kurzfassung
2. Einführung
  - a) Motivation
  - b) Ziele
3. Natural Language Processing
  - a) Übersicht
  - b) Anwendungen
4. Unterstützung ausgewählter IT-Prozessen durch die Anwendung von Natural Language Processing
  - a) Allgemeines
  - b) Analyse ausgewählter Prozesse
5. Fazit

## Literatur

- [1] Potolea Dinsoreanu, Rădulescu. Identification of spam comments using natural language processing techniques. *Intelligent Computer Communication and Processing*, 2014.
- [2] Bernhard M. Huber. *Managementsysteme für IT-Serviceorganisationen: Entwicklung und Umsetzung mit EFQM, COBIT, ISO 20000, ITIL*. dpunkt Verlag, 2009.
- [3] Dan Jurafsky and James H. Martin. *Speech and Language Processing*. Prentice Hall, 2008.
- [4] Christopher S Manning. *Foundations of Statistical Natural Language Processing*. MIT University Press Group Ltd, 1999.

- [5] Tetreault Nobata, Mehdad Thomas, and Chang. Abusive language detection in online user content. 2016.
- [6] Zhang Wang, Anvik Xie, and Sun. An approach to detecting duplicate bug reports using natural language and execution information. 2008.