# Exposé

**Titel der Arbeit**: Automatic Text Summarization using Deep Learning and Natural Language Processing

**Autor**: Daniel Vogel

**Forschungsfragen**: Wie lassen sich Texte automatisiert zusammenfassen? Wie qualitativ und skalierbar ist die Lösung?

**Inhaltsverzeichnis**

|  |  |
| --- | --- |
| Kapitel | Seiten |
| Einleitung | 5 |
| Deep Learning | 10 |
| Natural Language Processing | 10 |
| Datengrundlage | 3 |
| Extraktiver Ansatz | 10 |
| Abstraktiver Ansatz | 10 |
| Zusammenfassung | 3 |
| Diskussion & Ausblick | 3 |

**Abstract**

Zunächst ist eine kurze Einleitung erforderlich, welche den Kontext und die Notwendigkeit der Arbeit im Gesundheitswesen offenlegt. Hier wird beschrieben, dass diese Arbeit beispielsweise als Grundlage für eine nachgeschaltete Generierung und Verdichtung von Protokollen einer Videosprechstunde dienen könnte. Als Ziel wird die Zusammenfassung von Textdokumenten verstanden. Die Zusammenfassung von Dialogen ist somit nicht Teil dieser Arbeit ist, wohl aber eine potenzielle Fortsetzung. Auch der Forschungsstand und Referenzen gehören offengelegt.

Architekturen des Deep Learning, insbesondere neuronale Netze, werden für die Zusammenfassung der Texte genutzt, weshalb zunächst eine theoretisch fundierte Beschreibung aller relevanten Komponenten erfolgt. Die vielversprechendsten Architekturen und relevantesten Hyperparameter werden zudem beschrieben.

Auch das Natural Language Processing stellt eine unabdingbare Grundlage dieser Arbeit dar. Daher sind sowohl die dazugehörigen Vorverarbeitungsschritte allgemein zu beschreiben als auch die letztendliche Merkmalsextraktion zu konzipieren.

Bevor die beschriebenen Architekturen und Methoden tatsächlich Anwendung finden können, ist die Beschreibung der Datengrundlage erforderlich. Hierbei handelt es sich um unternehmensinterne fachspezifische und um frei verfügbare allgemeinsprachliche Textdaten.

Der extraktive Ansatz soll die Zusammenfassung der Texte mithilfe eines neuronalen Netzes angehen. Hierbei werden bestehende Sätze bewertet und entsprechend dieser Bewertung in die Zusammenfassung hineinkopiert.

Der abstraktive Ansatz hingegen kann auch neue Wörter für die zu erstellende Zusammenfassung generieren. Die Architektur und die Konfiguration einer ganzen Netzwerkarchitektur gehören hier entwickelt und beschrieben. Entsprechende Experimente, welche auch den Vergleich verschiedener Architekturen oder Konfigurationen impliziert, werden ebenfalls angestellt.

Die Zusammenfassung, die Diskussion und der Ausblick runden die Arbeit ab.

Literatur: Siehe Literaturverzeichnis (auf Anfrage per E-Mail zustellbar).

Stand: 14. Oktober 2020, 54 Seiten.