

Masterarbeit (Informatik, Medizintechnik, Data Science):

Semantische Annotation, Themen- und Ähnlichkeits-Analyse eines großen Corpus klinischer Dokumentationsformulare

Hintergrund:

Über die letzten 20 Jahre wurde am Universitätsklinikum Erlangen (UKER) ein großer Corpus von Dokumentationsformularen im klinischen Arbeitsplatzsystem (Siemens Soarian) aufgebaut. Die Implementierung der Formulare erfolgte nach den Vorgaben der verschiedenen klinischen Fachabteilungen, so dass sich vielfach Redundanzen ergeben haben (z.B. verschiedene Formulare zur Dokumentation des Raucherstatus mit teilweisen Überschneidungen, aber auch individuellen Unterschieden). Z.Zt. existiert noch keine semantische Annotation der Formulare mit standardisierten Terminologien

Problemstellung:

Zur Vorbereitung einer koordinierten Re-Implementierung der Formulare in einem neuen klinischen Arbeitsplatzsystem (KAS) sollen die bestehenden Formulare semantisch annotiert und auf dieser Basis systematisch in Bezug auf ihre Inhalte (Topics) und Ähnlichkeit miteinander analysiert werden. Zur semantischen Annotation sollen Natural Language Processing-Methoden (NLP) bzw. Large Language Models (LLMs) eingesetzt werden. Zur Analyse der Themen und Ähnlichkeiten sollen die Annotation in einer Graph-Datenbank abgelegt und entsprechende Distanzmaße bestimmt werden. Die Ergebnisse der Arbeit stellen einen wesentlichen Input für das Formularkonzept im zukünftigen KAS des UKER dar.

Fragestellungen:

- F1: Kann der KAS-Formularcorpus mit verfügbaren NLP- oder LLM-Methoden semantisch annotiert werden?
- F2: Welche Themenkomplexe und Ähnlichkeiten können aus dem annotierten Corpus abgeleitet werden?
- F3: können Bezüge zu öffentlich verfügbaren Formularsammlungen (z.B. [MDM-Portal](#)) hergestellt werden?

Aufgaben:

- A1: Einarbeitung in die verfügbaren Quelldaten & Literatur & Methoden
- A2: Semantische Annotation des Corpus & Ablage der Ergebnisse in einer Graph-Datenbank
- A3: Themen- & Ähnlichkeitsanalyse sowie Visualisierung der Ergebnisse
- A4: Analyse möglicher Bezüge zum MDM-Portal

Notwendige Vorkenntnisse:

- Notwendig: Programmierkenntnisse & -erfahrung z.B. in Python, gute deutsche Sprachkenntnisse, gute Selbstorganisation & Eigeninitiative
- Hilfreich: Erfahrung mit NLP- und/oder LLM-Methoden & Graph-Datenbanken

Literatur:

- z.B. <https://www.sciencedirect.com/science/article/pii/S1532046414000938>