

# Domänenspezifische Sprachmodelle in der semantischen Modellierung

## Wissenschaftliche Fragestellung und Zielsetzung der Arbeit

Durch die wachsende Menge an Daten und die Abhängigkeit von ihnen wird die effiziente Verwaltung dieser Daten immer wichtiger. Eine Möglichkeit, mit großen Mengen komplexer Daten umzugehen, ist das semantische Datenmanagement. Ein wesentlicher Bestandteil des semantischen Datenmanagements ist die Erstellung semantischer Modelle, d. h. von Wissensrepräsentationen, die zum Verständnis und zur Interpretation des gegebenen Datensatzes erforderlich sind. Methoden zur



Methoden zur automatischen Erstellung erster Versionen semantischer Modelle konzentrieren sich auf Beschriftungen oder Werte innerhalb des Datensatzes. Eine andere Möglichkeit, ein semantisches Modell automatisch zu erstellen, die in diesem Zusammenhang noch kaum erforscht ist, ist die Verwendung von beschreibenden Texten der Datenquelle. Zur Anwendung von NLP-Technologien ist die Identifikation eines geeigneten Sprachmodells von essentieller Bedeutung. Dabei kann es große Unterschiede machen, ob es sich um ein generalisiertes Sprachmodell handelt oder es an die zu Grunde liegende Ontologie angepasst ist.

## Vorgehensweise und Methodik

In dieser Arbeit soll der Einsatz domänenspezifischer Sprachmodelle zur semantischen Modellierung untersucht werden. Zur wissenschaftlichen Methodik gehört zunächst die Einarbeitung in den aktuellen Stand der Technik mit einem Fokus auf verschiedene Sprachmodelle. Weiter soll geeignetes Textmaterial identifiziert werden um Sprachmodelle für zwei unterschiedliche Domänen zu trainieren bzw. zu adaptieren. Abschließend sollen die Sprachmodelle mit Hilfe eines vorliegenden Datensatzes bezüglich ihrer Eignung zur semantischen Modellierung im Rahmen unterschiedlicher Experimente evaluiert werden um eine finale Empfehlung zum Einsatz von Sprachmodellen abzugeben.

## Benötigte Kenntnisse

- Programmierkenntnisse in Python
- Erfahrungen im Bereich der Natürlichen Sprachverarbeitung (NLP)
- Erfahrungen im Bereich des Maschinellen Lernens

## Art der Arbeit

Bachelor-/Masterarbeit

## Ansprechpartner

Andreas Burgdorf, M.Sc. | Tel.: +49 202 439 1047 | E-Mail: [burgdorf@uni-wuppertal.de](mailto:burgdorf@uni-wuppertal.de)