

Masterarbeit

Das semantische Curriculum: Semantic Web und Linked Data zur formalen Wissensrepräsentation teilstrukturierter deutscher Sätze.

Kurzbeschreibung

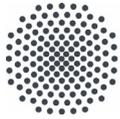
Diese Arbeit richtet sich an Studierende mit Interesse an Semantic Web Technologien, erfordert aber keine speziellen Vorkenntnisse in diesem Bereich. Um natürlich-sprachlich formulierte Gegebenheiten eindeutig maschinell auswerten zu können, müssen diese in einer *Controlled Natural Language (CNL)* vorliegen.

Ziel der Arbeit ist die Überführung von semi-strukturierten Sätzen in deutscher Sprache in maschineninterpretier- und querybare Formate. Als Datenbasis dienen dabei bestehende Lernzielformulierungen in einem Studiengang. Lernzielformulierungen verfügen über typische syntaktische Merkmale, was die Entwicklung einer einfachen und geeigneten *Controlled Natural Language (CNL)* erleichtert. Neben der Entwicklung einer geeigneten CNL liegt ein weiterer Schwerpunkt dieser Arbeit auf der Erarbeitung und Erprobung von Kennwerten zur Abbildungstreue der entwickelten CNL.

Hintergrund und Details

Nicht nur im Bereich Industrie 4.0 sowie beim Austausch von Informationen zwischen Maschinen und entlang komplexer Liefer- und Wertschöpfungsketten gewinnen interoperable Datenformate an Bedeutung: Die Möglichkeit große und komplexe Wissensbestände und Datensammlungen computerunterstützt eindeutig auswerten zu können, gewinnt im Zuge der Digitalisierung in einer Vielzahl von Kontexten an Bedeutung. Dafür müssen die entsprechenden Informationselemente in einem maschinell eindeutig interpretierbaren Format vorliegen. Als von individuellen technischen Realisierungen unabhängige formale Wissensrepräsentation und als Austauschformat zwischen Akteur*innen und Anwendungen setzt sich zunehmend die Modellierung des Wissens in Form von Linked-Data durch. Sogenannte Ontologien stellen dabei sowohl die konzeptionelle Modellierungsumgebung dar, als auch fertig modellierte Sachverhalte und Wissensbestände. Sollen bzw. müssen die Datensätze – wie es häufig der Fall ist – durch Menschen aktuell gehalten werden, stellen sich dabei die Probleme, dass die ontologiekonforme Dateneingabe (a) die Einarbeitung der Anwender*innen in die entsprechenden Editoren erfordert und (b) die Dateneingabe und Pflege von Metadaten zusätzlichen Aufwand für die Anwender*innen bedeutet, so dass in der Praxis die kollaborative Pflege eines formalen Wissensbestandes durch viele Akteure häufig wenig Anklang findet. Das gilt besonders für dynamische, häufigen oder regelmäßigen Veränderungen unterworfenen Wissensbestände und Datensätze, die häufig aktualisiert werden müssen.

Ein Lösungsansatz für dieses Problem besteht darin, den Anwendenden eine an die natürliche menschliche Sprache angelehnte *Controlled Natural Language (CNL)* zur Verfügung zu stellen, welche nur eindeutig interpretierbare Satzkonstruktionen zulässt. Damit können Sätze und Sachverhalte formuliert werden, die sowohl für Menschen natürlich wirken, als auch eindeutig maschinell interpretierbar sind.



Ziel der Arbeit ist die Überführung von vorliegenden, semi-strukturierten Sätzen in deutscher Sprache in maschineninterpretier- und querybare Formate. Als Datenbasis dienen dabei bestehende Lernzielformulierungen in einem Studiengang. Lernzielformulierungen verfügen über typische syntaktische Merkmale, was die Entwicklung einer einfachen und geeigneten Controlled Natural Language erleichtert. Gleichzeitig eröffnen sich durch eine eindeutig maschineninterpretierbare Abbildung der Lernziele von Bildungsangeboten hochinteressante Analysemöglichkeiten.

Aufgabenstellung

- Nutzung (optional auch: Auswahl) eines geeigneten Ontologie-Editors
- Anbindung (optional auch: Auswahl) geeigneter bestehender Vokabulare an den Ontologie-Editor
- Entwicklung einer ersten einfachen und vorläufigen CNL Version
- Erprobung der vorläufigen CNL an einem Set realer natürlich-sprachlich formuliert vorliegender Lernzielformulierungen und Weiterentwicklung der CNL anhand der Erfahrung mit der Umsetzung der realen Lernzielformulierungen
- Erarbeitung und Anwendung geeigneter Kennwerte um Aussagen treffen zu können über die Ausdrucksmächtigkeit der CNL und das Ausmaß der Abbildung des Sets realer Lernzielformulierungen
- Erprobung von Queries über dem nun maschinell eindeutig auswertbaren Set von Lernzielen und Erprobung und Veranschaulichung von Visualisierungsmöglichkeiten der Query-Ergebnisse

Profil der Arbeit

Themenart: theor./konzeptionell

Voraussetzungen

Interesse an Semantic Web Technologien. Spezielle Vorkenntnisse oder Programmierkenntnisse sind nicht notwendig, können aber selbstverständlich eingebracht werden, falls vorhanden.

Betreuung



M.A. Dipl.-Ing. (FH) Jan Wunderlich
jan.wunderlich@ids.uni-stuttgart.de
Tel.: 0711 – 685 60707
www.ids.uni-stuttgart.de