

Thesen zur Bachelorarbeit

Multimodale Datenakquisition, Veredelung und Visualisierung nationaler Wissenschaftsinstitutionen

Die CORDIS-Datenbank¹ enthält alle Datensätze zu Forschungsprojekten sucht. Die Datensätzen beziehen sich auf Projekte und Organisationen in unterschiedlichen Zeiträumen. Der für diese Arbeit erforderliche Datensatz findet sich im Rahmenprogramm „Horizont 2020“ für Forschung und Innovation von 2014 bis 2020. Dieser Datensatz enthält viele unterschiedliche Date. Die Datei 'H2020-Projekte' enthält die Informationen zu öffentlichen Zuschüssen für jedes Projekt. Die Datei 'H2020 Organisationen', die enthält die Organisationen, die von der Europäischen Union im Rahmen des Rahmenprogramms „Horizont 2020“ für Forschung und Innovation von 2014 bis 2020 finanziert werden. Die Datei enthält wichtige Informationen über die Organisationen. Die Daten sind im CSV-Format.

Arbeitsprozess:

1. Sparql Protocol And Rdf Query Language (SPARQL) ist Abfragesprache für RDF². Um nach dem richtigen Datensatz zu suchen, werden spezielle Techniken wie SPARQL³ verwendet.
2. Die Daten werden vom CSV-Format in das XML / RDF-Format konvertiert.
3. Die RDFLIB-Bibliothek⁴ bietet die Möglichkeit, RDF-Daten in einem Wissensgraph-Speicher zu speichern. Die Daten werden im Wissensgraph-Speicher gespeichert.
4. Nach der Speicherung der Daten mithilfe von RDFLib wird JSON-LD⁵ (JavaScript Object Notation for Linked Data) benutzt um zur Visualisierung zu gelangen.
5. Daten werden auf verschiedene Arten visualisiert werden. In dieser Arbeit wird die bekannte Bibliothek Vega⁶ eingesetzt, da sie ein Vielzahl an Visualisierungsmöglichkeiten bietet.
6. Nach alle Implementierungsschritte beschrieben werden, welche für die Zielerreichung nötig sind, werden die Ergebnisse im Detail untersucht und beurteilt.

1 CORDIS Europäische Kommission: Community Research and Development Information Service

2 RDF: Resource Description Framework

3 SPARQL Query Language for RDF: <https://www.w3.org/TR/rdf-sparql-query/>

4 RDFLIB ist ein RDF-JavaScript-Bibliothek, kann auch mit anderen Sprachen wie Python verwendet werden.

5 JSON-LD ist ein leichtes Linked Data-Format. Menschen können leicht lesen und schreiben. Es basiert auf dem bereits erfolgreichen JSON-Format.

6 Vega ist eine Javascript-Bibliothek, die interaktive Grafiken bietet.