

Semantic Web-Anwendung zur Analyse von Bevölkerungsdaten in deutschen Städten

Lukas Ruminski, Nived Stephen und Sarah Kiki

30. August 2023

gitlab.imn.htwk-leipzig.de/semantic-web-migration

HTWK Leipzig - Semantic Web

Inhaltsverzeichnis

1	Einleitung	1
1.1	Hintergrund	1
1.2	Zielsetzung	1
1.3	Methodik	1
2	Relevante Datenquellen	2
2.1	Bevölkerung insgesamt und Ausländeranteil	2
2.2	Ausländische Bevölkerung	2
2.3	Schutzsuchende	2
2.4	MediaWiki Action API	3
3	Ontologie	3
3.1	Vokabular	3
3.1.1	Präfixe	3
3.1.2	Klassendefinitionen (Class Definitions)	4
3.1.3	Objekteigenschaften (Object Properties)	4
3.1.4	Datentyp-Eigenschaften (Datatype Properties)	4
3.1.5	Ontologie (Graphisch)	5
4	Extraktion und Aufbereitung relevanter Daten	6
5	SPARQL-Anfrage	7
5.1	Kompetenzfragen und Ergebnisse	7
6	Ergebnisse und Bedeutung	10

1 Einleitung

Die vorliegende Forschungsarbeit widmet sich der Entwicklung und Umsetzung einer innovativen Semantic Web-Anwendung zur tiefgreifenden Analyse von Bevölkerungsdaten in verschiedenen Städten Deutschlands. Im Fokus steht die Verknüpfung von semantischer Modellierung und Datenanalyse, um Einblicke in die Bevölkerungsstruktur, Migrationstrends und demografische Veränderungen auf höchst präzise und bedeutungsvolle Weise zu gewinnen.

1.1 Hintergrund

Die zunehmende Verfügbarkeit großer Mengen an Bevölkerungsdaten eröffnet die Möglichkeit, relevante Erkenntnisse über soziale, wirtschaftliche und demografische Phänomene zu gewinnen. Allerdings erfordert die Interpretation solcher Daten eine robuste Methodik, die über herkömmliche statistische Ansätze hinausgeht. Das Semantic Web bietet hierbei eine vielversprechende Lösung, indem es strukturierte, vernetzte und semantisch angereicherte Datenmodelle ermöglicht.

1.2 Zielsetzung

Das Hauptziel dieser Forschungsarbeit besteht darin, die Synergie zwischen dem Semantic Web und der Analyse von Bevölkerungsdaten zu nutzen, um tiefgreifende Einblicke in das Verhalten, die Verteilung und die Veränderungen der Bevölkerung in verschiedenen deutschen Städten zu erhalten.

Durch die Entwicklung einer spezialisierten Ontologie und die Verknüpfung von Bevölkerungsdaten in einem semantischen Kontext streben wir danach, Fragen von sozialer und demografischer Bedeutung zu beantworten und bisher unentdeckte Zusammenhänge aufzudecken.

1.3 Methodik

Die Forschungsarbeit gliedert sich in mehrere Schritte:

1. Datenerfassung und Modellierung:
Bevölkerungsdaten von verschiedenen deutschen Städten, einschließlich Informationen zu Ausländern, Schutzsuchenden, Altersgruppen und Aufenthaltsdauern, werden erfasst und in einer eigens entwickelten Ontologie modelliert.
2. Tripelgenerierung und semantische Anreicherung:
Basierend auf der Ontologie werden Bevölkerungsdaten in Tripel um-

gewandelt und semantisch angereichert, um Beziehungen zwischen den Datenpunkten herzustellen.

3. Integration in Graph Datenbank:

Die semantisch angereicherten Daten werden in eine Graph Datenbank hochgeladen, um eine effiziente Speicherung, Abfrage und Verknüpfung der Daten zu ermöglichen.

4. SPARQL-basierte Analyse:

Mithilfe von SPARQL-Abfragen werden gezielte Fragestellungen beantwortet, um Erkenntnisse über Bevölkerungsstruktur, Migrationstrends und demografische Veränderungen zu gewinnen.

2 Relevante Datenquellen

Wir haben nur vollständige Datenquellen für das Jahr 2022 gefunden. Mit der Wayback Machine war es uns möglich, auch Daten für das Jahr 2021 zu erhalten. Deshalb haben wir uns nur auf diese Jahre fokussiert.

2.1 Bevölkerung insgesamt und Ausländeranteil

Die Daten stammen vom Statistischen Bundesamt (Destatis) und wurden bis zum 31. Dezember 2021 fortgeschrieben. Diese Daten sind auf den offiziellen Websites www.destatis.de und www.statistik-portal.de verfügbar.

2.2 Ausländische Bevölkerung

Die Daten wurden zum 31. Dezember 2022 im Ausländerzentralregister (AZR) erfasst und sind ebenfalls auf www.destatis.de zu finden.

2.3 Schutzsuchende

Die Daten wurden ebenfalls bis zum 31. Dezember 2022 im Ausländerzentralregister (AZR) erfasst. Informationen über registrierte Schutzsuchende können aus der GENESIS-Online-Datenbank des Statistischen Bundesamts gewonnen werden. Die Daten sind in Form von Tabellen im XLSX-, XLS-, CSV- und HTML-Format herunterladbar. Zudem bietet das Statistische Bundesamt die Möglichkeit, die Daten in Diagrammen, interaktiven Karten anzuzeigen und eine API-Programmierschnittstelle zur automatisierten Datenverarbeitung zu nutzen.

2.4 MediaWiki Action API

Um detaillierte Informationen über einzelne Städte zu erhalten, haben wir die MediaWiki-API genutzt, um Links zu den individuellen Städteseiten zu generieren. Diese Links ermöglichen den direkten Zugriff auf umfassende Informationen über jede Stadt, einschließlich Bevölkerungsstatistiken, geografischer Merkmale und kultureller Aspekte. Die Verwendung der MediaWiki-API hat es uns ermöglicht, gezielte Daten für verschiedene Städte abzurufen und eine umfassendere Analyse der regionalen Unterschiede und Entwicklungen vorzunehmen.

3 Ontologie

Unsere Ontologie sollte den demografischen Wandel in Deutschland darstellen. Um die Ontologie zu erstellen, haben wir die Methode nach Grüninger und Fox genutzt. Um die richtigen Klassen und Eigenschaften zu bestimmen, haben wir Kompetenzfragen entwickelt, die mithilfe der Ontologie beantwortet werden können. Wir haben den WebVOWL Editor erfolgreich genutzt, um unsere Ontologie zu erstellen. Diese Plattform bietet eine interaktive Web-Benutzeroberfläche, die es uns ermöglicht, die Ontologie einfach und visuell ansprechend zu entwickeln. Die intuitive Schnittstelle erlaubt es uns, komplexe Beziehungen und Konzepte leicht zu gestalten und Änderungen in Echtzeit zu visualisieren. Dies erleichtert den Entwicklungsprozess und führt zu einer klareren Repräsentation unserer Daten und Strukturen.

3.1 Vokabular

3.1.1 Präfixe

Die Präfixe, die für die Ontologie verwendet werden, sind:

```
mo: <https://gitlab.imn.htwk-leipzig.de/semantic-web-migration/
migration-backend/-/blob/main/Ontology/migrationOntology.ttl#> .
rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
owl: <http://www.w3.org/2002/07/owl#> .
xsd: <http://www.w3.org/2001/XMLSchema#> .
dc: <http://purl.org/dc/elements/1.1/#> .
```

3.1.2 Klassendefinitionen (Class Definitions)

In diesem Abschnitt werden verschiedene OWL-Klassen definiert, die im Ontologie-Modell verwendet werden. Jede Klasse repräsentiert einen bestimmten Konzepttyp wie Stadt, Bevölkerung oder Ausländer. Diese Klassen dienen als Grundbausteine für die Modellierung von Beziehungen und Eigenschaften.

rdfs:label	rdf:type
City	owl:Class
Population	owl:Class
Foreigners	owl:Class
Germans	owl:Class
ProtectionSeekers	owl:Class
Agegroup	owl:Class
Duration_of_stay	owl:Class
Residence_Status	owl:Class

3.1.3 Objekteigenschaften (Object Properties)

Hier werden die Beziehungen zwischen den OWL-Klassen durch Objekteigenschaften beschrieben. Diese Eigenschaften definieren, wie verschiedene Klassen miteinander in Beziehung stehen. Beispielsweise wird die Beziehung zwischen Stadt und Bevölkerung durch die Eigenschaft `hatBevölkerung` definiert, die angibt, welche Bevölkerung in welcher Stadt lebt.

rdfs:label	rdf:type	rdfs:domain	rdfs:range
hasPopulation	owl:ObjectProperty	City	Population
hasForeigners	owl:ObjectProperty	Population	Foreigners
hasGermans	owl:ObjectProperty	Population	Germans
PopulationAgegroup	owl:ObjectProperty	Population	Agegroup
ForeignersAgegroup	owl:ObjectProperty	Foreigners	Agegroup
GermanAgegroup	owl:ObjectProperty	Germans	Agegroup
hasDurationOfStay	owl:ObjectProperty	Foreigners	Duration_of_stay
hasResidenceStatus	owl:ObjectProperty	Foreigners	Residence_Status
hasProtectionSeekers	owl:ObjectProperty	Population	ProtectionSeekers

3.1.4 Datentyp-Eigenschaften (Datatype Properties)

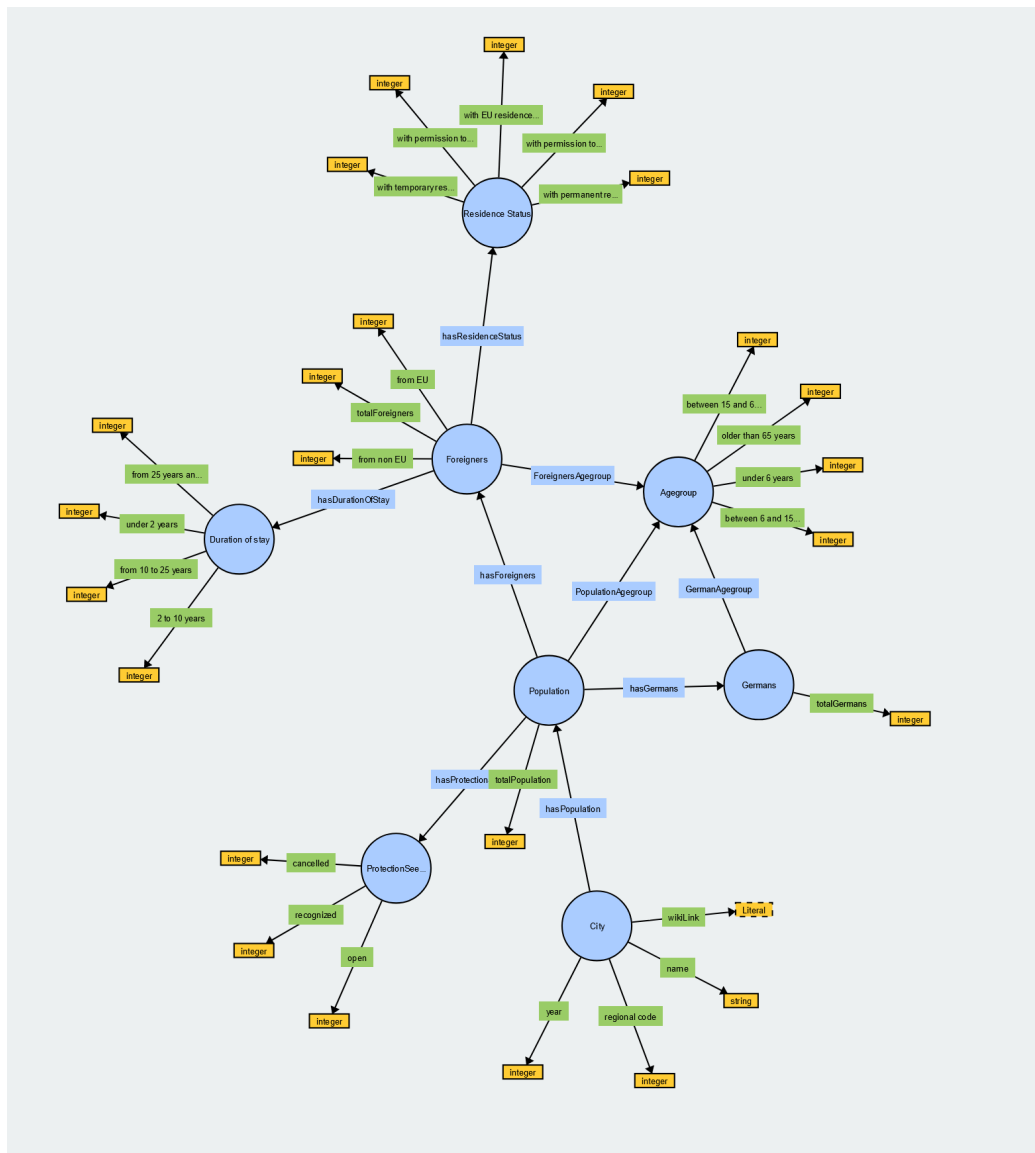
Dieser Abschnitt behandelt Eigenschaften, die konkrete Datenwerte speichern, wie z.B. den Namen einer Stadt, das Jahr ihrer Gründung oder die Gesamtbevölkerung. Diese Eigenschaften sind mit bestimmten Klassen und

Datentypen verknüpft, um die Informationen in der Ontologie genau zu definieren.

rdfs:label	rdfs:domain	rdfs:range
name	City	xsd:integer
year	City	xsd:integer
wikiLink	City	xsd:integer
regional_code	City	xsd:integer
totalPopulation	Population	xsd:integer
totalForeigners	Foreigners	xsd:integer
fromEU	Foreigners	xsd:integer
from_non_EU	Foreigners	xsd:integer
totalGermans	Germans	xsd:integer
with_temporary_residence_permit	Residence_Status	xsd:integer
with_permanent_residence_permit	Residence_Status	xsd:integer
with_permission_to_remain	Residence_Status	xsd:integer
with_permission_to_remain_unill_deported	Residence_Status	xsd:integer
with_EU_residence_permit	Residence_Status	xsd:integer
recognized	ProtectionSeekers	xsd:integer
cancelled	ProtectionSeekers	xsd:integer
open	ProtectionSeekers	xsd:integer
under2years	Duration_of_stay	xsd:integer
2to10years	Duration_of_stay	xsd:integer
from10to25years	Duration_of_stay	xsd:integer
from_25_years_and_more	Duration_of_stay	xsd:integer
under_6_years	Agegroup	xsd:integer
between_6_and_15_years	Agegroup	xsd:integer
between_15_and_65_years	Agegroup	xsd:integer
older_than_65_years	Agegroup	xsd:integer

3.1.5 Ontologie (Graphisch)

Alle für die Ontologie erforderlichen Dateien befinden sich im Ordner Ontologie in unserem Repository



4 Extraktion und Aufbereitung relevanter Daten

Die vorliegenden Daten wurden zunächst aus einer CSV-Datei extrahiert und anschließend in das JSON-Format umgewandelt. Diese Umwandlung erfolgte, um einen verbesserten Zugriff und eine flexiblere Verarbeitung der Daten zu ermöglichen.

ten zu ermöglichen. Die umgewandelten JSON-Daten bildeten die Grundlage für die nächste Entwicklungsphase. Hierbei wurde ein speziell entwickeltes TypeScript-Programm erstellt, das sich als äußerst leistungsfähig erwiesen hat. Das Programm ist in der Lage, die JSON-Daten zu lesen und gezielt darauf zuzugreifen. Dieser Zugriff bildet die Grundlage für die darauf folgende Datenverarbeitung.

In einem weiteren Schritt wurde die Funktionalität implementiert, um die JSON-Daten in sogenannte Tripel zu transformieren. Diese Transformation war entscheidend, um semantische Beziehungen und Zusammenhänge zwischen den Datenpunkten herzustellen. Dabei kam eine vordefinierte Ontologie zum Einsatz, die es ermöglichte, die Daten in ein strukturiertes und aussagekräftiges Format zu bringen.

Die generierten Tripel bilden die Grundlage für die Erstellung von Turtle-Dateien. Hierbei wurden nicht nur die eigentlichen Daten einbezogen, sondern auch Namensräume, die die semantische Bedeutung der Datenpunkte verdeutlichen. Die Turtle-Dateien stellen somit eine hochwertige und umfassende Darstellung der Daten dar, die sich ideal für weitere Analysen und Interpretationen eignet.

Die Quellcodedateien in TypeScript zur Realisierung dieser Prozesskette sind verfügbar und befinden sich im übersichtlichen `src` Ordner. Die Kompilierung der TypeScript-Dateien erfolgt mit dem Befehl `tsc`. Die Ausführung des Befehls `node tripleGenerator.js > data2022.ttl` führt zur Generierung der Turtle-Dateien, die die Daten in einer für Menschen lesbaren und maschinenverarbeitbaren Form präsentieren. Die generierten Tripel-Daten werden im Ordner `RDF Triples` in unserem GitLab-Repository gespeichert.

5 SPARQL-Anfrage

Unsere Turtle Dateien wurden erfolgreich in GraphDB hochgeladen. Dadurch können wir gezielte SPARQL-Abfragen auf Grundlage dieser Daten durchführen. Dies ermöglicht uns, spezifische Informationen zu unseren Kompetenzfragen zu extrahieren und zu analysieren. Die Interaktivität von GraphDB ermöglicht es uns, die Abfragen direkt zu testen und relevante Ergebnisse zu erhalten.

5.1 Kompetenzfragen und Ergebnisse

Die Kompetenzfragen, zusammen mit den dazugehörigen SPARQL-Abfragen und den erzielten Ergebnissen, sind wie folgt formuliert:

1. Wie viele Ausländer leben in Leipzig?

```
1 PREFIX mo: <https://gitlab.imn.htwk-leipzig.de/semantic-web-migration/migration-backend/-/blob/main/Ontology/migrationOntology.ttl#>
2 SELECT ?name ?totalForeigners
3 WHERE {
4   ?subject mo:name ?name;
5           mo:name "Leipzig"@en;
6           mo:year 2022;
7           mo:hasPopulation ?population ;
8           mo:regional_code 14713 .
9   ?population mo:hasForeigners ?Foreigners .
10  ?Foreigners mo:totalForeigners ?totalForeigners.
11 }
```

	name	totalForeigners
1	"Leipzig"@en	"80460"^^xsd:integer

2. Wie viele Ausländer in Hamburg sind älter als 65 Jahre?

```
1 PREFIX mo: <https://gitlab.imn.htwk-leipzig.de/semantic-web-migration/migration-backend/-/blob/main/Ontology/migrationOntology.ttl#>
2 SELECT ?year ?name ?older_than_65_years
3 WHERE {
4   ?subject mo:name ?name;
5           mo:year ?year;
6           mo:name "Hamburg"@en;
7           mo:hasPopulation ?population .
8   ?population mo:hasForeigners ?Foreigners .
9   ?Foreigners mo:ForeignersAgegroup ?Agegroup.
10  ?Agegroup mo:older_than_65_years ?older_than_65_years .
11 }
```

	year	name	older_than_65_years
1	"2022"^^xsd:integer	"Hamburg"@en	"28723"^^xsd:integer
2	"2021"^^xsd:integer	"Hamburg"@en	"27223"^^xsd:integer

3. Was ist die Wikiseite von Leipzig?

```
1 PREFIX mo: <https://gitlab.imn.htwk-leipzig.de/semantic-web-migration/migration-backend/-/blob/main/Ontology/migrationOntology.ttl#>
2 SELECT ?name ?wikiLink
3 WHERE {
4   ?city mo:name ?name;
5         mo:year 2021;
6         mo:regional_code 14713;
7         mo:name "Leipzig"@en;
8         mo:wikiLink ?wikiLink .
9 }
```

	name	wikiLink
1	"Leipzig"@en	https://en.wikipedia.org/wiki/Leipzig

4. Wie viele Asylsuchende gibt es in Deutschland?

1	PREFIX mo: <https://gitlab.imn.htwk-leipzig.de/semantic-web-migration/migration-backend/-/blob/main/Ontology/migrationOntology.ttl#>
2	SELECT (SUM(?recognized) as ?totalRecognized)
3	WHERE {
4	?subject mo:name ?name ;
5	mo:hasPopulation ?population .
6	?population mo:hasProtectionSeekers ?protectionSeekers .
7	?protectionSeekers mo:recognized ?recognized .
8	}
9	
10	
totalRecognized	
1	"3508550"^^xsd:integer

5. Wie viele Ausländer leben in Frankfurt schon länger als 25 Jahre?

1	PREFIX mo: <https://gitlab.imn.htwk-leipzig.de/semantic-web-migration/migration-backend/-/blob/main/Ontology/migrationOntology.ttl#>
2	SELECT ?name ?from_25_years_and_more
3	WHERE {
4	?subject mo:name ?name;
5	mo:year 2022;
6	mo:name "Frankfurt am Main"@en;
7	mo:hasPopulation ?population .
8	?population mo:hasForeigners ?foreigners .
9	?foreigners mo:hasDurationOfStay ?duration_of_stay .
10	?duration_of_stay mo:from_25_years_and_more ?from_25_years_and_more .
11	}
12	
name	
from_25_years_and_more	
1	"Frankfurt am Main"@en
	"67710"^^xsd:integer

6. Wie hat sich die Anzahl der Ausländer zwischen 2021 und 2022 in Leipzig verändert?

1	PREFIX mo: <https://gitlab.imn.htwk-leipzig.de/semantic-web-migration/migration-backend/-/blob/main/Ontology/migrationOntology.ttl#>
2	SELECT ?name ?year ?totalForeigners
3	WHERE {
4	?subject mo:name ?name;
5	mo:name "Leipzig"@en;
6	mo:year ?year;
7	mo:hasPopulation ?population ;
8	mo:regional_code 14713 .
9	?population mo:hasForeigners ?Foreigners .
10	?Foreigners mo:totalForeigners ?totalForeigners.
11	}
name	
year	
totalForeigners	
1	"Leipzig"@en
	"2022"^^xsd:integer
	"80460"^^xsd:integer
2	"Leipzig"@en
	"2021"^^xsd:integer
	"55665"^^xsd:integer

7. Wie viele Ausländer leben schon länger als 25 Jahre in Deutschland?

1	PREFIX mo: <https://gitlab.imn.htwk-leipzig.de/semantic-web-migration/migration-backend/-/blob/main/Ontology/migrationOntology.ttl#>
2	SELECT (SUM(?from_25_years_and_more) AS ?total_25)
3	WHERE {
4	?subject mo:year 2022;
5	mo:hasPopulation ?population .
6	?population mo:hasForeigners ?foreigners .
7	?foreigners mo:hasDurationOfStay ?duration_of_stay .
8	?duration_of_stay mo:from_25_years_and_more ?from_25_years_and_more .
9	}
total_25	
1	"2935165"^^xsd:integer

6 Ergebnisse und Bedeutung

Die vorliegende Forschungsarbeit hat eine wegweisende Semantic Web Anwendung zur Bevölkerungsdatenanalyse in deutschen Städten entwickelt. Die Verschmelzung von semantischer Modellierung und Datenanalyse erlaubte profunde Einblicke in Bevölkerungsstruktur, Migrationstrends und Demografie. Dieser Ansatz übertraf konventionelle statistische Methoden. Die eigens entwickelte Ontologie bildete das Grundgerüst, wobei durch Kompetenzfragen passende Klassen, Eigenschaften und Beziehungen definiert wurden. Der WebVOWL Editor erleichterte die Gestaltung komplexer Zusammenhänge visuell. Schritte wie Datenerfassung, Tripelgenerierung und Graphdatenbankintegration ermöglichten gezielte SPARQL-Abfragen.

Die Ergebnisse waren beeindruckend von Ausländeranteilen bis zu Asylsuchenden konnten relevante Daten gewonnen werden. Die Arbeit stellt nicht nur Antworten auf Forschungsfragen bereit, sondern auch wertvolle Erkenntnisse für soziale Dienste, Stadtplanung und Politik. Zusammenfassend eröffnet die Studie nicht nur eine effiziente Semantic Web-Anwendung zur Bevölkerungsdatenanalyse, sondern zeigt auch neue Wege, wie das Semantic Web zur besseren Interpretation komplexer sozialer Phänomene genutzt werden kann. Die Verbindung von semantischer Modellierung, Datenanalyse und SPARQL-Abfragen verleiht der Auswertung von Bevölkerungsdaten eine präzisere und tiefgreifendere Bedeutung, mit Potenzial für vielfältige Anwendungen in Forschung und gesellschaftlicher Gestaltung.