

Studierende auf dem
Leipziger Semantic Web Tag 2025
Future Knowledge Experts

Paulo do Carmo und Thomas Riechert
Hochschule für Technik, Wirtschaft und Kultur (HTWK) Leipzig

12.06.2025

Hochschule für Technik Wirtschaft und Kultur Leipzig

Main Campus located at Karl-Liebknecht-Straße, Leipzig

- ca. 6.500 students within 6 faculties

(Source: <https://www.htwk-leipzig.de/hochschule/ueber-uns>)

- 40 degree courses

- Faculty of Architecture and Social Sciences
- Faculty of Civil Engineering
- **Faculty of Computer Science and Media (ca.1.800 students)**
- Faculty of Engineering
- Faculty of Business Administration and Industrial Engineering
- Faculty of Digital Transformation

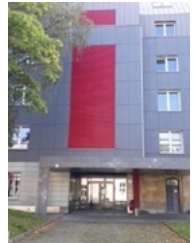


Quelle: Wikipedia, Lange Nacht der Computerspiele

Faculty Computer Science and Media

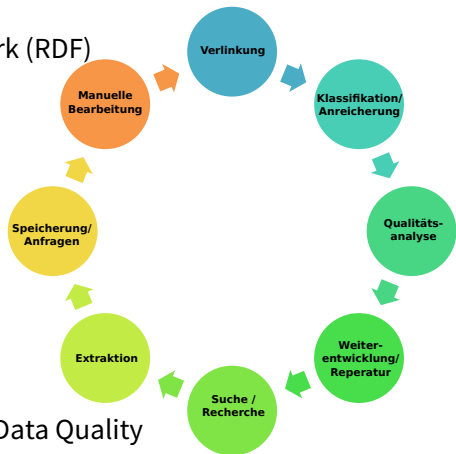
■ 18 degree courses

- Bibliotheks- und Informationswissenschaft (B.A., M.A.)
- Buch- und Medienproduktion (B.Eng.)
- Buch- und Medienwirtschaft (B.A.)
- Digitale Print-Technologien / Drucktechnik (B.Eng.)
- Druck- und Verpackungstechnik (M.Eng.)
- Fernsehproduktion (B.A.)
- **Informatik (B.Sc., M.Sc.)**
- **Medieninformatik (B.Sc., M.Sc.)**
- Medienmanagement (M.Eng.)
- Medientechnik (B.Eng.)
- Museologie (B.A.)
- Museumspädagogik | Bildung und Vermittlung im Museum (M.A.)
- Publishing Management (M.A.)
- Verpackungstechnologie und Nachhaltigkeit / Verpackungstechnik (B.Eng.)



Inhalte: Modul Semantic Web / 5. Semester B.Sc.

- ✓ Einführung / Linked Data
- ✓ Resource Description Framework (RDF)
- ✓ RDF Serialisierungen /
Abfragesprache SPARQL
- ✓ RDF-Schema
- ✓ OWL Syntax
- ✓ Linked Open Data
- ✓ Semantische Suche / Semantic Search
- ✓ Anwendungen
- ✓ Datenqualität im Web / Linked Data Quality
- ✓ Ontology Engineering
- ✓ Weitere Aspekte



25 Studierende / 25 participating students

Students Posters

- 15 Posters from 26 students talking about their projects of data semantification

Amazon book reviews and user preferences

■ Author: Erik Kessler

Amazon book reviews and users preferences

Sources



Kaggle is an online-community, that specializes on data science.

Format: RDF
Interface: SPARQL
License: CC0
Data: Amazon Reviews



DBpedia is a community project, that transforms wiki data into the Linked Open Data concept.

Format: RDF
Interface: SPARQL
License: GNU
Data: Background knowledge

Simplified Ontology



Workflow

1. Define the timespan

- Analyses of the recent reviewed books over the time of a year

2. Integrate outsource data

- Detect name of author and work
- Gather information of other works and life of the author

3. Interpret user data

- Search of other reviews from same user ID
- Establish link of interest between books

4. Secure results in datafile

Outlook

This ontology is made by a dataset from 2021. A updated statistic allows a current look into the link between users reviews and their interest. Additionally, the ontology can be expanded into:

- **Comparing sales to score and reviews**
- **Anticipating the purchase or review of the user**, in dynamic to the same author, authors of the same genre, authors of the same time period, etc.
- **Involvement of other sources, the user is influenced by**

Microsoft News Dataset: A Semantic Web Exploration

- Authors: Anna Liepelt, Tristan Sopauschke

Microsoft News Dataset

- A Semantic Web Exploration -

We aim to explore patterns in user behavior and news.com article popularity by transforming raw data into RDF and querying it semantically.

Data Source

- Dataset:** Microsoft News Dataset (MIND)
- Scale:** 160,000 news articles, 1 million users
- Format:** .json (5-Star Open Data)
- Content:** For each impression, it records which articles were shown and which were clicked.

Impressions are grid-based news recommendations dynamically generated for each user.

Example of the triple stored data:

```

mind:user a mind:Impression ;
mind:impressionid "95" ;
mind:impressionarticle mind:news07739 .

mind:news07739 a mind:newsarticle ;
mind:newsarticleid mind:news07739 ;
mind:newsarticleyear 2008 .

```

Ontology Overview

Our ontology models the relationships between users, impressions, and news articles.

The ontology uses both custom vocabulary (mind) and standard vocabularies (rdf, rdfs, xsd).

We link news article entities to Wikidata for semantic enrichment.

```

SELECT ?user ?news {
  ?user mind:impression ?imp .
  ?imp mind:impressionarticle mind:article .
  ?article rdfs:label ?label .
  ?article mind:category ?cat .
}

```

Example of a SPARQL query "Which users who clicked on specific articles also clicked on similarly rated?"

Procedure

Choose a dataset: MIND Dataset

Create a Turtle ontology to fit the data

Convert data into RDF with Python

Load RDF data into a triple store

Query the data using SPARQL

Ontology visualized using OWL and WebVOWL

HTWK Hochschule für Technik, Wirtschaft und Kultur Leipzig

Authors: Anna Liepelt, Tristan Sopauschke

Association: Semantic Web, HTWK Leipzig, SoSe25

Goodreads book review

- Authors: Anastasia Christabel Utomo, Minh Huy Nguyen, Coffi Gracia Wilfrid Akplogan



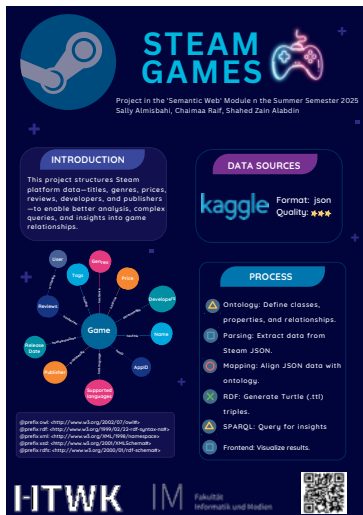
Urban Violence in LA & US Press Freedom

■ Author: Lisa Wilhelm



Steam Games

- Authors: Sally Almisbahi, Chaimaa Raif, Shahed Zain Alabdin



Most Streamed Spotify Songs 2024

■ Authors: Franca Lenz, Sofia Berdichevska

Most Streamed Spotify Songs 2024

What influence do TikTok and YouTube have on the streaming numbers of songs on Spotify – and which platform contributes more to their success?

Project for „Semantic Web“ im Sommersemester 2025
4. Sem. Medieninformatik Bachelor, Fakultät Informatik & Medien
Franca Lenz, Sofia Berdichevska

Datasource



Link www.kaggle.com
Data Song Statistics
CSV ★ ★ ★ ★

Ontology



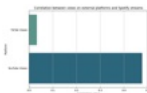
made with:



- Class
- Data Property
- Object Property
- Class Hierarchy

@prefa owl: <https://www.w3.org/2002/07/owl>
@prefa owl: <https://www.w3.org/2002/07/owl> Schema
@base <https://www.w3.org/2002/07/owl>

Data analysis

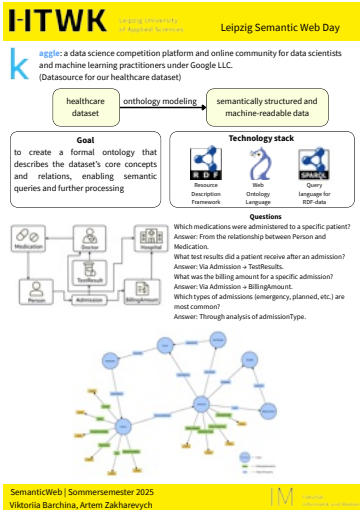


Conclusion:

YouTube has a clear influence on Spotify streams, while TikTok has almost no overall effect on the streaming numbers. YouTube has a significantly greater impact. TikTok only influences Spotify streams in individual cases.

Health care services and providers

■ Authors: Viktoriia Barchina, Artem Zakharevych



The Semantics of the Internet Movie Database - IMDb

■ Author: Pascal Henrici

The Semantics of the Internet Movie Database - IMDb

Data

Paracet

Structure

Semantic

Search Queries

Data Analysis

Machine Learning

HTWK

Problemlösung für Technik, Wirtschaft und Natur

Netflix Movies and TV Shows

- Authors: Ilyass Essar Rhini, Khalil Chafai

HTWK

Hochschule für Technik,
Wirtschaft und Kultur Leipzig

Netflix Movies and TV Shows

Mit der globalen Expansion von Netflix hat sich die Angebotsstruktur stark verändert. Inwiefern hängen auf Länderebene die Anzahl verfügbarer Netflix-Titel, der Anteil lokaler Produktionen sowie die durchschnittliche Zuschauerbewertung je Genre mit dem kulturellen Diversitätsindex und der Internetnutzung zusammen?

Question:

How can RDP and SPARQL be used to determine which Netflix genres are most popular in which countries and at what times?



Most Popular Netflix Genres by Country



Data Sources:



Procedure:


- Read in the CSV file from Kaggle (Netflix Movies and TV Shows) and convert to RDF (R1)
- Extend with additional metadata (e.g., via SPARQL query to IMDb)
- Import the R1 file into a local Apache Jena Fused server
- Formulate SPARQL queries for analysis by genre, country, and release year
- Visualization of the results (e.g., genre distribution in the frontend)

How Political Bias in Newspapers influences the spread of Fake News?

■ Authors: Carl-Moritz von Butler, Steve Rossow

HOW POLITICAL BIAS IN NEWSPAPERS INFLUENCES THE SPREAD OF FAKE NEWS ?

*In the digital age, fake news has become a **widespread phenomenon**. A 2023 survey by the German Federal Statistical Office found that nearly **half of all internet users** had encountered content online they deemed **untrustworthy or false**.*




HTWK

Hochschule für Technik,
Wirtschaft und Kultur Leipzig


DATASET

Fake News Detection Dataset

This dataset is designed for practicing fake news detection using machine learning and natural language processing (NLP) techniques. It includes a rich collection of 20,000 news articles, carefully generated to simulate real-world data scenarios. Each record contains metadata about the article and a label indicating whether the news is real or fake.



ONTOLOGY SCHEMA

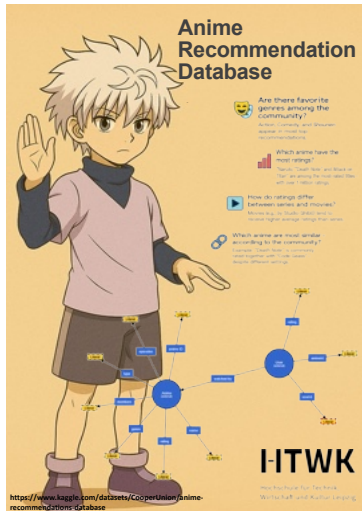


NEXT STEPS

1. Convert Data into RDF
2. Load into a Triple Store
Upload the RDF data to a SPARQL endpoint
3. Visualize and Explore
Explore relationships with graph visualizations and Linked Data tools
4. Ask Smart Questions
"5 authors with the most fake news articles"
"Fake vs. real news ratio of a media organization"

Anime Recommendation Database

■ Author: Maxim Strasser



Analysis of Youtube Trends with RDF and SPARQL

■ Authors: Anass Boustani, Mohammed Abdmugheni

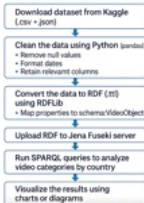
HTWK

Projekt im Modul „Semantic Web“ im
Sommersemester 2025
Anass Boustani, mohammed abdmugheni

ANALYSIS OF YOUTUBE TRENDS WITH RDF AND SPARQL

QUESTION: How can RDF and SPARQL be used
to identify the most popular YouTube video
categories across countries?

PROCEDURE



YouTube RDF Analysis



DATA SOURCES:

