

Master- und Bachelor Studierende auf dem 10. Leipziger Semantic Web Tag 2022

Minutes by Future Knowledge Experts

Vera Meister und Thomas Riechert
Technische Hochschule Brandenburg (THB) und
Hochschule für Technik, Wirtschaft und Kultur (HTWK) Leipzig

05.07.2022

Hochschulen – Universities of Applied Sciences

Technische Hochschule Brandenburg

- <http://www.th-brandenburg.de/>
- Brandenburg an der Havel
- 3 Fachbereiche
- ca. 3000 Studierende



**Technische Hochschule
Brandenburg**
University of
Applied Sciences

Hochschule für Technik, Wirtschaft und Kultur Leipzig (HTWK Leipzig)

- <http://htwk-leipzig.de/>
- Leipzig, Karl-Liebknecht-Straße 132
- 6 Fakultäten
- ca. 6500 Studierende

HTWK

Hochschule für Technik,
Wirtschaft und Kultur Leipzig

Präsentationen

Technische Hochschule Brandenburg

- Projektarbeit
- Modul: Enterprise Knowledge Engineering
- Betreut durch Prof. Dr. Vera G. Meister
- Master Wirtschaftsinformatik

Hochschule für Technik, Wirtschaft und Kultur Leipzig (HTWK Leipzig)

- Zwischenpräsentation der Projektarbeit
- Modul: Semantic Web
- Betreut durch Prof. Dr. Thomas Riechert
- Bachelor Informatik, Bachelor Medieninformatik und Studienrichtung Bibliotheks-informatik

Podcast gesucht? 1/3



Ich suche...

55%

Aller Podcasthörer
nutzen Spotify

Aber warum ist es
so **schwer** etwas
Passendes **zu**
finden?

© Statista 2021

Podcast gesucht? 2/3



**Viel getippt & doch
nichts gefunden?**



Steven Audrey, Theresa Schulz, Lisa Meinetsberger

**Podcast
gesucht?**

Spotify

RSS-Feed

Enterprise Knowledge
Engineering
Projekt SoSe 2022

 **Technische Hochschule
Brandenburg**
University of
Applied Sciences

Podcast gesucht? 3/3



Enterprise Knowledge
Engineering
Abschlussprojekt

Betreut durch: Prof. Dr.
Vera G. Meister

Steven Audrey
Theresa Schulz
Lisa Meinetsberger

SoSe 2022

Wollen Sie mehr wissen?



Besuchen Sie unser Poster*

Suchen

Podcast gesucht?

Oftmals fällt die Suche nach Podcasts in Spotify, Apple Podcasts oder anderen Plattformen schwer. In dieser Präsentation wird die Möglichkeit, Podcasts zu finden, die eine spezifische Thematik abdecken, vorgestellt. Dabei wird die Suche nach Podcasts in Spotify, Apple Podcasts oder anderen Plattformen erleichtert.

Wieher bekommen wir die Daten?

Idee: Die Idee ist, Podcasts zu finden, die eine spezifische Thematik abdecken. Dies wird durch die Analyse von Podcast-Titeln und Beschreibungen erreicht.

Lösung: Die Lösung ist, Podcasts zu finden, die eine spezifische Thematik abdecken. Dies wird durch die Analyse von Podcast-Titeln und Beschreibungen erreicht.

Wie transformieren wir die Daten?

Die Daten werden in eine Form gebracht, die für die Analyse geeignet ist. Dies wird durch die Verwendung von Python und Pandas erreicht.

Abfragen:

- Welche Podcasts sind über 'Data Science' & 'Machine Learning'?
- Welche Podcasts sind über 'Data Science' & 'Machine Learning'?
- Welche Podcasts sind über 'Data Science' & 'Machine Learning'?

*Der Umwelt zu liebe farblich invertiert

Tool: LearnSPARQL

05.07.2022

LEARN
SPARQL

LearnSPARQL:

eine Website zum Lernen und
Üben von SPARQL Abfragen

Leipzig Semantic Web Tag

LearnSPARQL ist eine Website, die den Einstieg in SPARQL vereinfachen, als auch bisherige Fähigkeiten festigen soll.

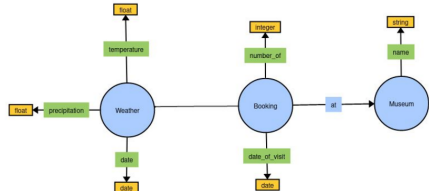
Recherche: Besucherzahlen der Hamburger Kunsthalle

Kann eine Korrelation zwischen der Besucherzahl der Hamburger Kunsthalle und dem Wetter festgestellt werden?

Datenquellen

- Buchungsdaten Hamburger Museen (2013 - 2017)
 - <https://codingdavinci.de>
 - Daten sind als SQL-Datei herunterladbar
 - Tripel: ~ 300.000
- Wetterdaten des Deutschen Wetterdienstes
 - <https://brightsky.dev>
 - Daten sind im JSON-Format über eine Api abrufbar
 - Tripel : 1662

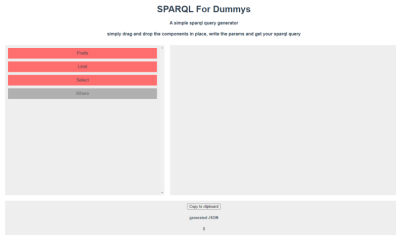
Vokabular



Tool: SPARQL For Dummies

SPARQL For Dummies

Ein einfacher SPARQL Query Generator



Überblick

- Grade für Anfänger kann es kompliziert sein in die Thematik einzusteigen.
- Dieses Tool kann dabei helfen Schnell und zuverlässig eine funktionierende SPARQL Querys zu erstellen.
- Es soll auch möglich sein, komplexere Querys problemlos zu erstellen.
- Per Drag-and-Drop lassen sich die benötigten Komponenten zusammen schieben
- Es werden dabei immer mögliche oder Empholene Schritte vorgeschlagen.

Tool: SparQLy

05.07.2022

Semantic Web

Tag 2022 in Leipzig.



a lightweight SparQL pretty print in JavaScript

SparQLy bietet drei Funktionen:

```
src > example.js
1 import * as sparql from './sparql.js';
2
3 console.log(sparql.pretty("PREFIX dbo: <http://dbpedia.org/ontology/> PREFIX res: \
4 <http://dbpedia.org/resource/> PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> \
5 SELECT DISTINCT ?string WHERE { res:Leipzig dbo:areaCode ?string . }"));
6
7
```

SparQL String

sparqly.pretty.([String])

Pretty String

```
[local@srv01-prd-opr-lpz sparql-pretty-print]$ node ./src/example.js
PREFIX dbo: <http://dbpedia.org/ontology/>
PREFIX res: <http://dbpedia.org/resource/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
SELECT DISTINCT ?string
WHERE {
  res:Leipzig dbo:areaCode ?string .
}
```

- ★Queries werden in Token umgewandelt
- ★Token werden nach Instruktionen sortiert

- ★Instruktionen werden formatiert ausgegeben
- ★Sehr schlank mit >300 Zeilen Code (incl. parser)

Recherche: Musikgenres

HTWK

Hochschule für Technik,
Wirtschaft und Kultur Leipzig

Korrelieren die Musikgenres der Top 10 der Billboard Charts in den Jahren 2010-2019 mit den Genres der erfolgreichsten Konzerttourneen der 2010er?

Datenquellen

last.fm

Link: <https://www.last.fm/api>

Datenformat: XML/JSON

Schnittstelle: Rest-API

Lizenz: Selbst definiert – Registrierung notwendig

Open Data: ***

→ Genres zu Artist*innen

kaggle

Link:

<https://www.kaggle.com/datasets/dhruvil-dave/billboard-the-hot-100-songs>

Datenformat: CSV

Schnittstelle: Download

Lizenz: CC BY-NC-SA 4.0

Open Data: **

→ Artist*innen Top 10 Charts



Link:

https://en.wikipedia.org/wiki/List_of_highest-grossing_concert_tours

Datenformat: HTML

Schnittstelle: JSOUP Parsing

Lizenz: CC BY-SA

Open Data: *

→ Erfolgreichste Konzerttourneen 2010er

Recherche: Bafög

Bafög

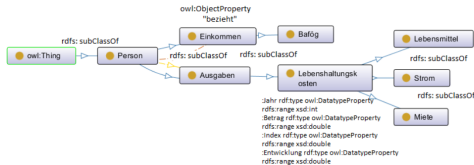
Mehr für dich

Semantic Web

„Wie verhält sich die Entwicklung des Bafög-Höchstsatzes im Vergleich zu der Entwicklung der allgemeinen Lebenshaltungskosten?“

Sommersemester 2022

Projektaufbau



Recherche: Denkmale

Semantische Suche: Liste von Denkmalen in Abhängigkeit vom Wetter

Datenquellen

Coding da Vinci – Denkmaldaten Niedersachsen

[https://codingdavinci.de/daten/
denkmaldaten-niedersachsens](https://codingdavinci.de/daten/denkmaldaten-niedersachsens)

Openweather – API, aktuelle Wetterdaten

<https://openweathermap.org>

Datenextraktion

Projekt: Life Knowledge Base

LIFE KNOWLEDGE BASE: MAKING BIODIVERSITY ACCESSIBLE THROUGH SEMANTIC WEB TECHNOLOGY Finn Mergenthal

Life Knowledge Base is part of an effort to promote and produce open and inter-operable information through a combination of modern technologies. It aims to make the Catalogue of Life, an online database of all known species and the biological hierarchy accessible to research and the public.

The project is work in progress, which means not all planned functionalities are fully implemented. The current state can be found at github.com/AKSW/life-knowledge-base

Catalogue of Life

Catalogue of Life (COL, catalogueoflife.org) is an international project, which aims to make information on all known species and the biological hierarchy accessible to research and the public. It comprehends information from 166 datasets and additionally publishes primary data.

Structure

The COL is only available in multiple column-based formats and cannot be queried automatically. However, it can be downloaded as a Darwin Core Archive using the Darwin Core Standard (DwC, dwc.tdwg.org), a vocabulary for exchanging information on biological diversity.

The Catalogue uses the DwC-Extensions "SpeciesProfile", "VernacularName" and "Distribution", which introduces Composite Primary Keys, thus providing additional Challenges to using the data in a knowledge graph.

Queryability

To let the data of the COL become part of the semantic web I decided to parse the data in triples and host it in a virtuoso triple store.
In the process part of the original data needs to be restructured to fit the triple-format (and avoid using meta-statements, retaining clarity).

Stronger semantics

To make the COL data even more useful to researchers we are trying to add stronger semantics to the original information using simple property exchanging as well as adding new properties and employing owl rules.

Recherche: DNB – Linked Data Services

HTWK

Hochschule für Technik,
Wirtschaft und Kultur Leipzig

IM

Fakultät
Informatik und MedienStudienrichtung Bibliotheks-informatik
Modul: Semantic WebLeipziger Semantic Web Tag
05.07.2022

Deutsche Nationalbibliothek: Linked Data Services

Weitere Verlinkungen

- **WorldCat**
 - (ohne Ontologie, via OCLC-ID)
- **K10plus** u.a. (Verbund-)kataloge
 - (mit owl:sameAs, nur einseitig vom K10plus-Eintrag zum DNB-Datensatz, nicht anders herum)
- **Buchhandel.de**
 - (vielleicht über ISBN oder GTINT)

Daten-Zugang

- kostenloser, gemeinfreier Zugang (CC0)
- **kein SPARQL-Endpoint**
- Gesamtabzüge als **RDF** u.a. (data.dnb.de/opendata/)
- Laufende Aktualisierung per Schnittstelle möglich

Gemeinsame Normdatei

- **Normdaten** als "kontrolliertes Vokabular":
- zur **eindeutigen** Zuordnung von **Entitäten** auch bei Namensgleichheit
- **systematisch** geordnete und **vernetzte** Schlagwörter zur Inhaltserschließung
- entstanden im **Bibliotheksbereich**
- zukünftig für alle **kulturellen und wissenschaftlichen** Sammlungen
- **9 Millionen Datensätze:**



verlinkte Data Provider

Beispiele

- im GND-RDF-Datensatz via owl:sameAs:



Visualisierung und Exploration der GND: explore.gnd.network

Titeldaten der DNB

Metadaten zu **> 42 Millionen** Medienwerken:

Bücher, Zeitschriften, Zeitungen,
CDs, Schallplatten,
Karten,
Online-Publikationen



VERLINKTE DATEN:

beteiligte Personen Autorin, Herausgeberin, Übersetzerin

beteiligte Körperschaften Herausgeber, Verleger, ...

Konferenzen Fachkonferenzen, Sanktionen

Werk-Datensatz bei Übersetzungen, als Thema von Sekundäts

Orte via GeoNames und GeoNames-Id

Sachschlagwörter zur Inhaltserschließung nach ISGDS

Tool: AdayAway

AdayAway - Projekt im Modul Semantic Web

Leipzig, 05.07.2022

Idee

Sommer, Sonne, Flugrecherche – unsere Projektidee! Wir erhielten von Herrn Professor Riechert die Aufgabe, uns eine Recherche/Applikation zu überlegen. Spätestens Corona lernte uns, flexibel zu bleiben. Auch bei Reisebuchungen. Eher spontan gebucht, als langfristig geplant. So war die Idee geboren, eine Website zu programmieren, die es Reisenden ermöglicht, günstige Flüge und schöne Orte für einen Ferienstart innerhalb der nächsten Stunden und Tage zu finden. Gleichzeitig ermöglicht dies Airlines auch unter Klimaschutzaspekten, jeden Platz zu besetzen.

Umsetzung

Für unsere Applikation benötigen wir aktuelle Flüge, Städte und ihre zugehörigen Flughäfen inklusive der IATA-Kennungen und die Wetterdaten an den entsprechenden Zielorten. Als Entwicklungsumgebung nutzen wir Docker mit 3 Containern. Eine MySQL-Datenbank mit den IATA-Kennungen und den dazu gehörigen Städten, eine Python-AdayAway hon-Flask API, die diese Daten verarbeitet und als

Datenquellen

Quelle 1: Selbstprogrammierte IATA-API

Für diese API haben wir die CSV-Dateien von <https://ourairports.com/data> heruntergeladen und in einen MySQL Docker Container geladen. Zusätzlich haben wir mit Python Flask eine Api programmiert, welche über den Python MySQL-connector die Daten aus der Datenbank abfragt und als JSON-Objekt returned. In Bild 01 sieht man eine beispielhafte Abfrage. Hier werden die Städte in IATA-Codes der Flughäfen umgewandelt und umgekehrt.

Quelle 2: OpenWeather API

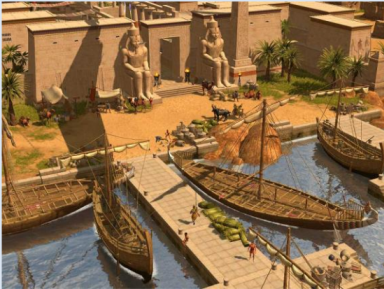
Bei <https://openweathermap.org/api> werden die Wettervorhersage der Zielorte für die nächsten fünf Tage abgefragt. Diese API returned ebenfalls ein JSON-Objekt, welches Vorhersagen aller drei Stunden, also 40 Werte, beinhaltet, aus denen die Temperatur und das Weathericon ausgelesen wird. In Bild 02 sieht man eine beispielhafte Antwort.

Quelle 3: Flight Data

Diese API ist von <https://rapidapi.com/Travelpayouts/api/flight-data>.

Recherche: Schwierigkeit von Computerspielen

Schwierigkeit von Computerspielen



**Gibt es bei Computerspielen
einen Zusammenhang zwischen
Schwierigkeitsgrad, Genre
und Alterseinstufung?**

HTWK Leipzig • Semantic Web • 2022



Postersession und Best Poster Awards

Wir freuen uns auf interessanten Diskussionen bei der anschließenden Posterpräsentation in der Wandelhalle.



**Voten Sie für das Best Poster in Semantic Web (grün/green)
and Design (rot/red)!**

16:00 eccenca Poster Award Ceremony

