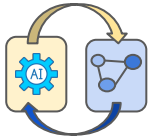
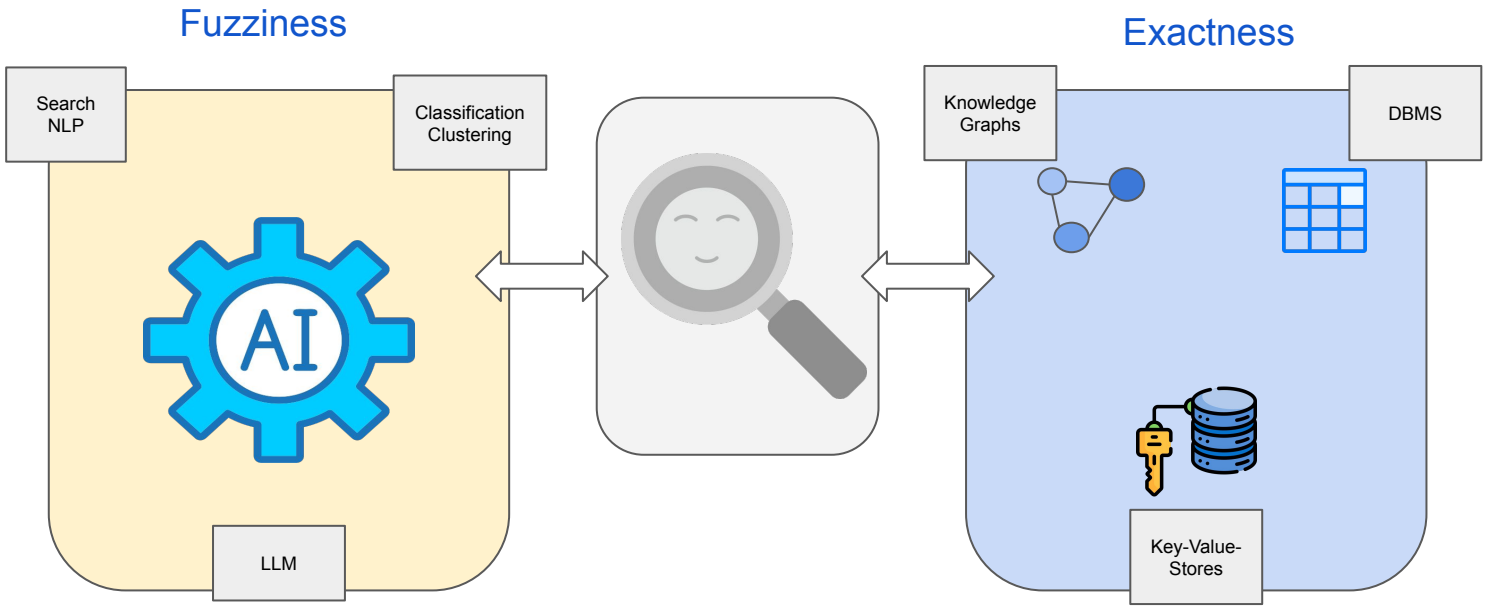
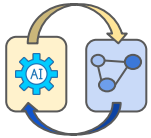
SPARQL*ing with AI

Lisa Wenige
Hochschule Merseburg
lisa.wenige@hs-merseburg.de



Information Search & Processing - Two Paradigms





Information Search & Processing - LLM and Knowledge Graphs

Large Language Models (LLM)

Cons
Implicit Knowledge
Hallucination
Indecisiveness
Black-box
Lacking Domain-specific/ new knowledge

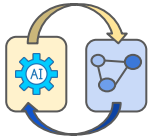
Pros
General Knowledge
Language Processing
Generalizability

Knowledge Graphs

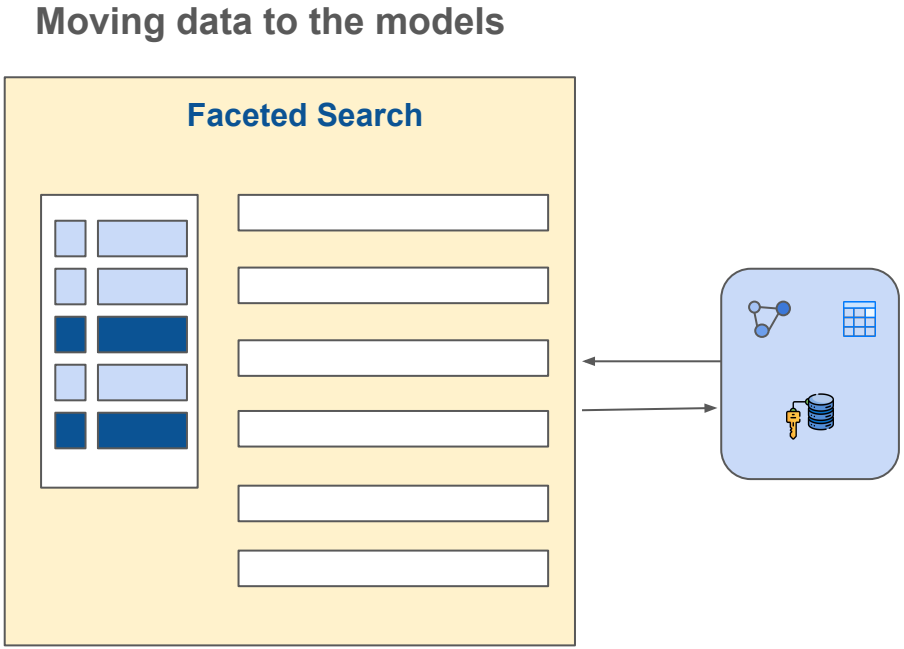
Pros
Structural Knowledge
Accuracy
Decisiveness
Interpretability
Domain-specific Knowledge
Evolving Knowledge

Cons
Incompleteness
Lacking Language Understanding
Unseen Facts

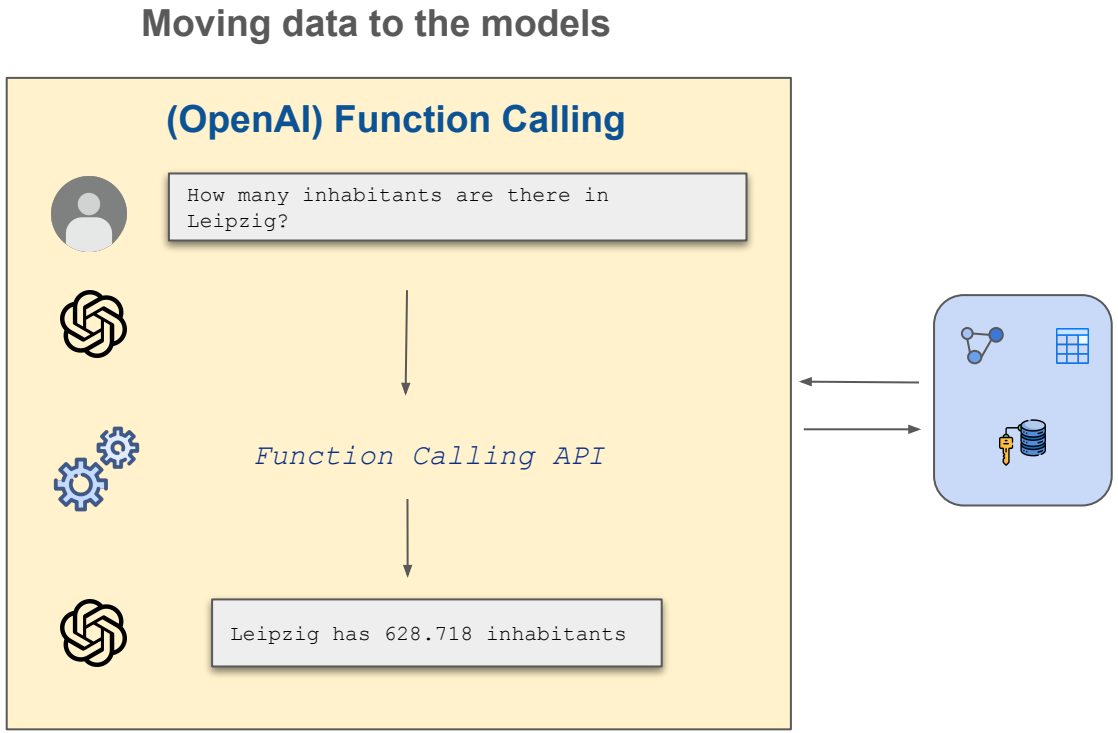
Source: Pan, Shirui, et al. "Unifying large language models and knowledge graphs: A roadmap."



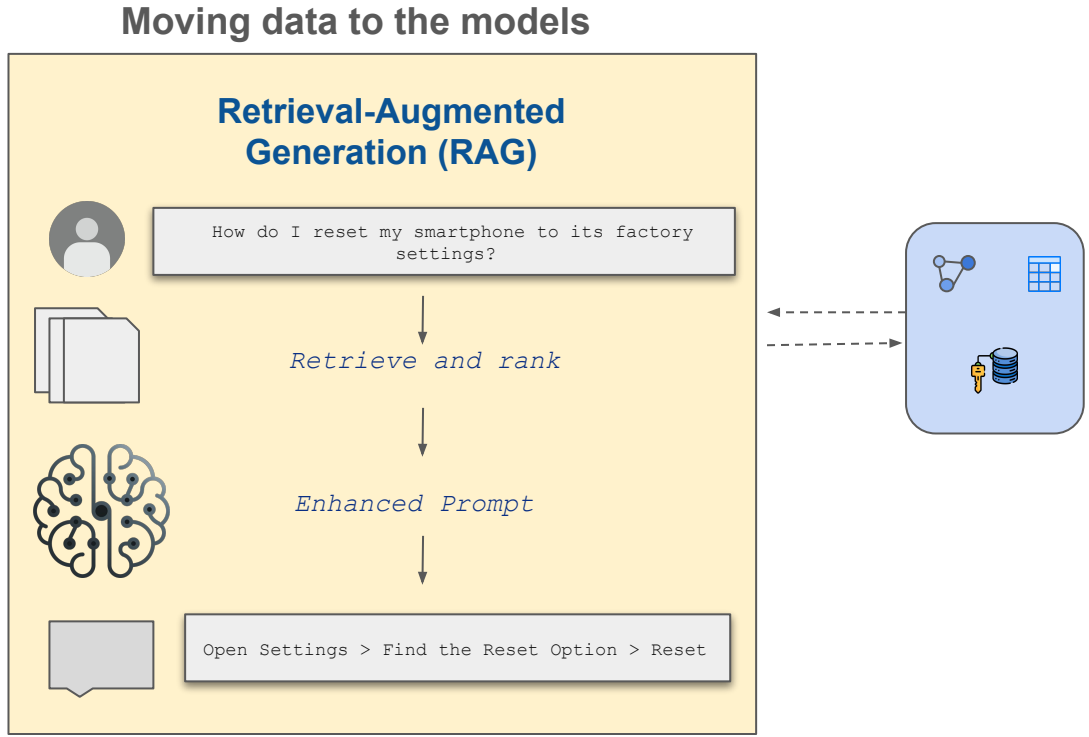
Information Search & Processing



Information Search & Processing

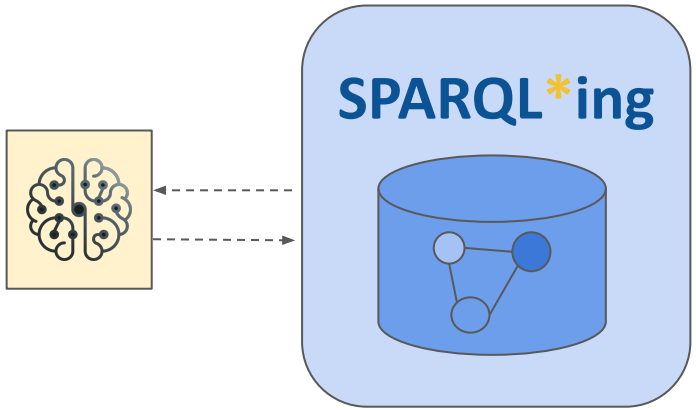


Information Search & Processing



Information Search & Processing

Moving models to the database



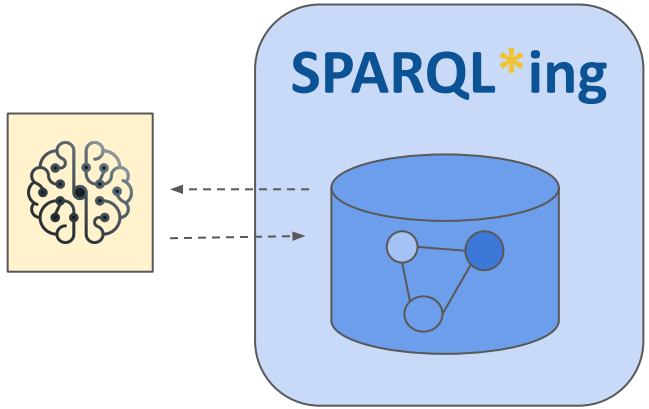
Information Search & Processing

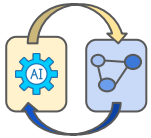
Moving models to the database: Advantages

Search & Retrieval
- Increase in recall
- Personalized filtering
- Bridging gaps in data (model) understanding

Machine Learning Operations (MLOps)
- Reproducibility
- Reusability
- Explainability

Workflows
- Autonomous process execution
- Efficiency (reducing prompt chaining)
- On-the-fly data integration & combination
- Complexity hiding





Related Work

Open Issues
- (Knowledge Graph) structured inputs
- Advanced workflows
- More functions for Knowledge Graph querying
- Complex return types

neo4j

CIPHER

```
CALL apoc.ml.openai.completion(
'What color is the sky? Answer in one word: ',
<API_KEY>,
{endpoint: <ENDPOINT_URL>, apiType:
<API_TYPE>, model: <MODEL>, path: ''})
```

https://neo4j.com/labs/apoc/5/ml/openai/

PostgresML

SQL

```
SELECT pgml.transform(
  task   => '{
    "task": "fill-mask",
    "model": "bert-base-uncased",
    "trust_remote_code": true
  }'::JSONB,
  inputs => ARRAY[
    'Paris is the [MASK] of France.'
  ]
);
```

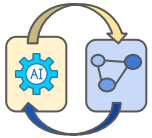
https://postgresml.org/

metaphactory

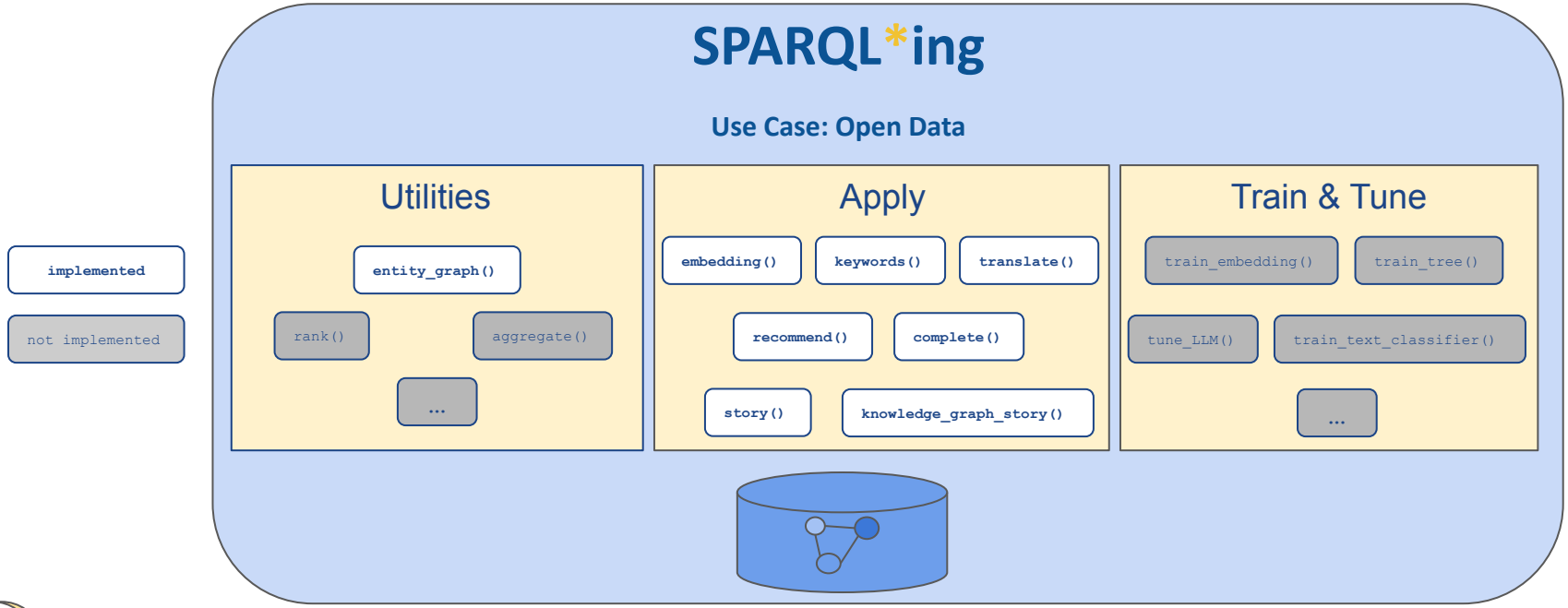
SPARQL

```
SELECT ?artist ?label WHERE
{ SERVICE mpfed:wikidataWord2Vec
          { wd:Q5598 mph:hasSimilar ?artist . }
  ?artist wdt:P106 wd:Q1028181 .
  ?artist rdfs:label ?label .
}
```

Haase, Peter, et al. "metaphactory: A platform for knowledge graph management." *Semantic Web* 10.6 (2019): 1109-1125.

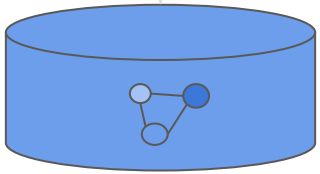


Family of AI-enhanced SPARQL Custom Functions



Function embedding()

```
SELECT DISTINCT ?dataset ?title  WHERE {
  BIND ('Verkehr' as ?keyword)
  ?dataset dct:title ?title;
           rdf:type dcat:Dataset .
  FILTER (CONTAINS(?title,?keyword))
}
```



DCAT Catalog

Table | Response | 2 results in 0.085 seconds

	dataset	title
1	<https://opendata.leipzig.de/dataset/3140cb43-0960-440f-b125-bc3ae6ac41d8>	Verkeh
2	<https://opendata.leipzig.de/dataset/9ffe4c39-5ad9-4765-a298-65b7adbfc4dd>	Verkeh

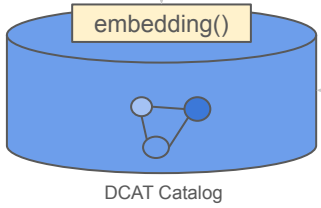
Showing 1 to 2 of 2 entries

Traffic Accidents (Quarterly Figures)

Traffic Accidents (Yearly Figures)

Fasttext-embeddings for query expansion

```
SELECT DISTINCT ?dataset ?title  WHERE {
  BIND (sparql-ai:embedding(
    "Verkehr",
    '{"model": "cc.de.300.bin","index": true, "top_n": 10}'
  ) as ?keyword)
  ?dataset dct:title ?title;
  rdf:type dcat:Dataset .
  FILTER (CONTAINS(?title,?keyword))
}
```



DCAT Catalog

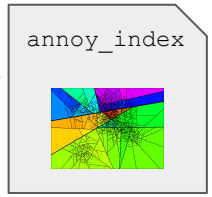


Table | Response | 7 results in 0.703 seconds

	dataset	title
1	<https://opendata.leipzig.de/dataset/23635389-0aaa-49a5-a52c-a0fc43b7f956>	Luftve
2	<https://opendata.leipzig.de/dataset/4b0eb8cb-a123-4d05-addd-e2401625bb2e>	Dauerz
3	<https://opendata.leipzig.de/dataset/73c5690c-07f2-41eb-abef-6a998c5b84d9>	Dauerz
4	<https://opendata.leipzig.de/dataset/d5c2baac-7e9f-471d-968a-12a15e8e8e5a>	Luftve
5	<https://opendata.leipzig.de/dataset/6c7a199f-e274-4942-84e0-a44c9d4ab230>	Person
6	<https://opendata.leipzig.de/dataset/3140cb43-0960-440f-b125-bc3ae6ac41d8>	Verkeh
7	<https://opendata.leipzig.de/dataset/9ffe4c39-5ad9-4765-a298-65b7adbfc4dd>	Verkeh

Showing 1 to 7 of 7 entries

Avian Traffic (Yearly Figures)

Permanent Bike Counter - Peak day last year

Permanent Bike Counter - No. of bikes yesterday

Public Transportation (Yearly Figures)

Traffic Accidents (Quarterly Figures)

Traffic Accidents (Yearly Figures)

Function recommend()

TFIDF similarities for recommendation

```
SELECT DISTINCT ?title ?recommended  WHERE {
  VALUES ?dataset {
    le:3140cb43-0960-440f-b125-bc3ae6ac41d8
    le:44543285-378f-40f2-aefc-50e0a27be2a0
    le:98579d91-49d0-4b2a-8c6c-87bcb52f9f29}
  ?dataset dct:title ?title
  BIND(sparql-ai:recommend(
    ?dataset,
    '{"model": "tfidf", "index": true, "top_n": 1}'
  ) as ?recommendation)
  ?recommendation dct:title ?recommended .
}
```

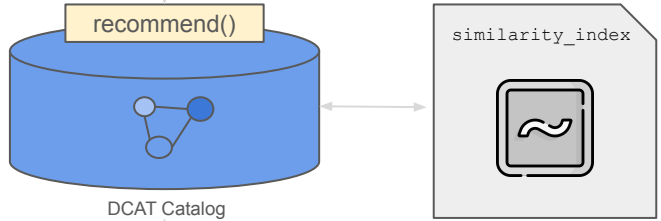


Table | Response | 3 results in 0.111 seconds

	title	recommended
1	Verkehrsunfälle (Quartalszahlen)	Verkehrsunfälle (Jahreszahlen)
2	Jahresstatistik LeipzigGiesst gegoss	Baumkataster, Stadt Leipzig
3	Durchschnittliche Anzahl obdachlos	Verweildauer in Notunterkünften

Traffic Accidents (Quarterly Figures)

Traffic Accidents (Yearly Figures)

Annual Statistics LeipzigGießt (Liters Watered)

Tree Registry, City of Leipzig)

Avg. No. of Homeless Individuals (Yearly Figures)

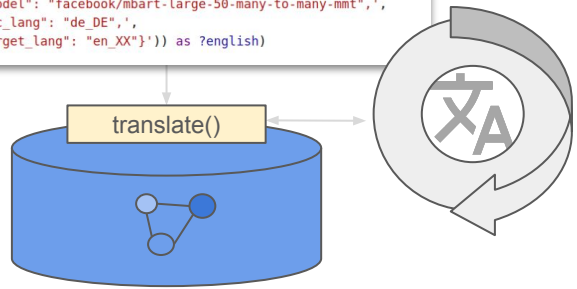
Stay duration in shelters (Yearly Figures)

Function translate()

Translation
with multilingual models

```
SELECT ?desc ?english ?ukrainian WHERE {
  ?dataset dct:publisher le-orga:31bbc58f-6e0f-44ac-a289-aebf488ba726;
           dct:description ?desc .

  BIND(sparql-ai:translate(
    ?desc,
    CONCAT(
        '{"model": "facebook/mbart-large-50-many-to-many-mmt",',
        '"src_lang": "de_DE",',
        '"target_lang": "uk_UA"}')) as ?ukrainian)
  BIND(sparql-ai:translate(
    ?desc,
    CONCAT(
        '{"model": "facebook/mbart-large-50-many-to-many-mmt",',
        '"src_lang": "de_DE",',
        '"target_lang": "en_XX"}')) as ?english)
}
```



	desc	english	ukrainian
1	Der Gutachterausschuss in der Stadt Leipzig h…	The Advisory Committee in the City of Leipzig …	На 02.03.2022 the Gutachterausschuss в місті Лей
2	Die öffentlichen Toiletten in der Stadt Leipzig.…	The public toilets in the city of Leipzig. The dat…	Громадські туалети в місті Лейпциг. Дані зробле
3	Der Gutachterausschuss in der Stadt Leipzig h…	The Advisory Committee in the City of Leipzig …	На 02.03.2023 the Gutachterausschuss в місті Лей
4	Die Digitale Stadtgrundkarte im Maßstab 1:1.…	The digital city map in the scale 1:1.000 serves …	Цифрова картка в масштабі 1:1.000 serves як ос
5	Die Digitale Stadtkarte 1:25000 beinhaltet die …	The digital city map 1:25000 covers the topics o…	Цифрова мапа 1:25000 - це тема дорожної і Schi
6	Die Digitale Stadtkarte im Maßstab 1:5.000 di…	The digital city map in scale 1:5.000 serves as a…	Цифрова карта на рівні 1:5.000 serves як основа

Function keywords()

Pipeline with embeddings and keywords

```
SELECT DISTINCT ?title ?keywords ?enhanced_keyword WHERE {
  le:30943d88-4ac9-4968-84bc-35e9b337a85d dct:title ?title.
  BIND(sparql-ai:keywords(
    ?title,
    '{"model": "bert-base-german-cased","stopwords": "german", "max_ngram": 2}'
  ) as ?keywords)
  BIND(sparql-ai:embedding(
    ?keywords,
    '{"model": "cc.de.300.bin","index": true, "top_n": 3}'
  ) as ?enhanced_keyword)
}
ORDER BY ASC(?enhanced_keyword)
```

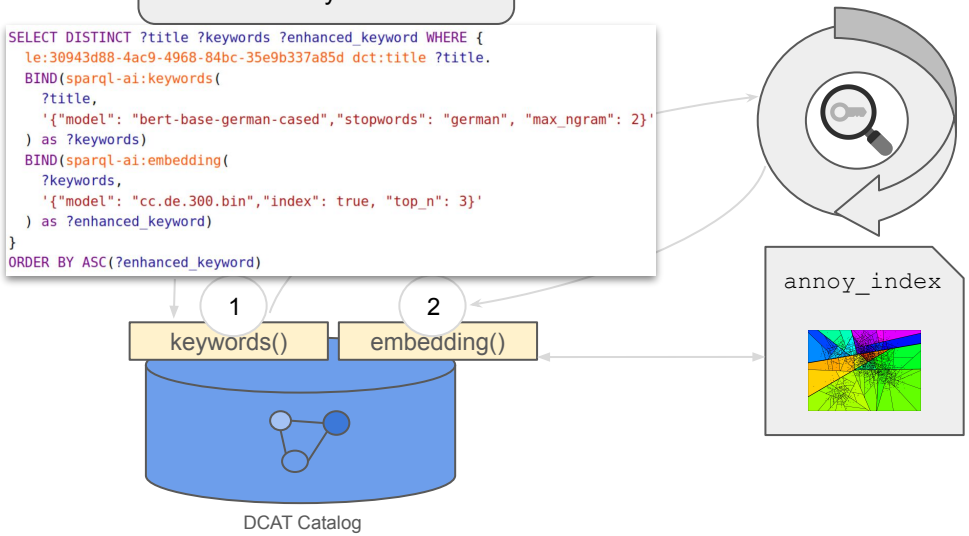


Table | Response | 9 results in 0.207 seconds

	title (Unemployed (Annual figures small-scale))	keywords ([Annual figures small-scale, unemployed annual figures, small-scale, annual figures, unemployed])	enhanced_keyword
1	Arbeitslose (Jahreszahlen, kleinräumig)	['jahreszahlen kleinräumig', 'arbeitslose jahreszahlen', 'kleinräumig', 'jahreszahlen', 'arbeitslose']	arbeitssuchende
2	Arbeitslose (Jahreszahlen, kleinräumig)	['jahreszahlen kleinräumig', 'arbeitslose jahreszahlen', 'kleinräumig', 'jahreszahlen', 'arbeitslose']	arbeitsuchende
3	Arbeitslose (Jahreszahlen, kleinräumig)	['jahreszahlen kleinräumig', 'arbeitslose jahreszahlen', 'kleinräumig', 'jahreszahlen', 'arbeitslose']	beschäftigungslo

job seekers
jobseekers
unemployed persons

Function complete()

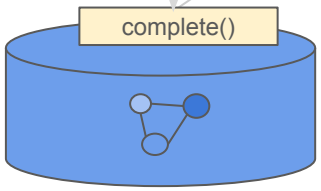
Few-shot prompting to generate dataset descriptions

```
SELECT (sparql-ai:complete(
            ?few_shots,
            "Zu- und Wegzüge syrische Bevölkerung 2023",
            CONCAT(
            '{"model": "gpt-4-0613",',
            '"prompt_start": "Gegeben sind folgende Beispiel-Daten in der Struktur Input, Output ",',
            '"prompt_end": "Generiere eine dct:description für den Titel ",',
            '"temperature": 0.8,',
            '"max_tokens": 200}'
            )
        ) as ?completed_description)
        ?few_shots
WHERE {
    SELECT (GROUP_CONCAT(CONCAT(?title, ', ', ?description); separator="\n") as ?few_shots)
    WHERE {
        ?dataset dct:title ?title;
                 dct:description ?description;
                 dcat:keyword "Bevölkerung" .
    }
}
```

```
prompt_start: Example
data with structure

Input, output
title_1, description_1
title_2, description_2

prompt_end: Generate a
description for the
title Arrivals and
departures of Syrian
individuals 2023
```



DCAT Catalog

completed_description

This data represents the number and trends of arrivals and departures of Syrian individuals in Leipzig in 2023. The information is based on residents' registration office data and offers insights into the mobility [...] of the Syrian population group in the city…

	completed_description	few_shots
1	"Diese Daten stellen die Anzahl und Trends der Zu- und Wegzüge syrischer Personen in Leipzig im Jahr 2023 dar. Die Informationen basieren auf Einwohnermelderegisterdaten und bieten Einblicke in die Mobilität und Veränderungen innerhalb der syrischen Bevölkerungsgruppe der Stadt. Es wird auch die Differenz zwischen prognostizierten und tatsächlichen Bewegungen aufgezeigt."	Prognostizierte und tatsächliche Altersstruktur, Entwicklung der tatsächlichen Altersstruktur 2018 und 2022; Vergleich prognostizierte und tatsächliche Altersstruktur 2022 Lebenserwartung, Angaben zur durchschnittlichen Lebenserwartung im gesamten Stadtgebiet sowie in einzelnen Clustergebieten. Bevölkerungsvorausschätzung und tatsächliche Einwohnerentwicklung, Bevölkerungsvorausschätzung 2019 (Hauptvariante) und tatsächliche Einwohnerentwicklung Bestand ukrainischer Einwohnerinnen und Einwohner in Leipzig am 31.12.2022 nach Alter, Bestand ukrainischer Einwohnerinnen und Einwohner in Leipzig am 31.12.2022 nach Alter Relative Entwicklung der Einwohnerzahl nach Ortsteilen, Relative Entwicklung der Einwohnerzahl nach Ortsteilen, 2018 bis 2022 Innerstädtische Umzüge ukrainischer Personen innerhalb der Stadt Leipzig 2022, Innerstädtische Umzüge ukrainischer Personen innerhalb der Stadt Leipzig 2022 Anteil von Kindern mit Migrationshintergrund in Kindertageseinrichtungen, Anteil von Kindern mit Migrationshintergrund unter Kinderkrippenkindern und Kindergartenkindern in Leipziger Kindertageseinrichtungen in öffentlicher und freier Trägerschaft sowie in der Leipziger Bevölkerung im Alter von 1 - unter 3 Jahren und im Alter von 3 - unter 7 Jahren zwischen 2012 und 2022 Kinderkrippen- und Kindergartenplätze in Kindertageseinrichtungen, Anzahl der Kinderkrippen- und Kindergartenplätze in Leipziger Kindertageseinrichtungen in öffentlicher oder freier Trägerschaft zwischen 2012 und 2022 und Anzahl der Kinder im entsprechenden Alter Relative Bevölkerungsentwicklung - Kinder und Jugendliche, Relative Entwicklung der Einwohnerzahlen 2018 bis 2022 für folgende Altersgruppen: 0 bis unter 6 Jahre, 6 bis unter 10 Jahre und 10 bis unter 19 Jahre Prognostizierte und tatsächliche Bevölkerungsentwicklung, Prognostizierte und tatsächliche Bevölkerungsentwicklung Wanderungssaldo nach Herkunfts- und Zielgebieten, Wanderungssaldo Leipzigs nach Herkunfts- und Zielgebieten 2018 bis 2022 Bei den hier dargestellten Daten zu Zuzügen und Wegzügen handelt es sich um revidierte Angaben, die von bisher veröffentlichten Angaben abweichen. Aufgrund einer signifikanten Zunahme von Korrekturbuchungen und Änderungen von Amts wegen im Leipziger Einwohnermelderegister in den zurückliegenden Jahren, insbesondere seit 2015, sind die betreffenden Zu- und Wegzüge in die hier dargestellten Daten neu eingerechnet worden. Bisher waren diese Vorgänge nicht mit ausgewertet worden. Die Meldebehörde nimmt Änderungen von Amts wegen unter anderem vor, wenn Personen in Erstaufnahmeeinrichtungen des Freistaates Sachsen in Leipzig aufgenommen wurden oder wenn bekannt wird, dass Personen ins Ausland verzogen sind. Die Daten wurden rückwirkend seit 2015 bereinigt. Die [bisher veröffentlichten, unrevidierten Angaben](https://statistik.leipzig.de/statcity/table.aspx?cat=3&rub=6) sind bis auf weiteres abrufbar. Registrierungen ukrainischer Personen in Leipzig im Jahr 2022, Registrierungen ukrainischer Personen in Leipzig im Jahr 2022 Bevölkerungsvorausschätzung 2023, Schätzung der Einwohnerzahlen für die Stadt Leipzig bis zum Jahr 2040 ausgehend von der Bevölkerungsstatistik 2023. Zu- und Wegzüge ukrainischer Personen nach/von Leipzig im Jahr 2022, Zu- und Wegzüge ukrainischer Personen nach/von Leipzig im Jahr 2022 Bestand ukrainischer Einwohnerinnen und Einwohner in Leipzig am 31.12.2022 nach Ortsteilen, Bestand ukrainischer Einwohnerinnen und Einwohner in Leipzig am 31.12.2022 nach Ortsteilen

Function story()

Opening Hours Special Markets

`if dct:publisher == "Market Office"`

Retrieve opening hours for the next special market

First Name Statistics 2023

`if dct:publisher == "Market Office"`

Determine the top rankings for first names

CKAN Datastore API

```
role: I am a PR officer. I write news that go viral.
prompt: Write a twitter post for the following data
```

```
SELECT ?title ?publisher ?soci
    VALUES ?dataset {le:ddba07
    ?dataset dcat:distribution
        dcat:contactPoint ?con
        dct:title ?title .
    ?contactPoint a vcard:Orga
        vcard:fn ?publisher .
    ?distribution dcat:mediaType "text/csv" ;
    BIND(sparql-ai:story(
        ?distribution,
        ?title,
        ?publisher,
        CONCAT(
            '{"model": "gpt-4-0613",',
            '"role": "Ich bin PR-Mitarbeiter*in. Ich schreibe Nachrichten, die viral gehen.",',
            '"prompt": "Schreibe einen Twitter-Post zum Datensatz",',
            '"temperature": 0.75,',
            '"max_tokens": 280}'
        )
    ) as ?social_media_post)
}
```

story()

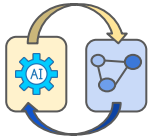
DCAT Catalog

Table | Response | 2 results in 14.071 seconds

	title	publisher	social_media_post
1	Leipziger Spezialmärkte - Öffnungszeiten und Termine	Marktamt	"🍷 Liebe Leipziger kommt mit tollen We men auf die guten Zeiten anstoßen! 🥂 #Leipzig #Weinfest #Markt"
2	Vornamenstatistik 2023	Standesamt	📊👶 Die Vornamen Bei den Mädchennan eweils 34. Welcher N

Dear Leipzig wine lovers, mark your calendars for May 15th to May 26th! The #LeipzigWineFestival is coming to the market with great wines and good vibes 🎉. Opening hours are from 12:00 PM to 11:00 PM. Let's toast together to the good times! 🥂 #Leipzig #WineFestival #Market

📊👶 The 2023 first name statistics are here! Emil and Noah are the frontrunners for boys' names with 50 and 48 mentions! For girls' names, Charlotte and Emilia share the top spot with 40 mentions, closely followed by Mila and Ella with 34 each. Which name is your favorite? #BabyNames2023 🍼🎉

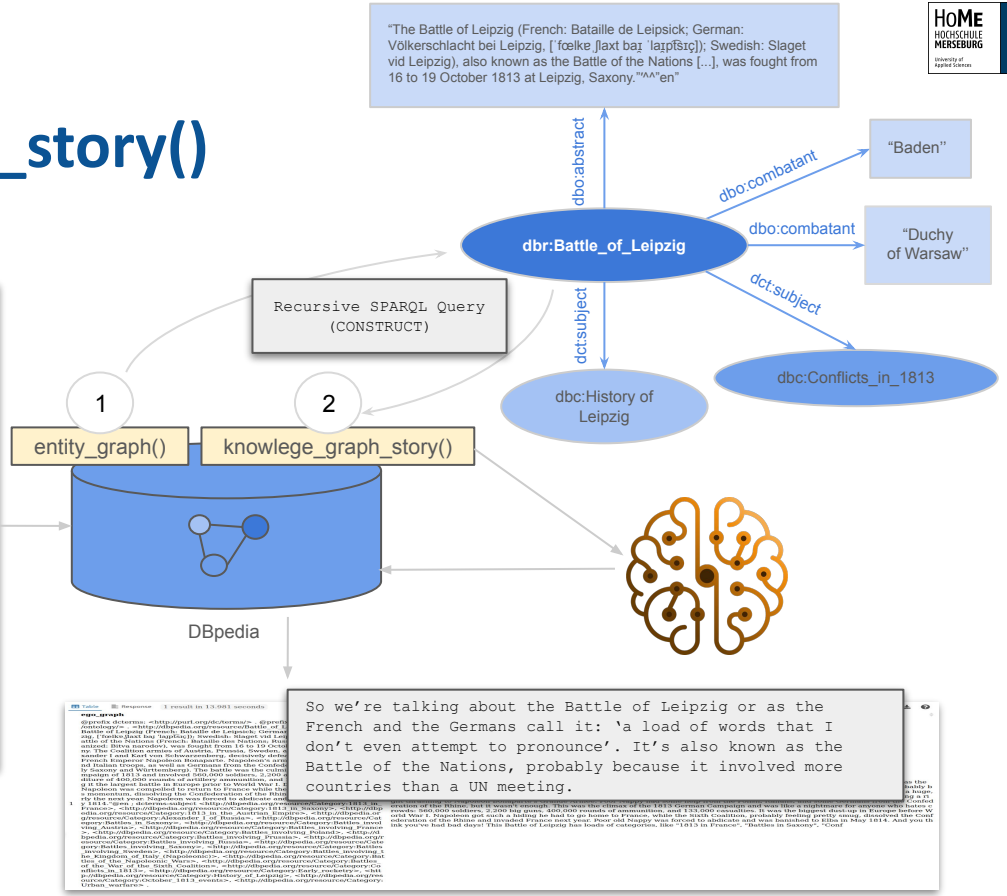


Writing mockumentary-style stories with knowledge graphs

Function knowledge_graph_story()

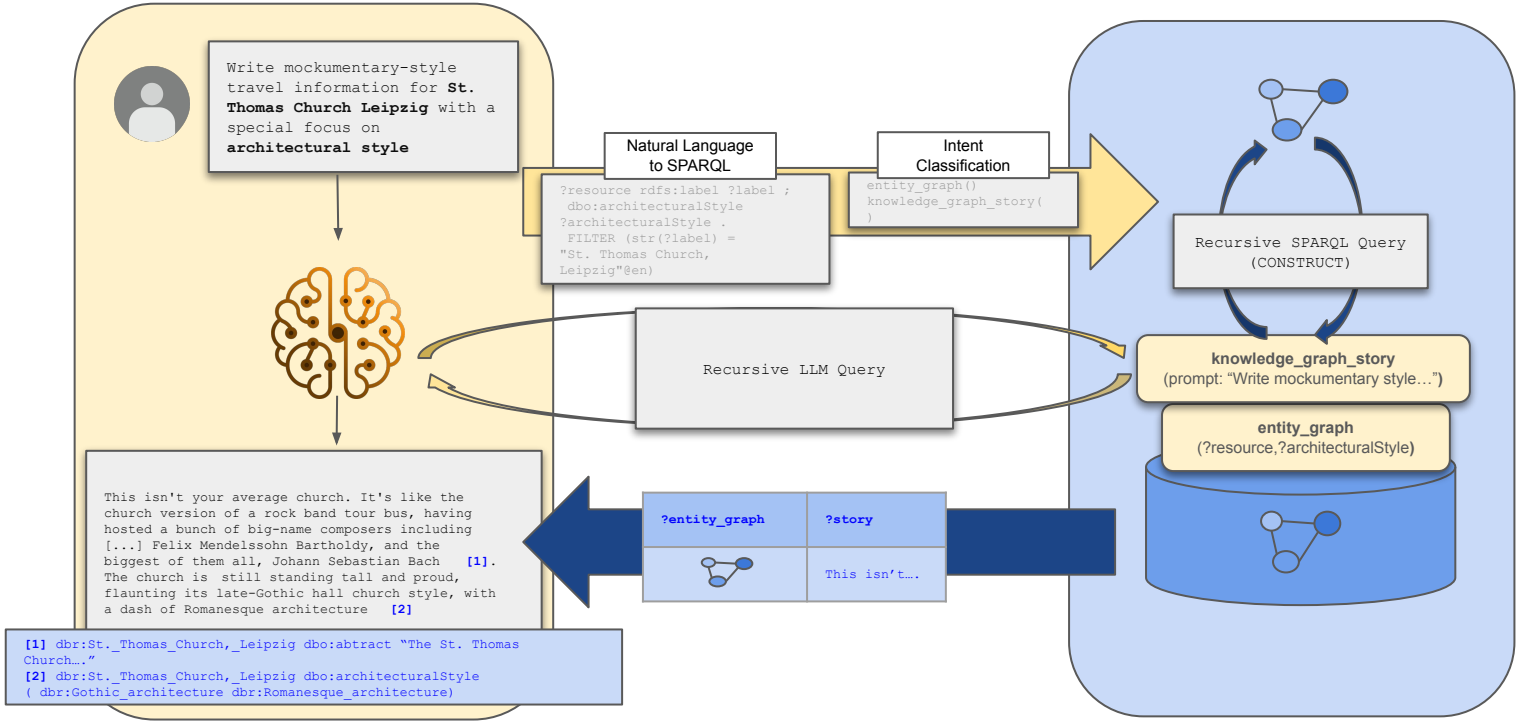
```
SELECT ?entity_graph ?story
WHERE {
  BIND(sparql-ai:entity_graph(
    <http://dbpedia.org/resource/Battle_of_Leipzig>,
    dbo:abstract,
    dbo:combatant,
    dcterms:subject,
    CONCAT(
        '{"outgoing": false,',
        '"lang": "en"}'
    )
  ) as ?entity_graph)

 BIND(sparql-ai:knowledge_graph_story(
    ?entity_graph,
    CONCAT(
        '{"model": "gpt-4-0613",',
        '"role": "I am Philomena Cunk. I give funny mockumentary-style travel information.",',
        '"prompt": "Write a cunky expanation for this data",',
        '"temperature": 0.65,',
        '"max_tokens": 350}'
    )
  ) as ?story)
}
```



Application Example - GuestBot

Vision - Full-circle Integration



Thank you for your attention

Questions?