Hybrid Enterprise Knowledge Graphs

Peter Haase
ISWC 2019, Industry Track
28.10.2019, Auckland, New Zealand
COMPANY FACTS

- metaphacts GmbH
- Founded in Q4 2014
- Headquartered in Walldorf, Germany
- Currently ~20 people

- Independent software vendor
- Privately-held, owner-managed company
- Platform for knowledge graphs and knowledge graph applications
Hybrid Enterprise Knowledge Graphs

Knowledge graph as integration hub
"Master data” about business-relevant entities
Interlinked and federated with:

Variety of data sources
- Local/remote RDF stores
- Virtual graphs over e.g. relational
- Enterprise APIs
- Data streams, sensor data

Variety of data modalities
- Text
- Temporal, geospatial
- Images
- Custom indices

Variety of data processing techniques
- Graph analytics
- Machine learning, statistics
- Domain-specific, e.g. genome sequence similarity

Graph database
Relational Database
Specialized indices
Machine learning model
Custom API
Hybrid Enterprise Knowledge Graphs in metaphactory

Ephedra Federation Engine: Main Principles

- Extends RDF4J API
- Data sources and compute services are wrapped into “virtual” RDF4J repositories
  - Graph patterns are transformed into API calls
  - SPARQL 1.1 federation using the SERVICE keyword
- Service wrapper repositories are explicitly described in the service registry
  - Explicit mappings between graph patterns and service input/output parameters
  - Built-in wrappers for relational, REST API
  - Easy SDK for implementing own Service wrappers
- Static and runtime optimizations for hybrid queries
• Explicit description of resources in the knowledge graph

• Enrichment of the graph with an implicit model, often a learned model using machine learning

• Typical example: Similarity of products, items, ...

• Range of different models
  • Graph embeddings, word embeddings
  • Integration of model via API
  • Can be computed at runtime or pre-trained

Use Case: Similarity Search & Machine Learning

```sparql
# Select a company similar to BMW
SELECT ?company WHERE {
  SERVICE ephedra:word2vec {
    :BMW word2vec:isSimilar ?company .
  }
  ?company :headquarters ?location
}
```

```
word2vec
embeddings

:BMW
in: URI
out: URI[]

?company = :Audi
```
Use Case: Sensor Data

- Knowledge graphs of smart objects, example: trains
- Sensors make tempo-spatial observations
- Interpretation of sensor and image data requires
  - Time and location
  - Weather conditions
- Federation service to include weather data via Dark Sky API
- Contextualization of sensor data with weather information

```
SELECT ?weather WHERE {
  ?observation geo:lat ?latitude; geo:long ?longitude; :time ?time
  SERVICE ephedra:weatherapi {
    ?results weatherapi:latitude ?latitude;
    weatherapi:longitude ?longitude;
    weatherapi:time ?time;
    weatherapi:summary ?weather.
  }
}
```
• Life science knowledge graphs e.g. in drug research: compounds, proteins, ...

• Integration of knowledge graph with special purpose databases operating on the chemical structures

• Chemical Structure Search API (REST) integrated via Service wrapper

• Example: finding exact, similar, or sub-structures

SELECT ?substance ?similarity ?id ?inchi WHERE {
  SERVICE ephedra:chemsimilarity {
    ?results chem:hasSMILES ?smilesCode .
    ?results chem:hasSimilarityThreshold ?similarityThreshold .
    ?results chem:hasMoleculeChEMBLID ?id .
    ?results chem:hasSimilarity ?similarity .
    ?results chem:hasStandardInChIKey ?inchi .
  }
  ?substance :chemblID ?id
}
Demo: Wikidata & Federated Sources

Poster & Demo Session
Today at 18:00

https://wikidata.metaphacts.com/
• Hybrid enterprise knowledge graphs
  • Core knowledge graph in RDF as integration hub
  • Integration with enterprise sources (databases, APIs)
  • Semi-/Unstructured: text, images, geo, ...
  • Compute services
  • Machine learning models

• Ephedra: Federation over (virtual) SPARQL endpoints and compute services
  • Explicit mappings between SPARQL nodes and service input/output parameters
  • Static and runtime optimizations for hybrid queries
  • Using SPARQL 1.1 federation

Future work
• SPARQL 1.2 federation
  • Community working group
  • Generalization of SERVICE for non-SPARQL endpoints

• Integration with FedX federation framework
  • Recent integration into RDF4J
  • Enabling transparent federation
  • Improved optimization techniques
metaphacts GmbH
Daimlerstraße 36
69190 Walldorf
Germany
p +49 6227 6989965
m +49 157 50152441
e info@metaphacts.com
@metaphacts
# Select a company similar to BMW

```
SELECT ?company WHERE {
  SERVICE ephedra:word2vec {
    BMW word2vec:isSimilar ?company .
  };
  ?company :headquarters ?location
}
```