



ontotext
THE GRAPHDB COMPANY

Knowledge Graph Embeddings in the Industry

Delivering the Future – Together!

Knowledge Graph Forum – PANTOPIX

About the Partner



With our **expertise**, we support our customers in **optimizing their information processes** and ensure the **maximum added value** of their digital data.

We accompany our customers in the **analysis of existing processes**, in the **development of new concepts** and in the **introduction and integration** of new systems, methods and tools.



About the Speaker



Nikhil Acharya
Knowledge Engineer
at PANTOPIX

” Having been associated in the Knowledge Graph Industry for 4 years ,I believe this Industry promises a strong future in the era of Large Language Models. I wish to bring the best ideas to the Industry and help solve Industrial problems.



What we're going to cover and why

1

Use of the Representation of Technical Data in the form of Semantic Relations

2

Methods currently used to analyse Semantic Relations

3

Usage of these methods to solve industrial problems

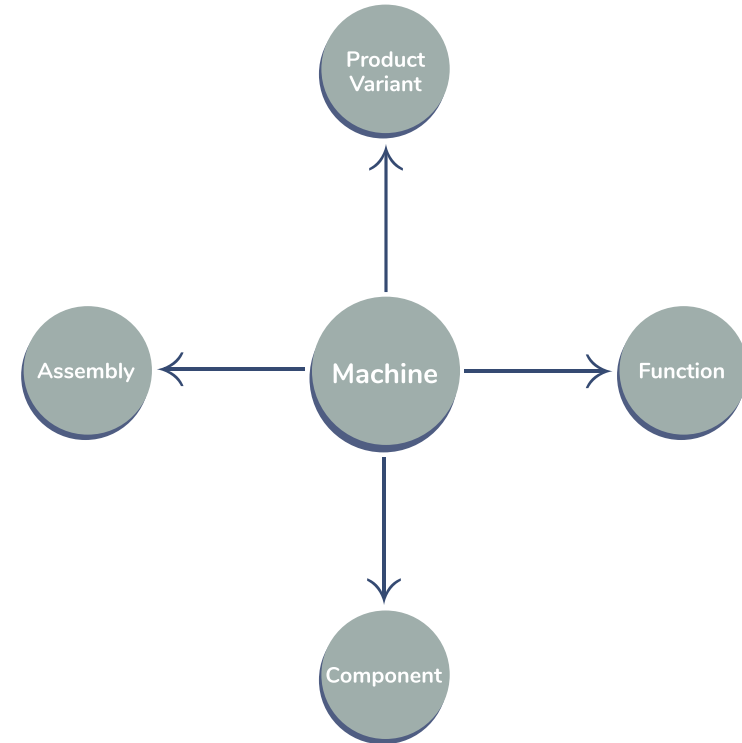
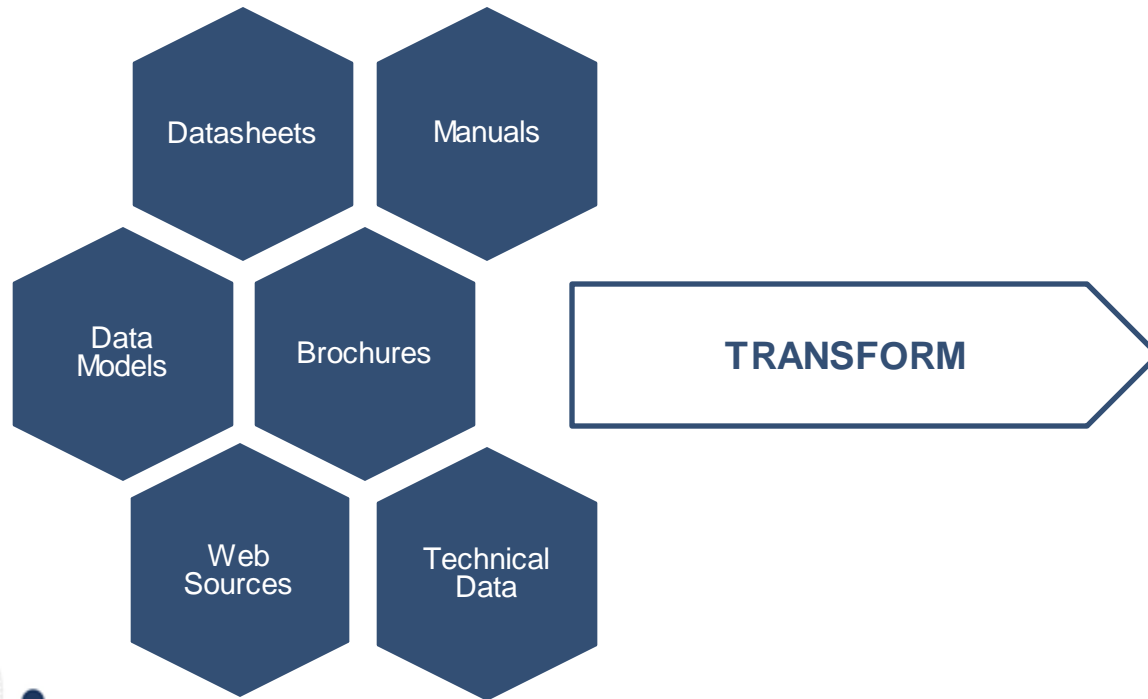


Representing and Analyzing Technical Data

- Technical Data from heterogenous sources should be represented semantically to facilitate unification, integrations, analysis and retrieval
- The retrieved data faces various challenges like incompleteness and correctness
- Meaningful representation of technical data enriched with Knowledge Graph Embeddings helps overcome these challenges



Representing Technical Data



Use Properties of Knowledge Graphs to enrich data. Represent machine properties like mass, dimensions etc in the Knowledge Graph.

Potentials of KG Embeddings

USED EQUIPMENT DATA SHEET

489

EQUIPMENT TYPE

Woodwaste Burning Industrial Hot Air Generator

MAKE

Fagida

MODEL

GX25D

SERIAL NO

NPL489/ CH-4905

YEAR MF

2012

VOLTAGE

220V Single phase

DIMENSIONS MAX 900mm (35.5") W x 1700mm (67") D x 1830mm (71.75") H

WEIGHT APPROX 418 Kg

SPECIFICATION

Woodwaste burning space heater

Generates 116,000 BTU/hr 34 kW/hr

200mm Flue outlet diameter

Double safety loading door

Loading door access 350mm x 525mm

Thermostatically controlled

Fan blows clean hot air generated by fire bricklined heat chamber to surrounding area

Lower ash grate for easy cleaning

Please note: This machine does not comply with smoke free zones. Check with your local authority for compliance in your area.

EQUIPMENT CONDITION New un-used showroom demonstrator

WARRANTY 12 Months See Used Equipment Terms Of sale

CHALLENGES

- Incomplete information (Year field)
- Faulty QA component
- Poor information visualization

-> Embeddings driven approach

POTENTIALS

- Stronger knowledge base
- Complete information
- Great visualization service
- Validated QA component

Example Datasheet

<https://www.yumpu.com/en/document/read/42827500/used-equipment-data-sheet-napoleon-products-limited>

Datasheet Information Retrieval

- OCR and LLM driven data extraction from datasheets
- XML mapping to Knowledge Graphs

Growth:	Adherent
Organism:	Homo sapiens (human)
Morphology:	Fibroblastoid

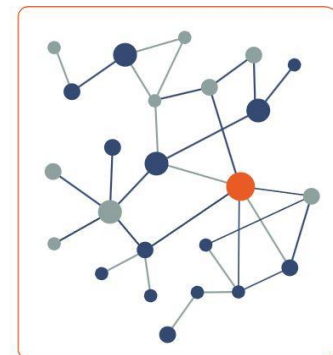
Comments:	ASC/TERT1 was developed from human adipose tissue-derived mesenchymal stem cells by transduction with a retroviral expression vector (pLXSN) containing the hTERT gene. The cell line was continuously cultured for more than 70 population doublings without showing signs of growth
-----------	---

Structured Technical Data

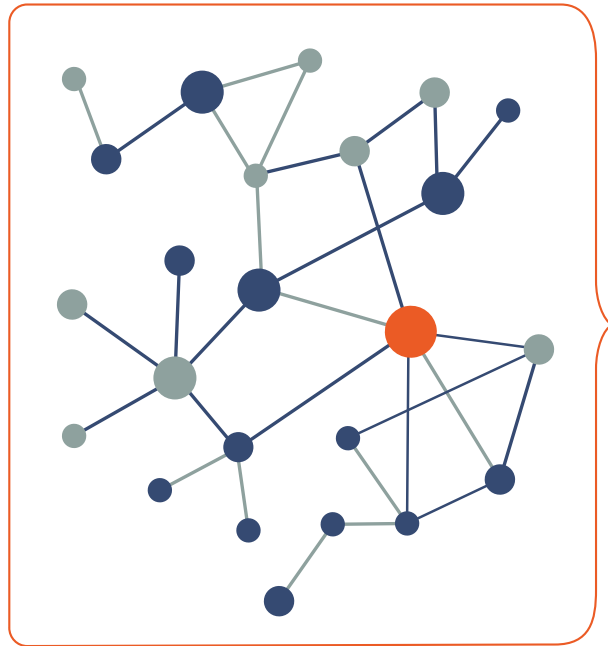
Unstructured Technical Data

Pretrained OCR and Language Models

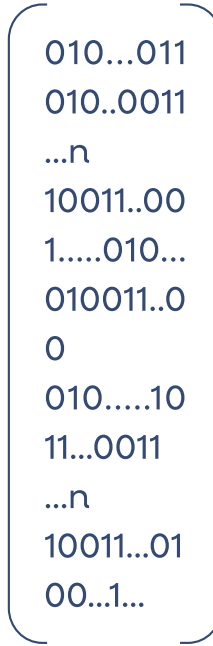
XML Mapping



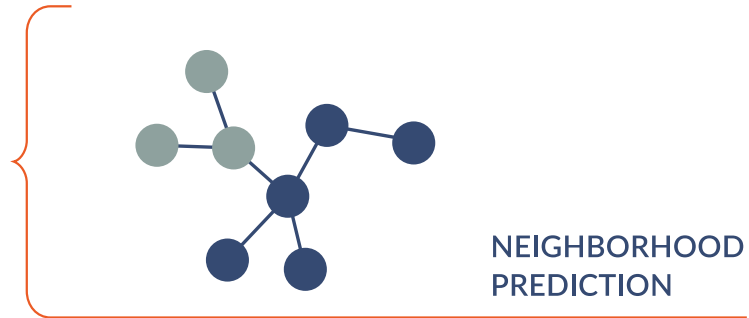
Knowledge Graph Embeddings



KNOWLEDGE GRAPH



KG EMBEDDINGS

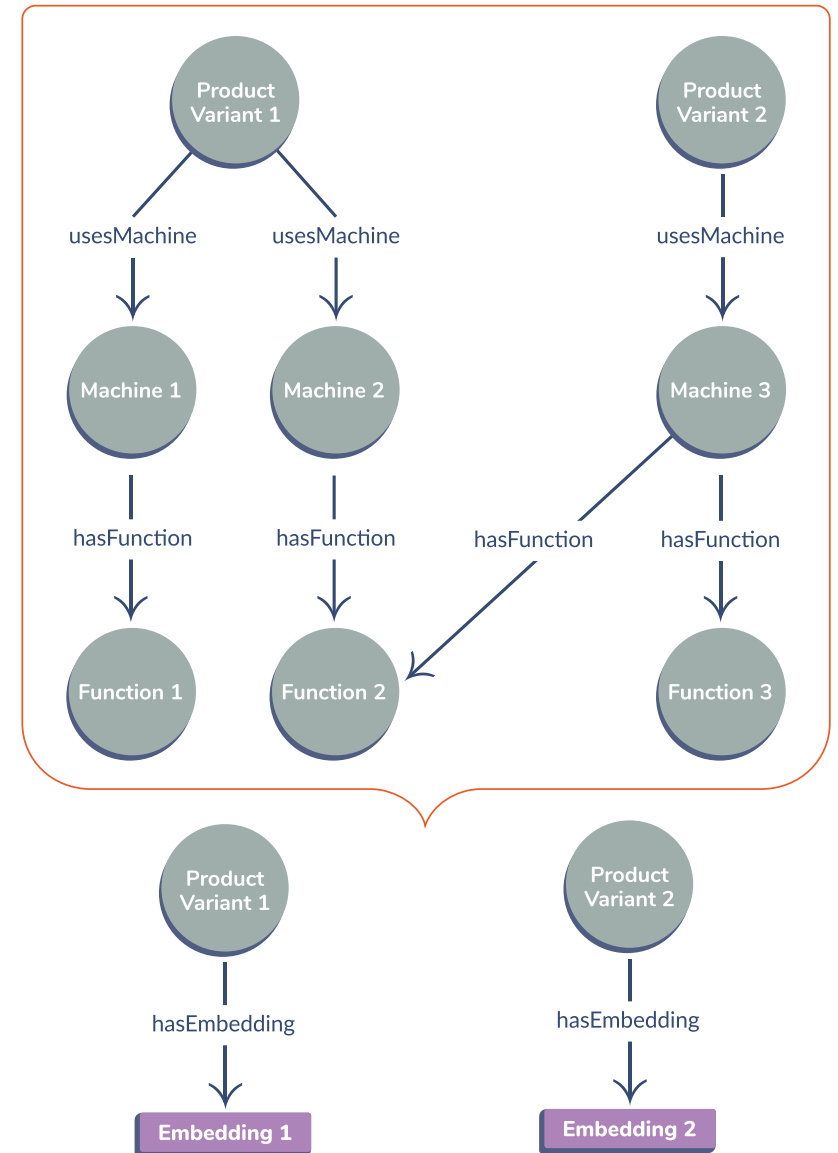


Information Analysis

- How do we compare two product variants?
- Can we represent the product variants numerically?

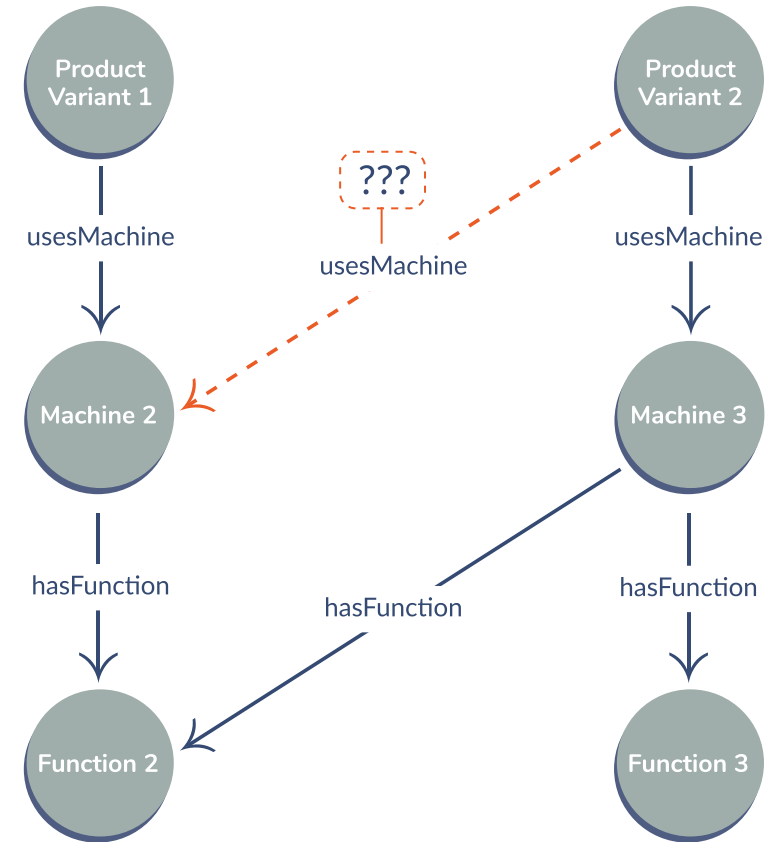
Yes!! Each product variant can be represented via an embedding.

(Hierarchy and ancestry is conserved.)

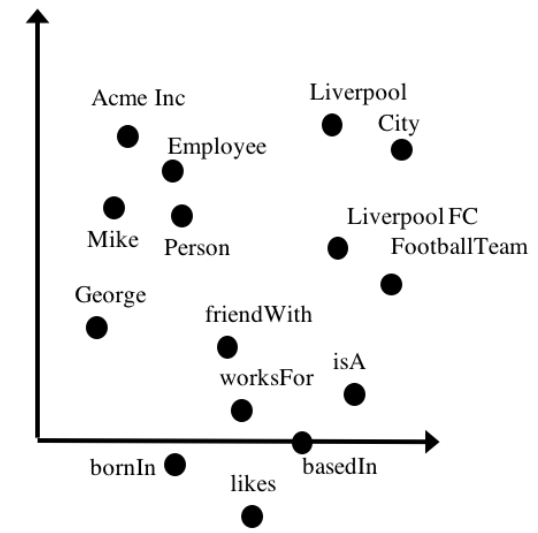
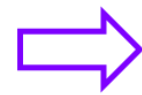
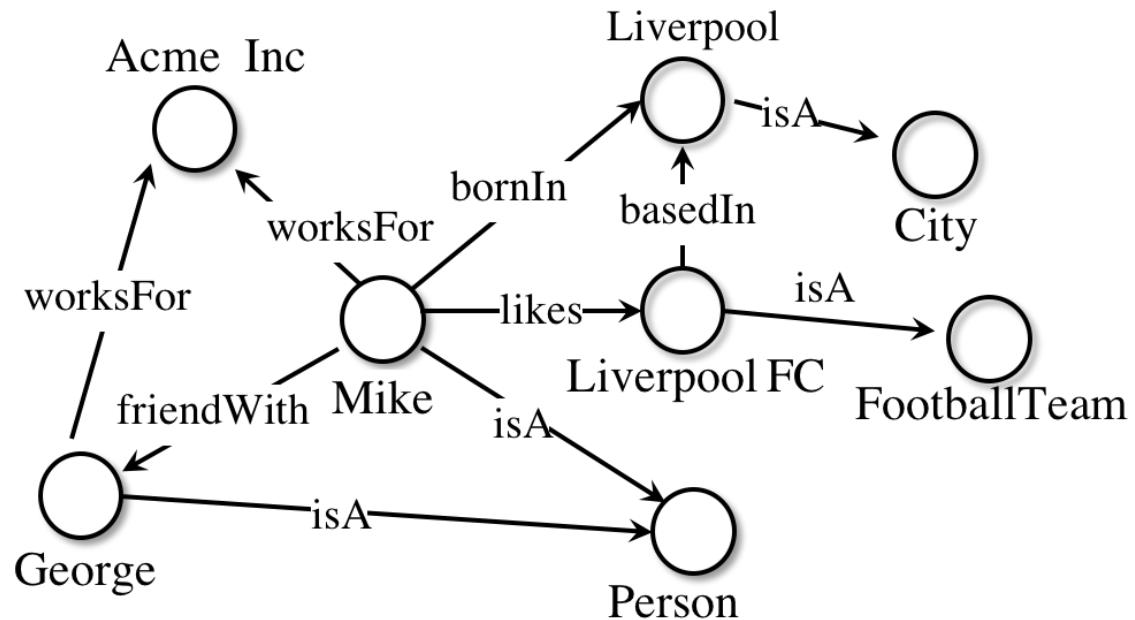


Information Enhancement

- What if we missed adding more information during Information Retrieval?
- Can we infer a link from product variant 2 to machine 2?
- Embeddings help us complete missing information from technical documents



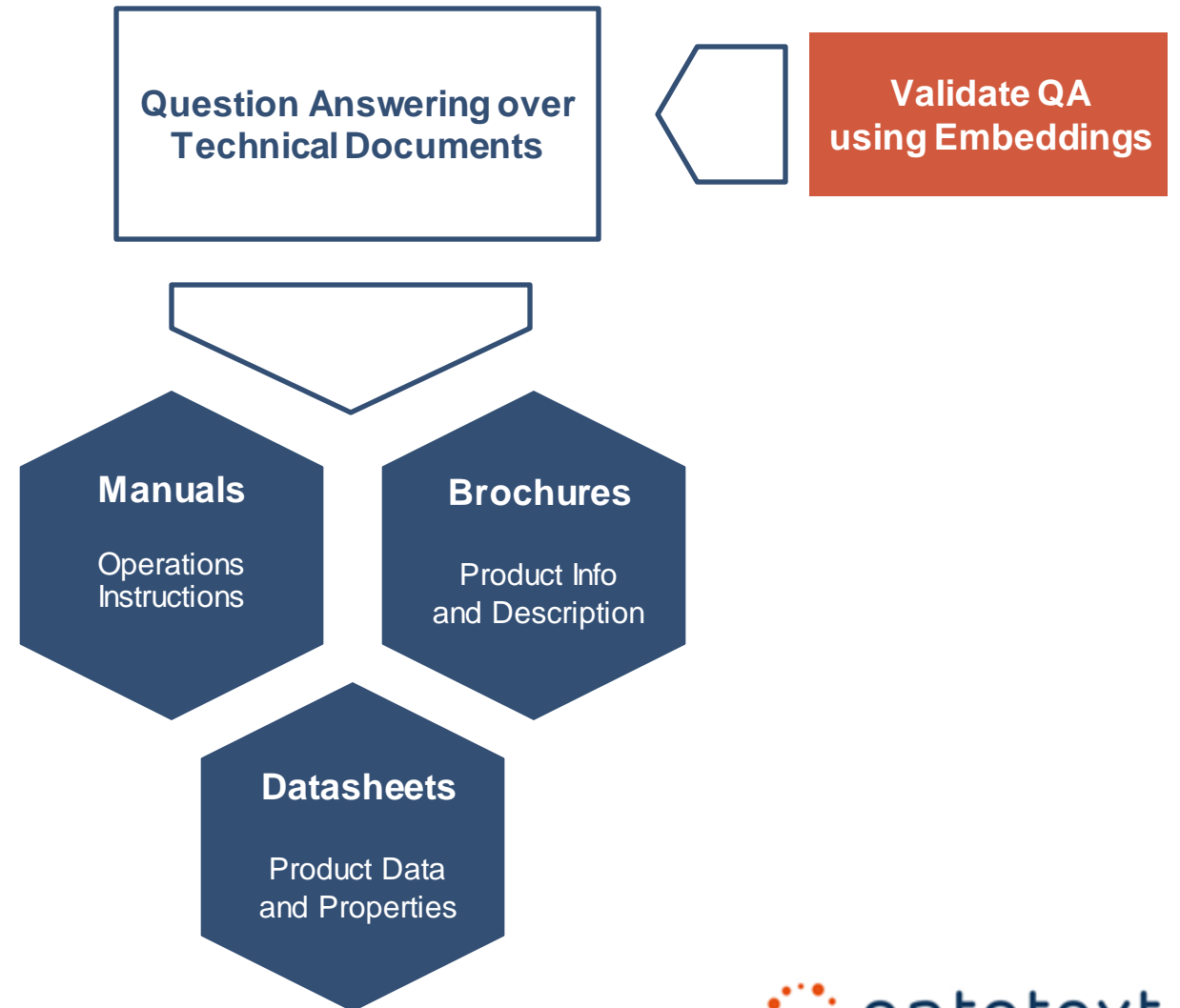
Information Visualization



Source: <https://docs.ampligraph.org/en/1.1.0/>

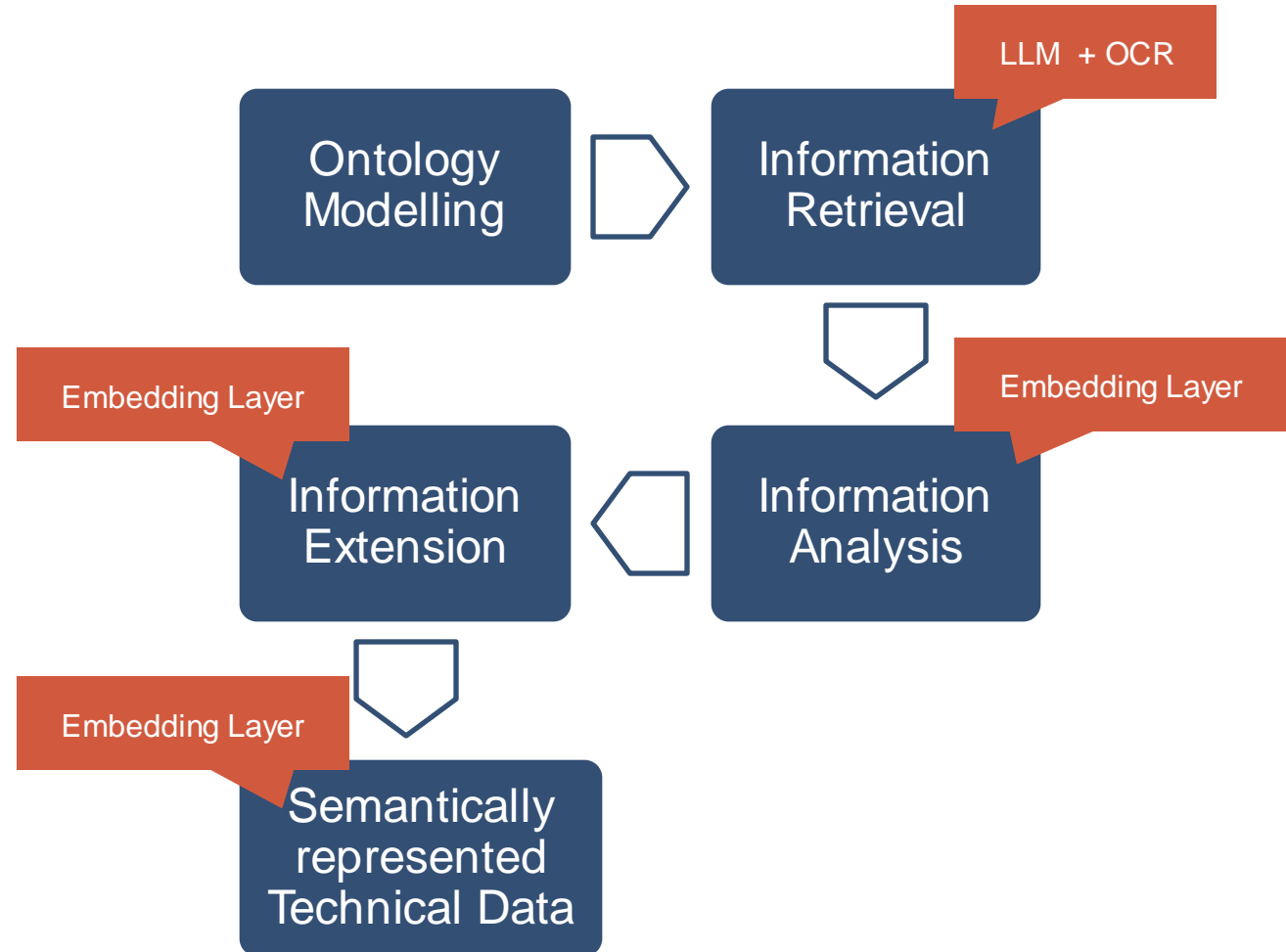
Information Validation

- Chat with technical data and know more
- LLMs are fine tuned over technical data
- Retrieved answer is validated on the Knowledge Graph



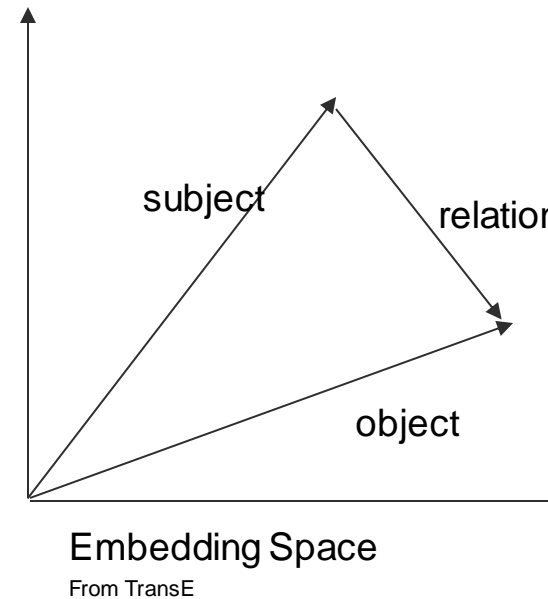
AI driven Workflow

- Contrary to conventional methods purely driven by rule-based mapping rules, AI driven methods perform better
- Information Analysis with embeddings adds another layer above SPARQL reports
- Information extension adds another layer along with Linked Data.



Existing KG Embedding Methods

- TransE
- RotatE
- CompGCN



<https://arxiv.org/abs/1911.03082>

https://proceedings.neurips.cc/paper_files/paper/2013/file/1cecc7a77928ca8133fa24680a88d2f9-Paper.pdf

<https://arxiv.org/abs/1902.10197v1>



Key Take Aways

AI driven methods need good connectivity and less dangling nodes for Knowledge Graph extraction and retrieval.

Training Models is an important part of AI and fine tuning should be carried out periodically.

KG Embeddings driven chatbot for technical documents would be another extension to the idea.



Bibliography

- https://proceedings.neurips.cc/paper_files/paper/2013/file/1cecc7a77928ca8133fa24680a88d2f9-Paper.pdf
- <https://arxiv.org/pdf/1911.03082.pdf>
- <https://arxiv.org/abs/1902.10197v1>
- <https://readthedocs.org/projects/roboy-sw-documentation-template/downloads/pdf/lite/>
- <https://www.pharmaceutical-networking.com/ascert1-product-data-sheet/>
- <https://docs.ampligraph.org/en/1.1.0/>
- <https://www.yumpu.com/en/document/read/42827500/used-equipment-data-sheet-napoleon-products-limited>





**Thank you for
your time!**