

A network diagram in the top left corner consisting of various sized nodes (circles) connected by solid and dashed lines, representing a graph structure.

Property Modelling for Product Ontology using Vector Embeddings driven by LLMs and OCR

PoolParty Summit 2024
Nikhil Acharya



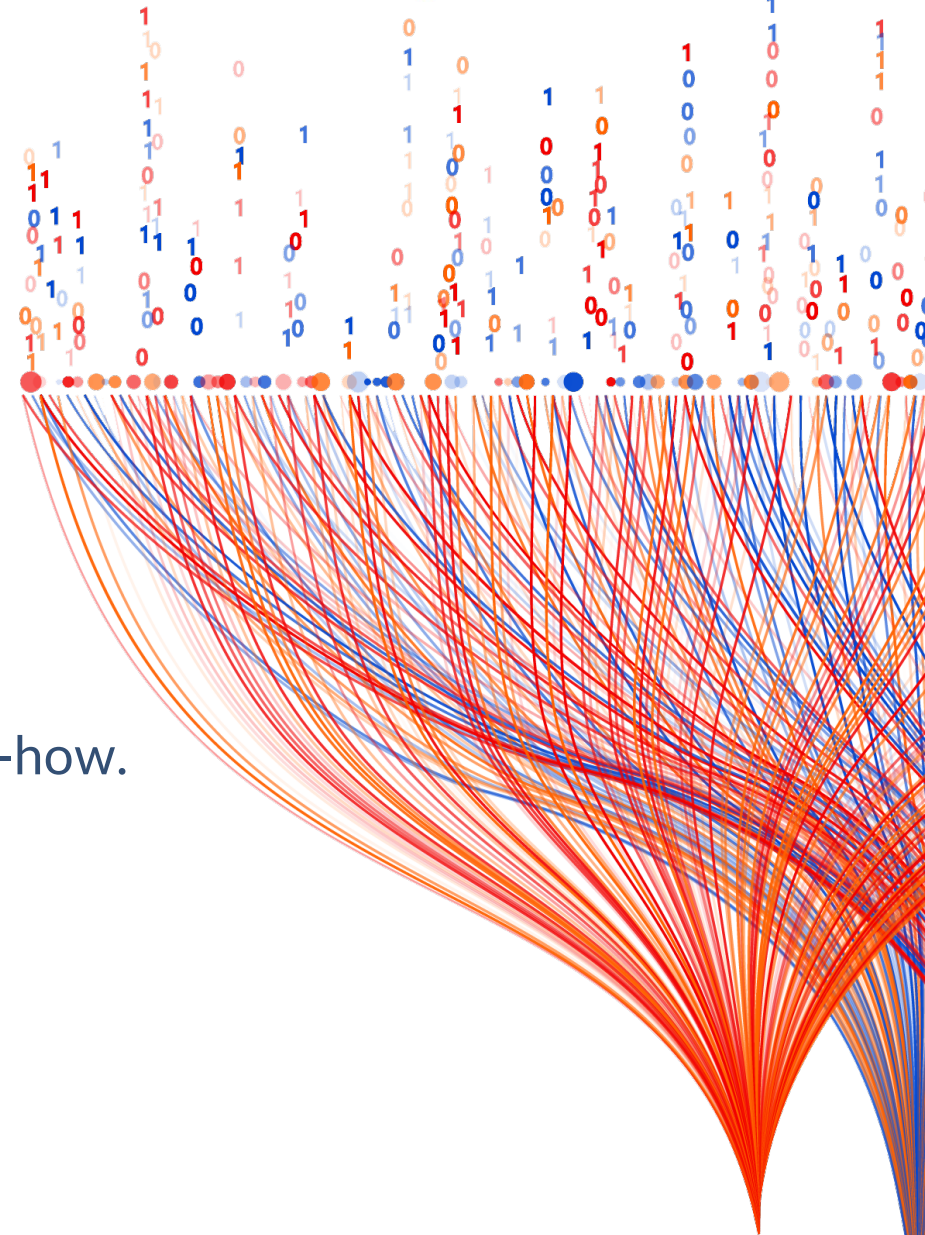
About PANTOPIX



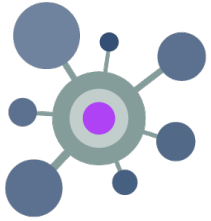
Intelligent Solutions for Technical Communication.

We guide our clients on their journey into the digital future.
Jointly creating the intelligent information landscape of tomorrow.

With a high level of consulting proficiency and technological know-how.

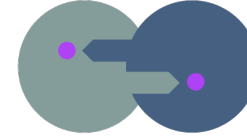


Turning Information into Value



Knowledge Graphs

KGs are used to describe different types of information in a metadata model.



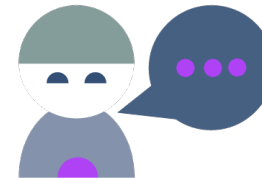
iiRDS

The technical standard enables a content-neutral transfer of information.



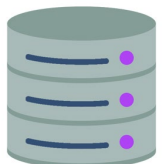
Artificial Intelligence

Ai as a driver for intelligent information provision.



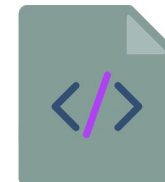
Service Information

Optimize service processes with the right information delivery tools.



PIM

Central management of product information for output to various channels.



DITA

Topic-based creation, distribution and use of technical information.

About me



Nikhil Acharya
Knowledge Engineer

”

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 to solve the challenges of our customers in the manufacturing industry”

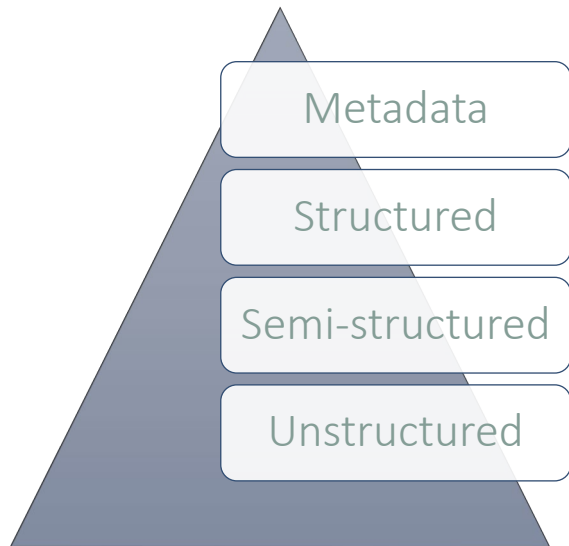
Introduction

Key Points

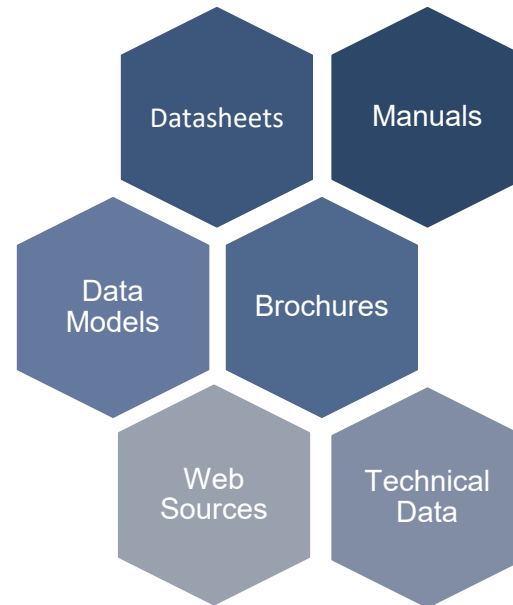
- Industries aims to manage, distribute and access their product information across different channels
- Product information comes with high degrees of heterogeneity
- Navigating and accessing product information across data sources is a huge challenge
- Strong need to extract and store dynamically changing product information accurately in a central location

Challenges

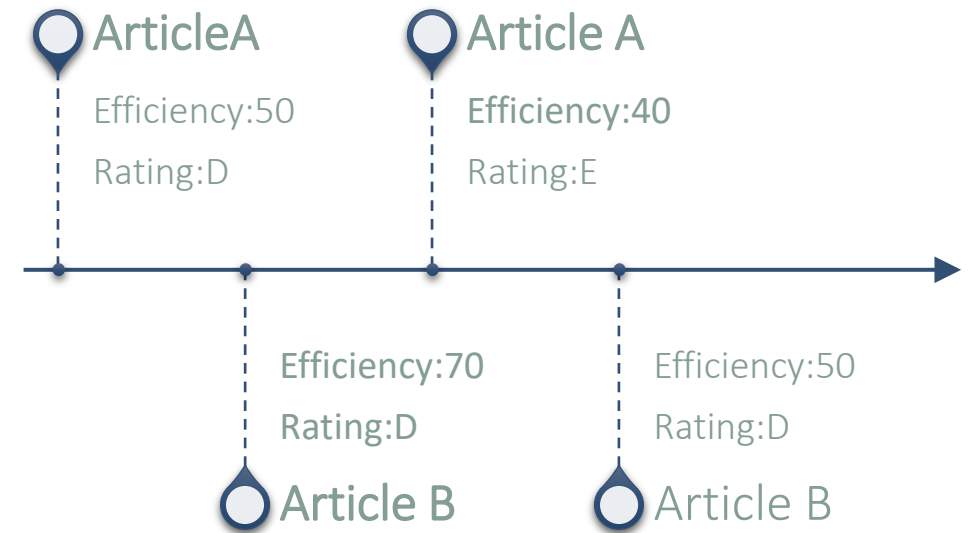
Model, capture and represent knowledge continuously



Complexity



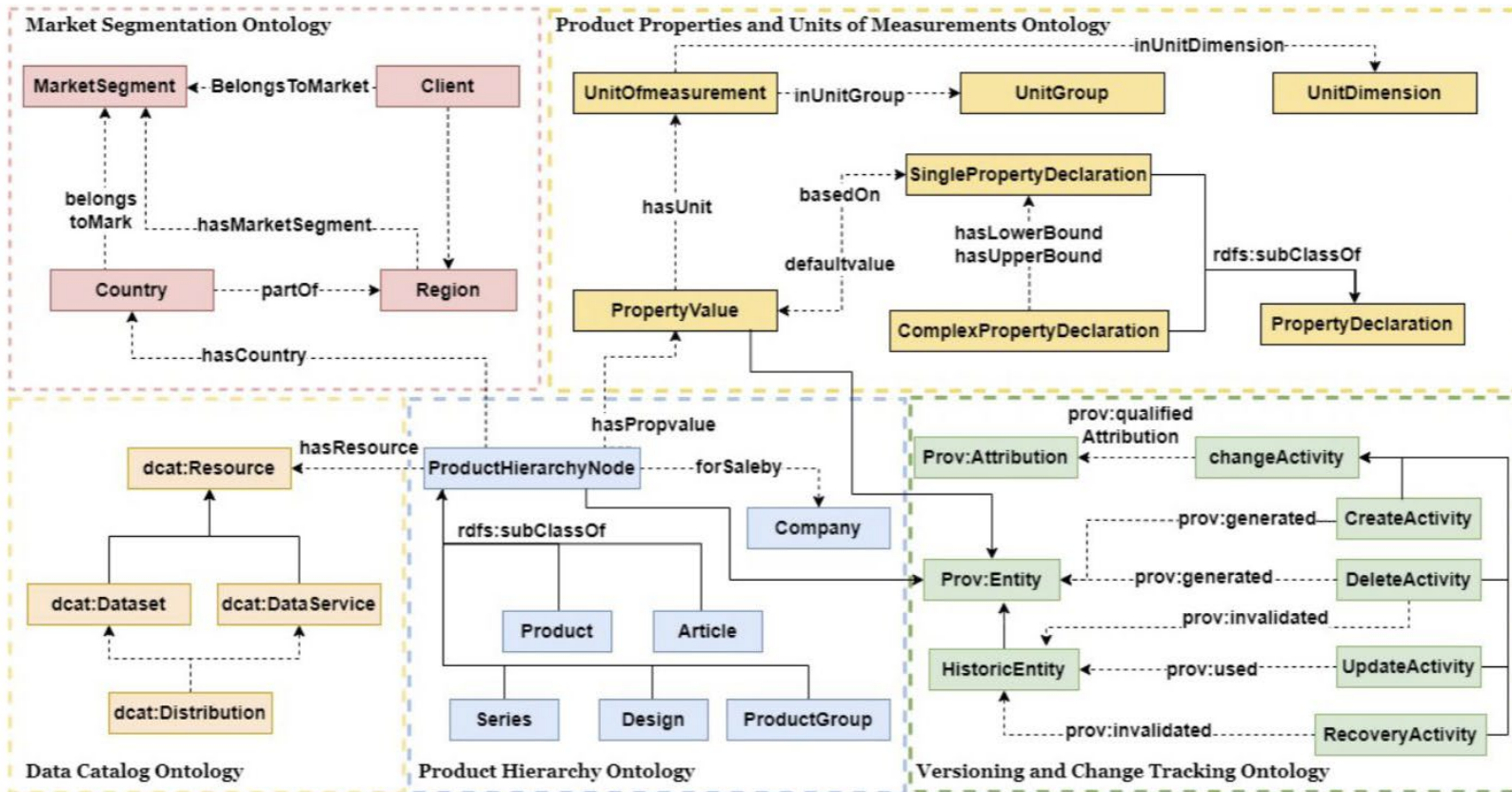
Heterogeneity



Evolution

Data Modeling

Modelling Information



Define set of Ontologies.
Keeping in mind:

- Product hierarchy
- Product properties and functions
- Product change tracking
- Product maintainability

Example

USED EQUIPMENT DATA SHEET

489

EQUIPMENT TYPE

Woodwaste Burning Industrial Hot Air Generator

MAKE Fagida
MODEL GX25D
SERIAL NO NPL489/ CH-4905
YEAR MF 2008
VOLTAGE 220V Single phase
DIMENSIONS MAX 900mm (35.5") W x 1700mm (67") D x 1830mm (71.75") H
WEIGHT APPROX 418 Kg

SPECIFICATION

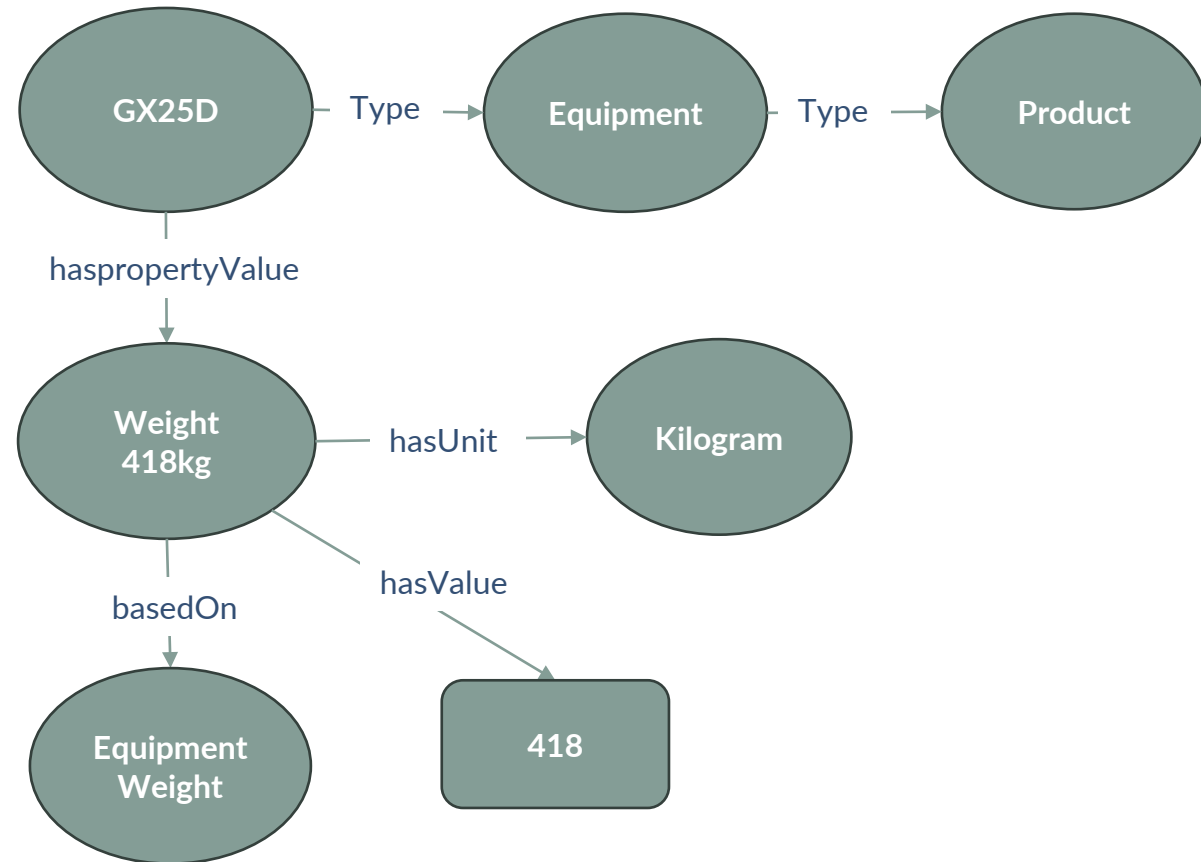
Woodwaste burning space heater
Generates 116,000 BTU/hr 34 kW/hr
200mm Flue outlet diameter
Double safety loading door
Loading door access 350mm x 525mm
Thermostatically controlled
Fan blows clean hot air generated by fire bricklined heat chamber to surrounding area
Lower ash grate for easy cleaning
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

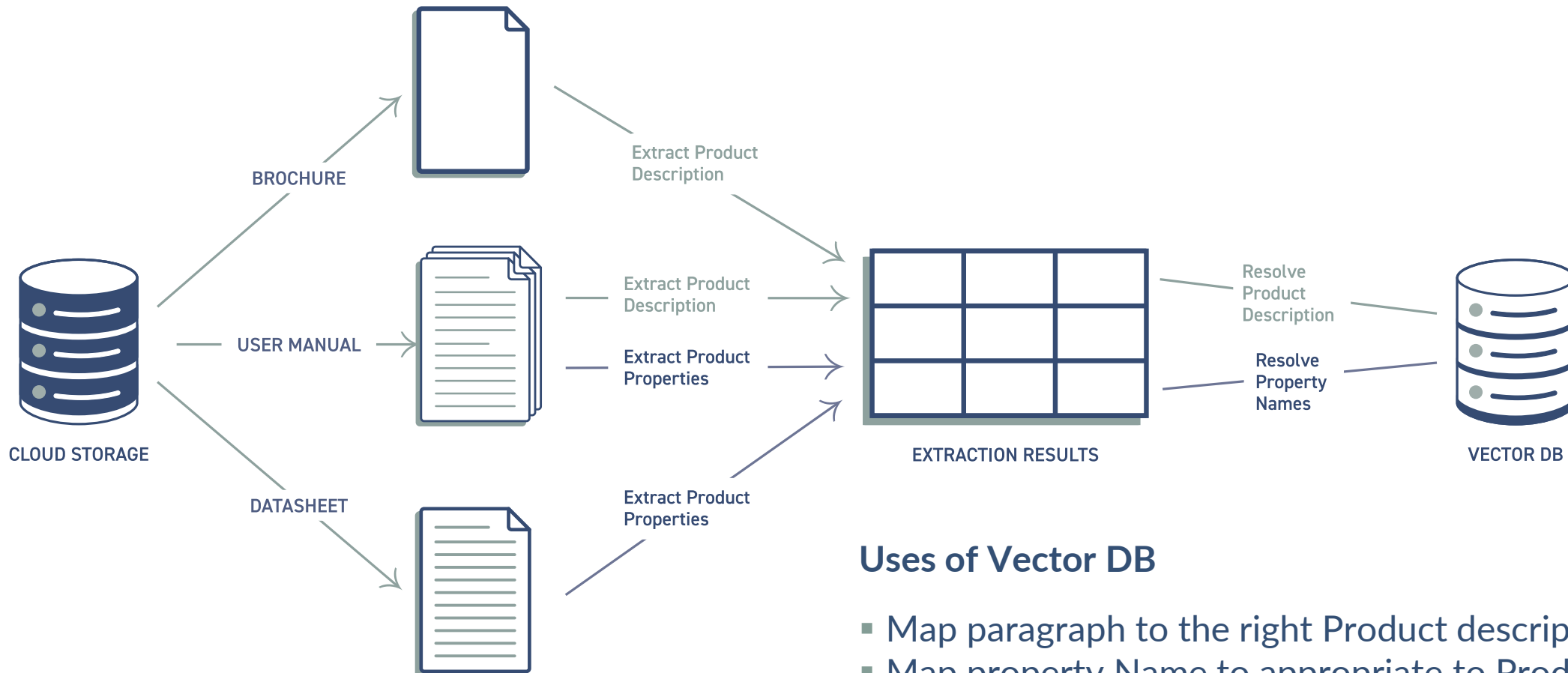


Data Extraction

Classifying Information

Type	Information	Format	Technique	Predefined properties	Predefined Entities
Brochure	Specific product description	Unstructured paragraphs	Entity recognition	No	Yes
Operation Manuals	Product operation procedures and product properties	Unstructured Paragraphs and Tabular data	Table Parser, Entity recognition	No	Yes
Product Datasheets	Contains Product properties	Tabular data	Table Parser, Form recogniser	Yes	Yes

Extracting Information



Uses of Vector DB

- Map paragraph to the right Product description
- Map property Name to appropriate to Product property
- Filter out irrelevant paragraphs and properties

Mapping Information

Extracted Information..

..is structured information about the product, the description and the properties.

- Map product to its appropriate Entity URI in KG
- Map product description to RDF description of Entity URI
- Validate and evaluate extracted properties for the entity (check domain and range, transform)
- Map extracted property value to appropriate relation and literal value

Transforming Extracted Information



USED EQUIPMENT DATA SHEET

489

EQUIPMENT TYPE

Woodwaste Burning Industrial Hot Air Generator

MAKE Fagida

MODEL GX25D

SERIAL NO NPL489/ CH-4905

YEAR MF 2008

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

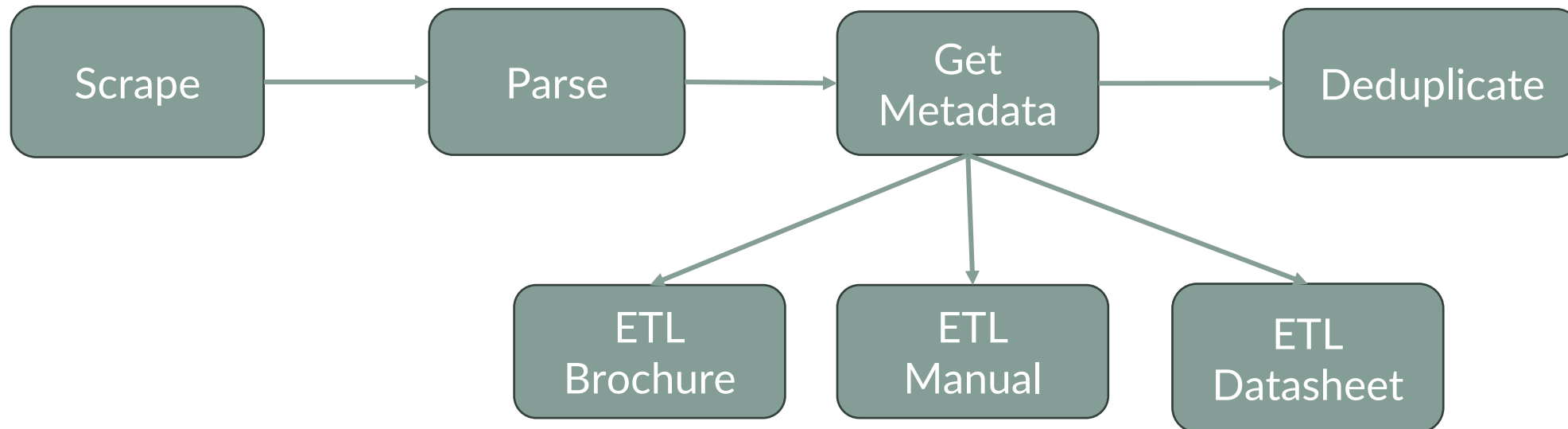
WARRANTY 12 Months See Used Equipment Terms Of sale

```
Extraction_489 = {  
    Equipment GX25D  
    Model: GX25D  
    Make: Fagida  
    Voltage: 220 V  
    Year: 08  
}
```

- Properties Model ,Make, Voltage, Year have property declaration
- Model, Make and Voltage values are successfully validated and will be converted to RDF
- Year property fails data validation and is discarded

Scaling Information

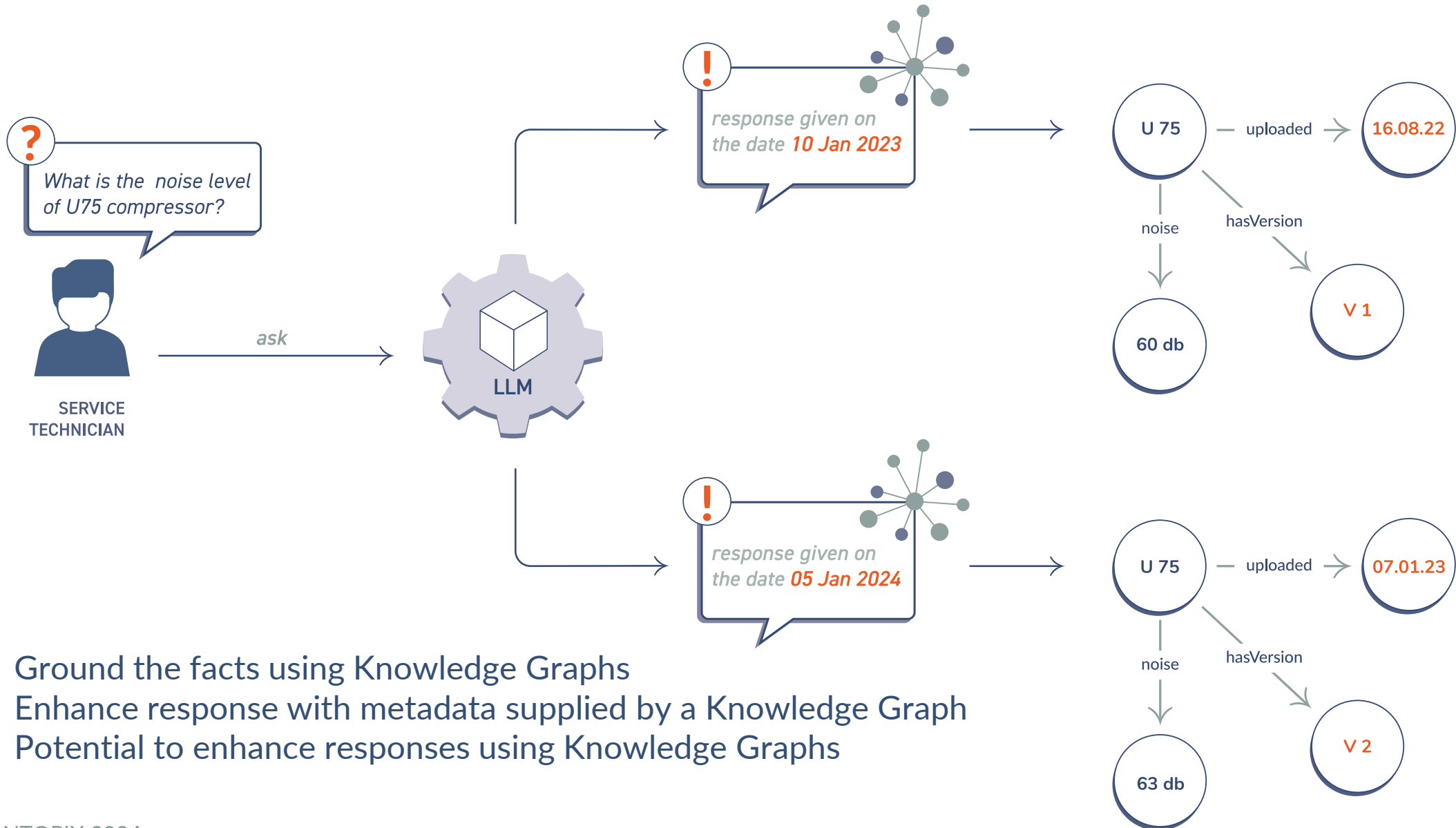
Big Data Pipeline



- Customers have large amounts of data. It demands a big data infrastructure.
- Scheduled scraping process looks out for new sources
- Metadata is crucial for Knowledge Graph Versioning (DCAT)
- Deduplication identifies and removes duplicates
- 3 separate ETL pipelines for each kind of data running in parallel

Use Case

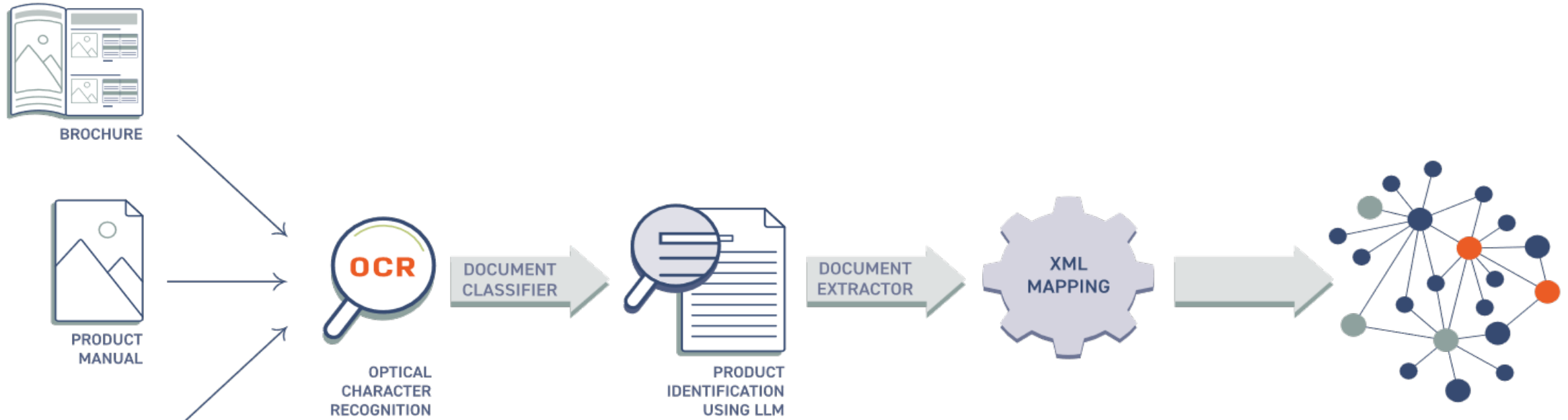
Question Answering



- Ground the facts using Knowledge Graphs
- Enhance response with metadata supplied by a Knowledge Graph
- Potential to enhance responses using Knowledge Graphs

Summary

The Full Pipeline



Conclusion

- Knowledge Graphs helps us overcome complexity, Heterogeneity and evolution of data
- Product Information, hierarchies, versions are very well captured by Product Information Models
- In the era of LLMs there is a strong need to ground responses and Knowledge Graphs come across as a useful tool for explaining the responses
- OCR models are very handy for extracting product information

Thank You for your
time
Questions?

Contact us



nikhil.acharya@pantopix.com

Visit our website at pantopix.com
Follow us on [LinkedIn](#) und [Twitter!](#)