Welcome to INFO216:
Knowledge Graphs

Spring 2024

Andreas L Opdahl
<Andreas.Opdahl@uib.no>

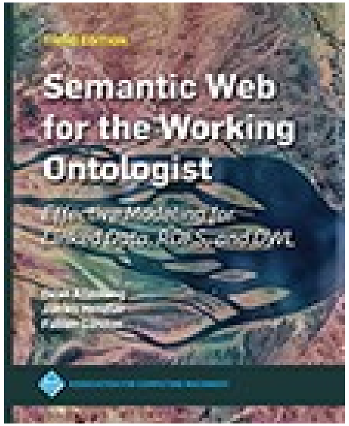
Session 7: Enterprise Knowledge Graphs

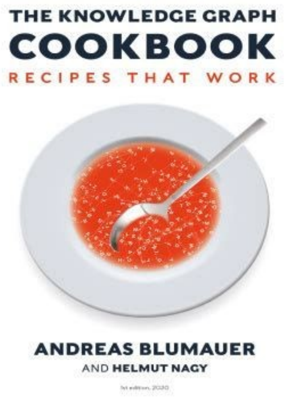
- Themes:
  - Open Knowledge Graphs *(← S05-S06)*
    - Linked Open Data resources / datasets
    - Wikidata, DBpedia, GeoNames, GDELT, WordNet, BabelNet, ConceptNet...
  - Enterprise Knowledge Graphs (EKGs)
    - Google's Knowledge Graph
    - Amazon's Product Graph
    - *Bosch' Line Information System (LIS)*
    - *the News Hunter platform*

Maybe later...

Readings

- Sources (suggested):
  - Blumauer & Nagy (2020):
    Knowledge Graph Cookbook – Recipes that Work:
    parts 2 and 4
- Resources in the wiki <http://wiki.uib.no/info216>:
  - *Introducing the Knowledge Graph: Things not Strings*, Amit Singhal, Google (2012)
  - *A reintroduction to our Knowledge Graph and knowledge panels*, Danny Sullivan, Google (2020)
  - *How Amazon's Product Graph is helping customers find products more easily*, Arun Krishnan, Amazon (2018)





Is anyone really using Knowledge Graphs?

Is anyone really using this?

Yes!

Tencent 腾讯
UniProt
USGS
Google
Bing
Alibaba.com
Bai du 百度
PubMed
facebook
ANTONI VAN LEEUWENHOEK FOUNDATION
MAASTRO
DEUTSCHE NATIONAL BIBLIOTHEK
The New York Times
BBC
europeana
GETTY
NXP
REUTERS
KUNGL. BIBLIOTEKET
National Library of Sweden
RENAULT
EPA
United States Environmental Protection Agency
IOS Press
European Commission
Walmart
SIEMENS
BEST BUY
Deloitte.
ELSEVIER
SPRINGER NATURE
amazon.com
accenture

Is anyone really using this?

Yes!

- But...
  - not quite as in the semantic web vision
  - not quite as in the LOD vision either
- Knowledge graphs are (additionally) becoming:
  - company internal
  - based on other technologies
    - such as general graph databases
  - not always linked to the LOD cloud

Is anyone really using this?

Yes!

- But...
  - not quite as in the semantic web vision
  - not quite as in the LOD vision either
- Knowledge graphs are (additionally) becoming:
  - company internal
  - based on other technologies
    - such as general graph databases
  - not always linked to the LOD cloud

Many of these ideas are widely adopted too, such as:

- microdata / schema.org
- RDF / SPARQL / … for semantic data exchange
- graph representations in general

Is anyone really using this?

Yes!

- But...
  - not quite as in the semantic web vision
  - not quite as in the LOD vision either
- Knowledge graphs are (additionally) becoming:
  - company internal
  - based on other technologies
    - such as general graph databases
  - not always linked to the LOD cloud

Similar ideas,
adapted to new uses
and business contexts,
using a combination of
standard and other
technologies

Google's Knowledge Graph

Google's Knowledge Graph

- Google Knowledge Graph (from 2012)
  - “Things, not Strings”
  - seeded from Freebase
  - facts from Wikipedia, Wikidata, CIA World Factbook
    - a growing number of other sources
  - used internally for many purposes
  - visible in Google Search results (Knowledge Panels)
  - question answering in Google Assistant / Home
  - semantic API (https://developers.google.com/knowledge-graph)
    - “returns only individual matching entities,
      rather than graphs of interconnected entities”

Caution: *The public documentation is limited, so this is compiled based on presentations, technical notes, forums etc.*

Google's Knowledge Graph

- Coverage:
  - claimed *(but be cautious)*
    - 18 billion facts (18G, norsk: 18 milliarder)
      about 570 million entities
    - 70 billion facts claimed in (2016)
    - *500 billion facts about five billion entities (2020)*
      - ...more than 3 times the size of the LOD cloud
  - from English to multiple languages
- Critiques:
  - source attribution, incl. Wikipedia / Wikidata
  - zero-click searches (around 25% of desktop searches)

Caution: *The public documentation is limited, so this is compiled based on presentations, technical notes, forums etc.*

Google's Knowledge Vault Project

- Attempt to extend the Knowledge Graph
  - covered resources not from open semantic datasets
  - facts extracted from the whole web
    - NLP of text documents
    - HTML trees and tables
    - human annotated pages (e.g., schema.org)
  - reported size
    - 1.6 billion facts
    - 271 million “confident” ones (90%)
  - *not put in production (never achieved 99% confidence target)*

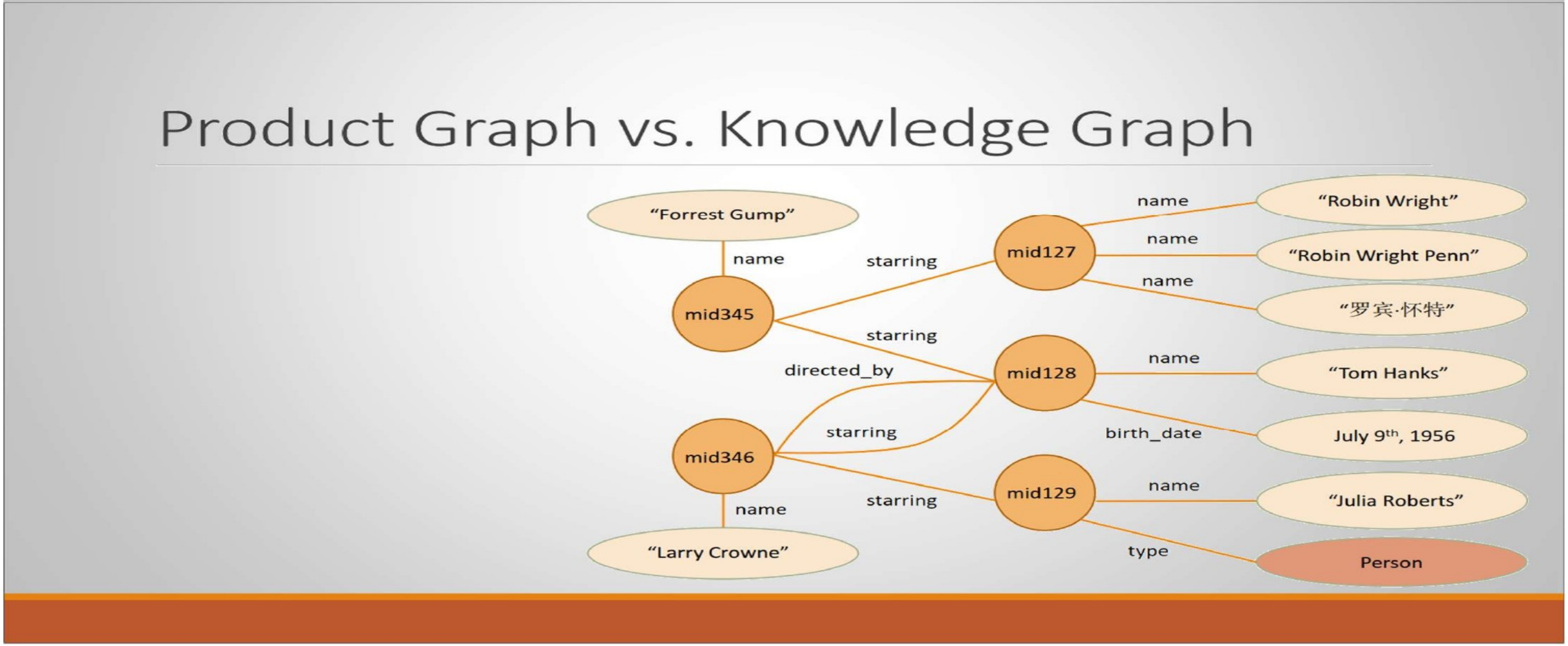
Caution: *The public documentation is limited, so this is compiled based on presentations, technical notes, forums etc.*

Amazon's Knowledge Graph

Amazon’s ambition (← S01)

- Let shoppers find the best products that fit their needs
  - allow greater variation in search terms
  - allow complex queries
- Ambition: *to structure all of the world’s information as it relates to everything available on Amazon*
- Describe every product on Amazon
  - both products and non-products
  - both concrete and abstract concepts
  - link related entities, both internal and external
- Enhanced customer experience
  - visit Amazon to see what’s new or interesting
  - discover ways to simplify and enrich their lives

Amazon



Product Graph vs. Knowledge Graph

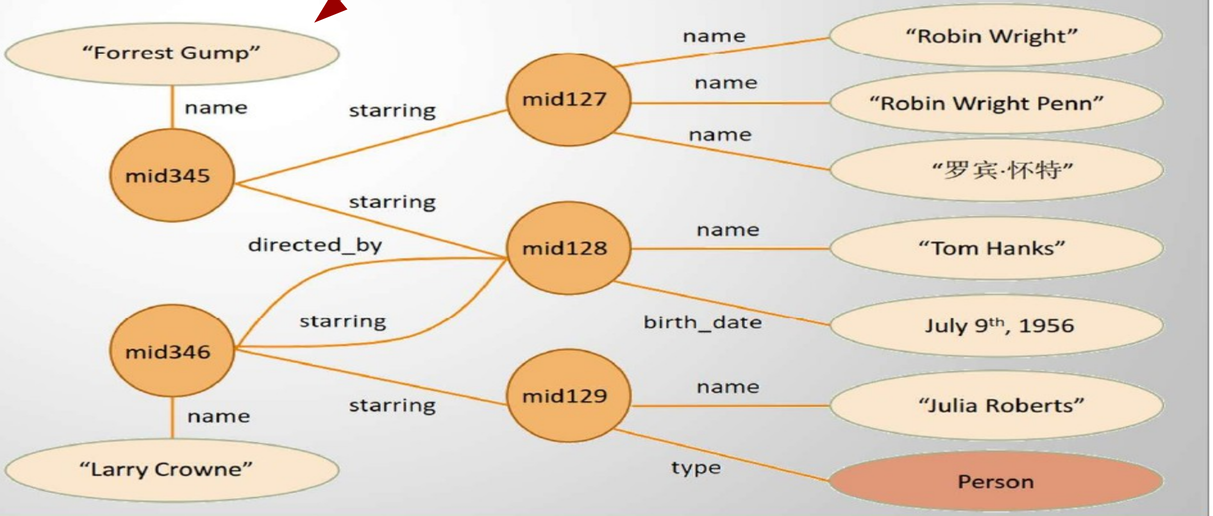
Frank van Harmelen (2018): Keynote at CAiSE'18

Amazon

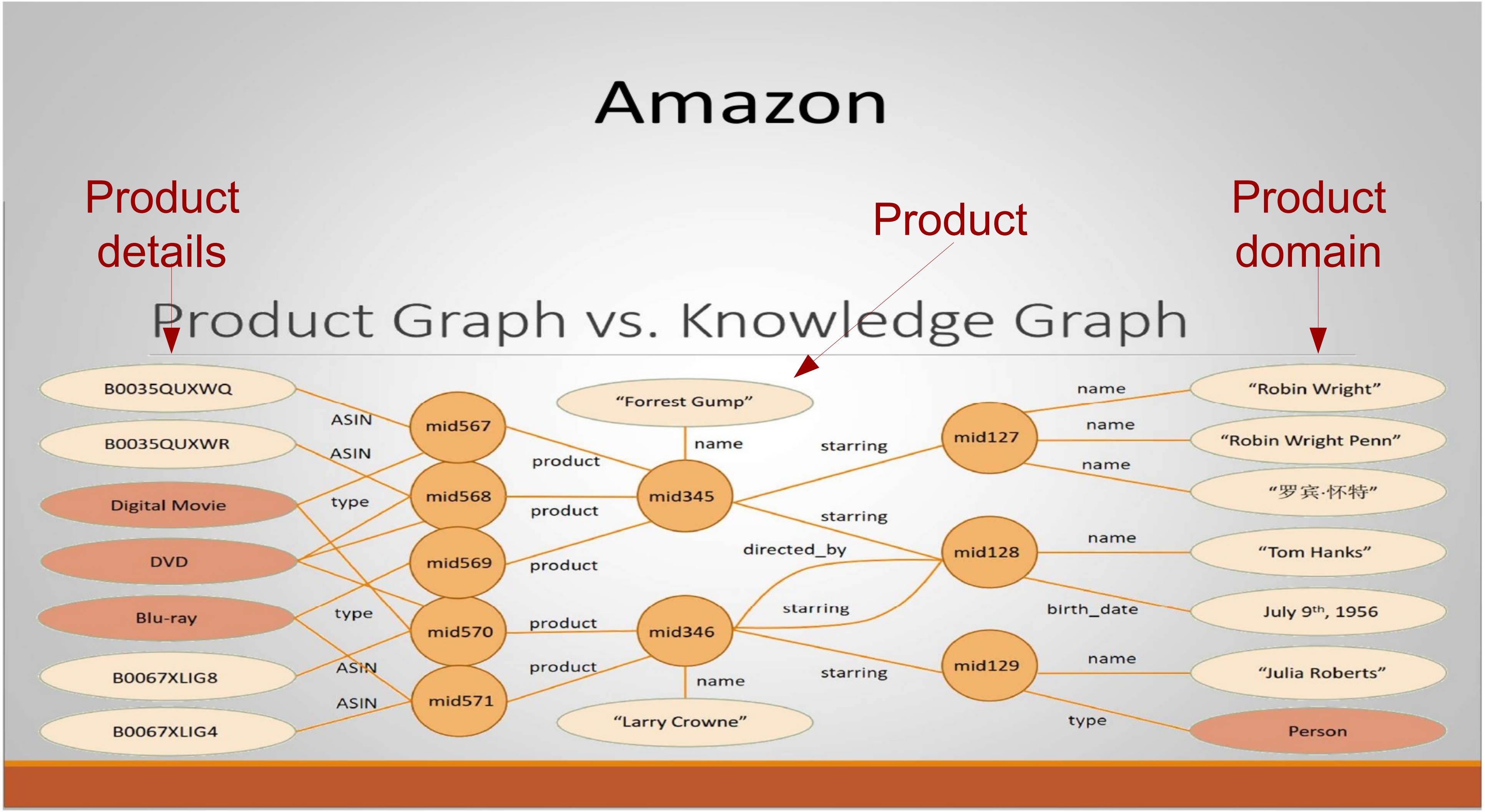
Product

Product domain

Product Graph vs. Knowledge Graph



Frank van Harmelen (2018): Keynote at CAiSE'18



Amazon

Product details

Product

Product domain

Product Graph vs. Knowledge Graph

Frank van Harmelen (2018): Keynote at CAiSE'18

Ratings & reviews

Amazon

Customers

Delivery services

Product details

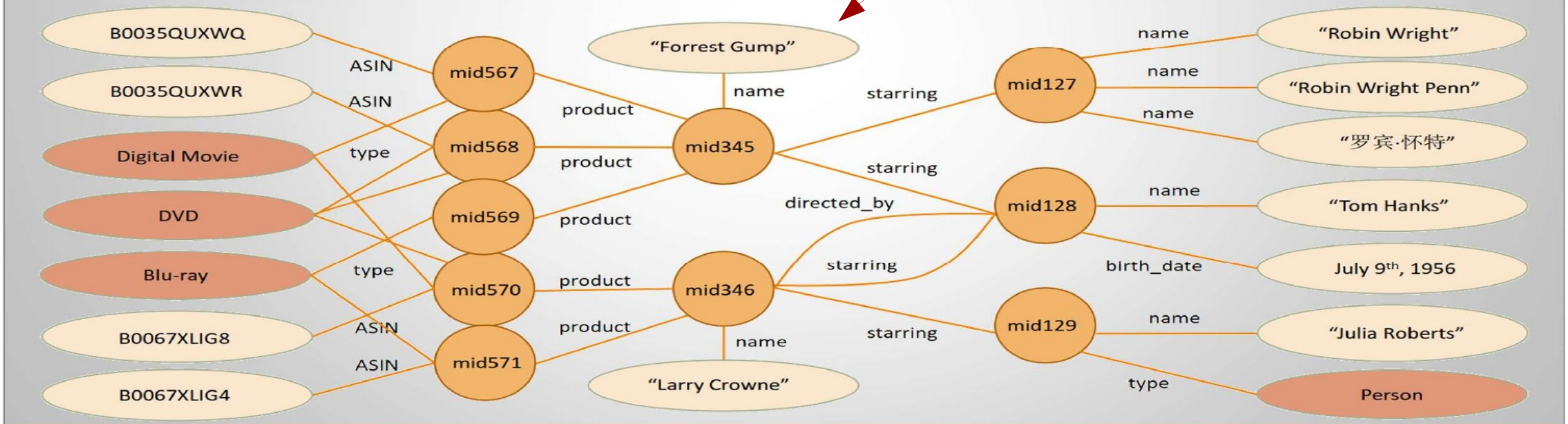
Suppliers

Support

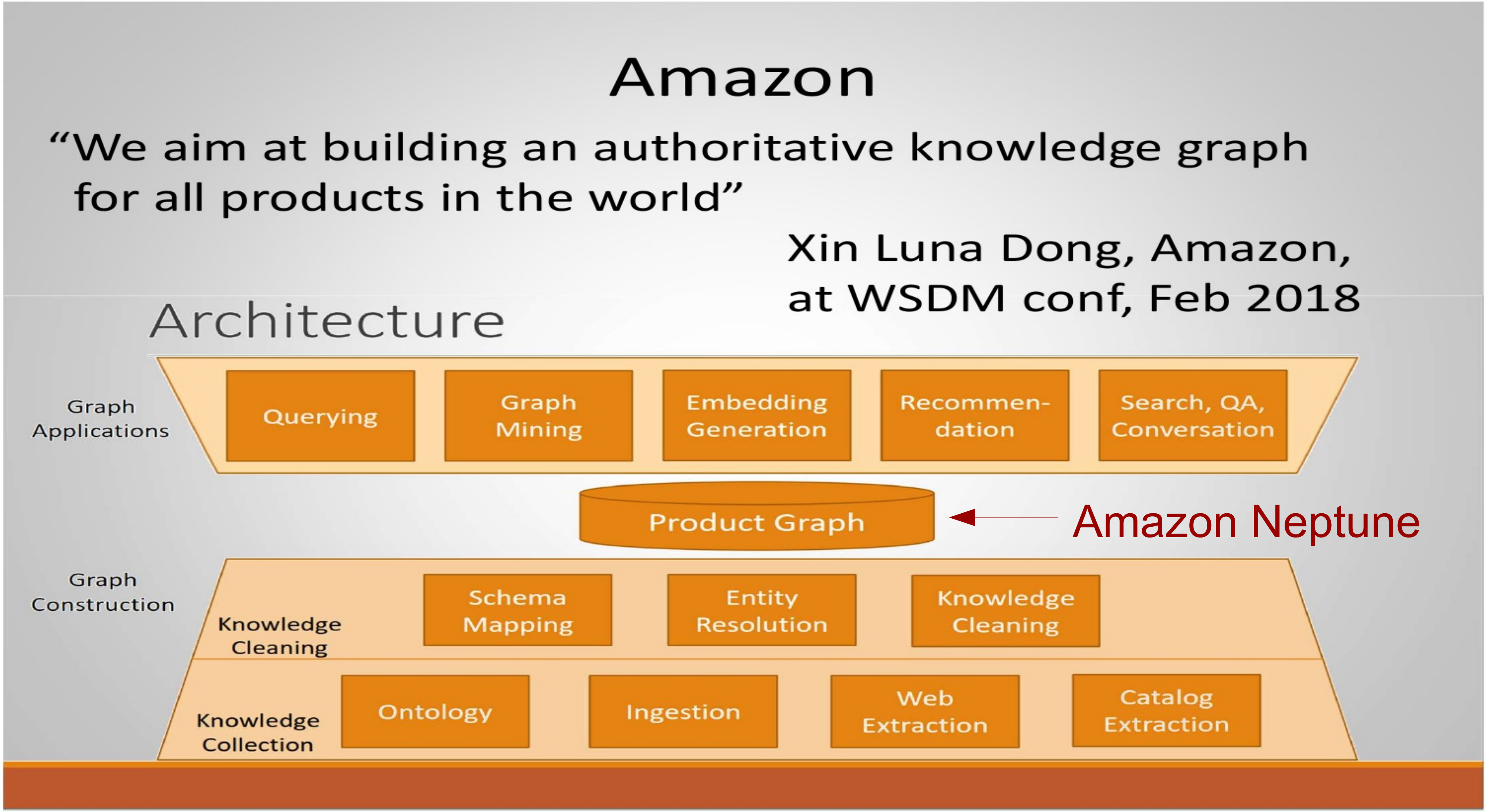
Product

Product domain

Product Graph vs. Knowledge Graph



Frank van Harmelen (2018): Keynote at CAiSE’18



Amazon

“We aim at building an authoritative knowledge graph for all products in the world”

Xin Luna Dong, Amazon,
at WSDM conf, Feb 2018

Architecture

Graph Applications

Querying

Graph Mining

Embedding Generation

Recommen-dation

Search, QA, Conversation

Product Graph

Amazon Neptune

Graph Construction

Knowledge Cleaning

Schema Mapping

Entity Resolution

Knowledge Cleaning

Knowledge Collection

Ontology

Ingestion

Web Extraction

Catalog Extraction

Frank van Harmelen (2018): Keynote at CAiSE’18

Challenges

- Ingest product-related information from Amazon's detail pages and from the Internet at large
  - product information is largely unstructured
  - trustworthiness of sources
- Machine learning techniques for
  - knowledge extraction, linkage and cleaning
  - distantly supervised learning (distant supervision)
    - use existing structured data to generate weak training data
    - train model on text data
  - open information extraction
  - graph mining techniques to identify interesting hidden patterns (buying product-X → buying product-Y)

Amazon AutoKnow

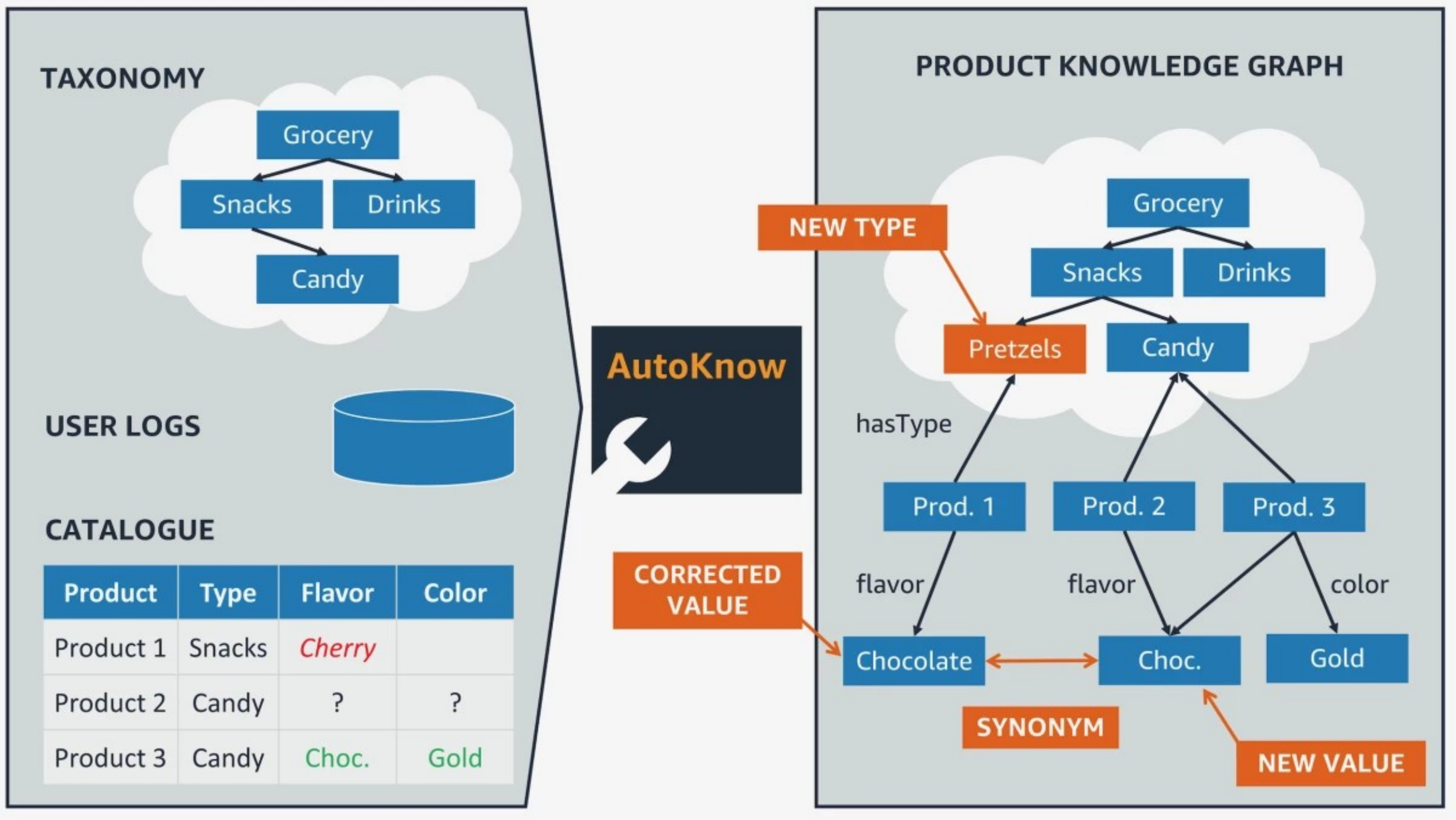
How to build a Product KG?

- Amazon’s AutoKnow:
  - a suite of techniques for automatically augmenting product KGs with both structured data and data extracted from free-form text sources
- Tasks:
  - combining data from different sources into a product graph
  - adding new product types to the taxonomy
  - adding new values for product attributes
  - correcting errors
  - identifying synonyms
- *“With AutoKnow, we increased the number of facts in Amazon’s consumables product graph (which includes the categories grocery, beauty, baby, and health) by almost 200%, identifying product types with 87.7% accuracy.”*

Challenges

- Retail information is hard:
  - the number of product types tends to grow as the graph expands
  - each product type has its own set of attributes
  - attributes vary widely, e.g.,
    color and texture versus battery type and effective range
  - the types of relationships between data items are essentially unbounded
  - vital product information exists in free-form text, e.g.,
    user reviews or question-and-answer sections

AutoKnow architecture



TAXONOMY

CATALOGUE

Product	Type	Flavor	Color
Product 1	Snacks	Cherry	
Product 2	Candy	?	?
Product 3	Candy	Choc.	Gold

AutoKnow architecture

- Inputs:
  - an existing product taxonomy
    - a graph structure
  - a product catalogue
    - structured information, such as labelled product names
    - unstructured product descriptions
  - user logs
    - free-form textual product-related information: customer reviews, product-related questions and answers; and product query data
- Output:
  - Amazon’s product graph

AutoKnow architecture

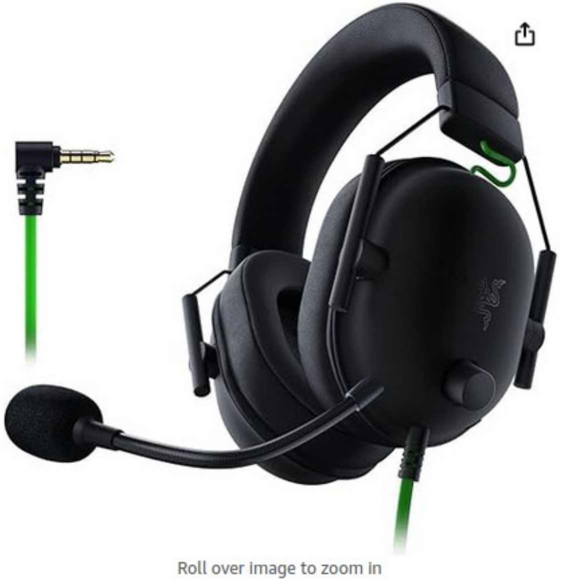
- Five modules in two suites:
  - Ontology suite
    1) taxonomy enrichment: identify and classify new entity types
    2) relation discovery: identifies (1) attributes of products, (2) their range of possible values, and (3) their importance to customers
  - Data suite
    3) data imputation: uses the entity types and relations to determine whether free-form text associated with products contains any information missing from the graph
    4) data cleaning: sorts through existing and newly extracted data to see whether any of it was misclassified
    5) synonym finding: identifies entity types and attribute values with identical/similar meaning

Taxonomy enrichment module

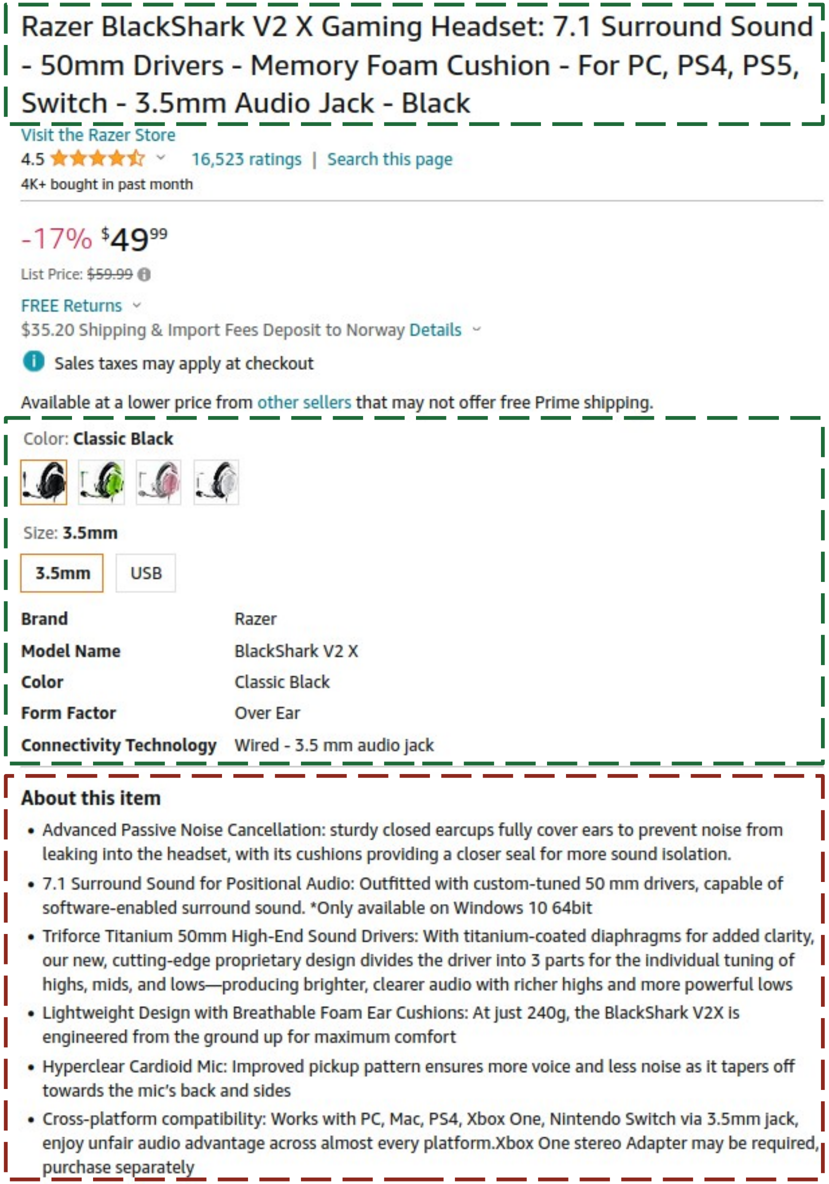
- Identification of new product types:
  - ML model labels substrings of product titles in the source catalogue.
    - also labels substrings that indicate product attributes
    - for use during the relation discovery step.
  - trained on product descriptions with hand-labelled types and attributes
- Classification of product types according to their hypernyms (i.e., the broader product categories that they fall under):
  - ML classifier uses data about customer interactions, such as which products customers viewed or purchased after a single query
  - trained on product data hand-labelled according to an existing taxonomy

Distant supervision

- How to auto-generate training labels from
  - free-text product descriptions and
  - semi-structured product data:
    - product type
    - attributes
    - attribute values


Roll over image to zoom in

More details, e.g., in product sheets and user manuals!



Razer BlackShark V2 X Gaming Headset: 7.1 Surround Sound - 50mm Drivers - Memory Foam Cushion - For PC, PS4, PS5, Switch - 3.5mm Audio Jack - Black

Visit the Razer Store

4.5 ★★★★☆ 16,523 ratings | Search this page

4K+ bought in past month

-17% $49.99

List Price: ~~$59.99~~

FREE Returns

$35.20 Shipping & Import Fees Deposit to Norway Details

Sales taxes may apply at checkout

Available at a lower price from other sellers that may not offer free Prime shipping.

Color: **Classic Black**

Size: **3.5mm**

3.5mm USB

Brand	Razer
Model Name	BlackShark V2 X
Color	Classic Black
Form Factor	Over Ear
Connectivity Technology	Wired - 3.5 mm audio jack

About this item

- Advanced Passive Noise Cancellation: sturdy closed earcups fully cover ears to prevent noise from leaking into the headset, with its cushions providing a closer seal for more sound isolation.
- 7.1 Surround Sound for Positional Audio: Outfitted with custom-tuned 50 mm drivers, capable of software-enabled surround sound. *Only available on Windows 10 64bit
- Triforce Titanium 50mm High-End Sound Drivers: With titanium-coated diaphragms for added clarity, our new, cutting-edge proprietary design divides the driver into 3 parts for the individual tuning of highs, mids, and lows—producing brighter, clearer audio with richer highs and more powerful lows
- Lightweight Design with Breathable Foam Ear Cushions: At just 240g, the BlackShark V2X is engineered from the ground up for maximum comfort
- Hyperclear Cardioid Mic: Improved pickup pattern ensures more voice and less noise as it tapers off towards the mic's back and sides
- Cross-platform compatibility: Works with PC, Mac, PS4, Xbox One, Nintendo Switch via 3.5mm jack, enjoy unfair audio advantage across almost every platform.Xbox One stereo Adapter may be required, purchase separately

Distant supervision from product descriptions

- Advanced Passive Noise Cancellation: sturdy closed earcups fully cover ears to prevent noise from leaking into the headset, with its cushions providing a closer seal for more **sound isolation**.
- **7.1 Surround Sound** for Positional Audio: Outfitted with custom-tuned **50 mm drivers**, capable of software-enabled surround sound. *Only available on Windows 10 64bit
- Triforce Titanium 50mm High-End Sound Drivers: With titanium-coated diaphragms for added clarity, our new, cutting-edge proprietary design divides the driver into 3 parts for the individual tuning of highs, mids, and lows—producing brighter, clearer audio with richer highs and more powerful lows
- Lightweight Design with Breathable Foam Ear Cushions: At just **240g**, the **BlackShark V2X** is engineered from the ground up for maximum comfort
- Hyperclear Cardioid Mic: Improved pickup pattern ensures more voice and less noise as it tapers off towards the mic’s back and sides
- Cross-platform compatibility: Works with PC, Mac, PS4, Xbox One, Nintendo Switch via **3.5mm jack**, enjoy unfair audio advantage across almost every platform.
  Xbox One stereo Adapter may be required, purchase separately

Distant supervision from product descriptions

- Generate labelled training data:
  - text, where attribute values from the structured data are labelled with attribute type
  - examples of attributes and values:
    - surround sound: “7.1 Surround Sound”
    - audio driver: “50 mm drivers”
    - noise control: “sound isolation”
    - weight: “240g”
    - model name: “BlackShark V2X”
    - compatibility: “PC”, “PS4”, “Switch”
    - connector type, headphones jack: “3.5mm jack”
- Fine-tune language model:
  - fine-tune a large language model on the labelled product descriptions
- Use language model (inference mode):
  - use to identify similar attributes in unlabelled descriptions

Distant supervision does not generate high-quality labels, but it is inexpensive and gives large training sets

Relation discovery module

- Classification of product attributes by two criteria and ML classifiers:
  - whether the attribute applies to a given product
    - example: flavour (an attribute) applies to food but not to clothes
  - how important the attribute is to buyers of a particular product
    - example: brand name (an attribute) is more important to buyers of snack foods than to buyers of produce
- Input data:
  - product descriptions from providers (attribute frequencies per product and per product type)
  - reviews and Q&As from customers (attribute frequencies per product)
  - manually-labelled data that match attributes with products

Data imputation module

- Identification of terms in product descriptions
  - that may fit the new product and attribute categories
  - but which are not yet represented in the KG
  - the product type is included among the inputs
- *Word embeddings* represent descriptive terms as points in a *vector space*
  - example terms:
    - product type: **Gaming Headset**
    - attributes: **model name**, **connector type**, **weight**, ...
    - attribute values: **BlackShark V2X**, **3.5mm jack**, **240g**, ...
  - the vector space is trained to group together related terms
  - some terms are labelled with product type or attribute:
    - if many terms in a cluster share the same label, should all the terms in the cluster have that label too?

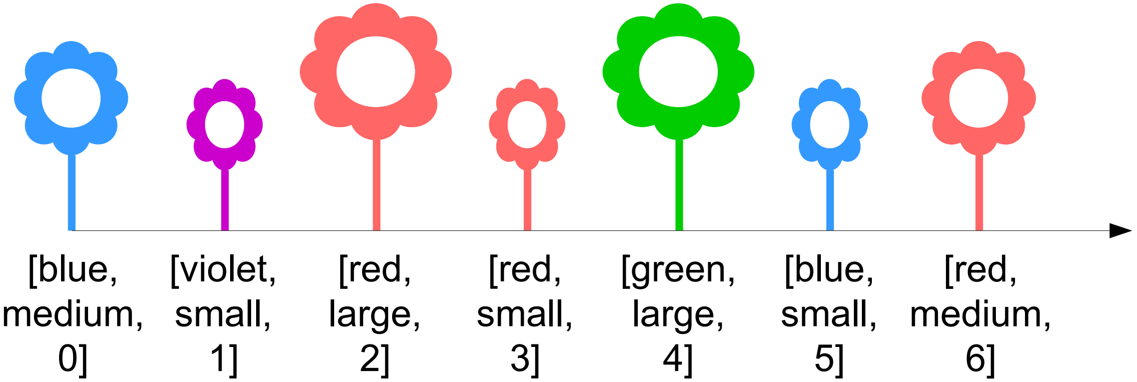
How can we represent the meaning of words?

- By designation (e.g., textual descriptions in a dictionary)
- As nodes in a network (e.g., in a knowledge graph in in WordNet)
- Formally (e.g., using an OWL ontology)
- As *vectors* in a *latent semantic space*!

How can we represent the meaning of words?

- By designation (e.g., textual descriptions in a dictionary)
- As nodes in a network (e.g., in a knowledge graph in in WordNet)
- Formally (e.g., using an OWL ontology)
- As *vectors* in a *latent semantic space*!
- Example:
  - *FlowerWorld™*
  - *“Everything is a flower!”*
  - *each flower has exactly three attributes:*
    - *colour*
    - *size*
    - *position*

Everything in FlowerWorld™ can be uniquely described by its position along three dimensions!

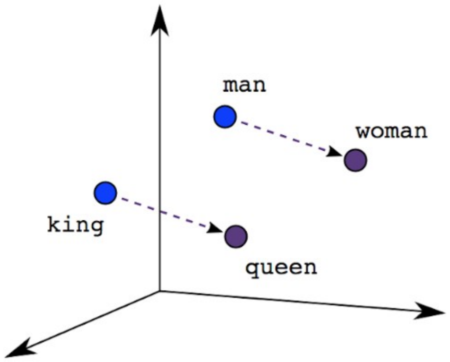


How can we represent the meaning of words?

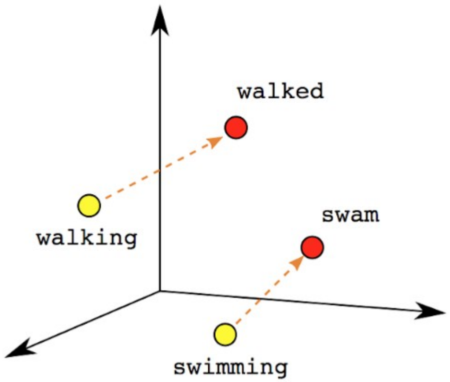
- By designation (e.g., textual descriptions in a dictionary)
- As nodes in a network (e.g., in a knowledge graph in in WordNet)
- Formally (e.g., using an OWL ontology)
- As *vectors* in a *latent semantic space*!
- (Our conceptualisations of) Things in the “real world”:
  - a bit more complex...
  - not fully describable by positions along dimensions
  - but perhaps we can describe them usefully by adding more dimensions?
  - but which dimensions to add?
    - use machine learning / neural networks to analyse large text corpora!

How can we represent the meaning of words?

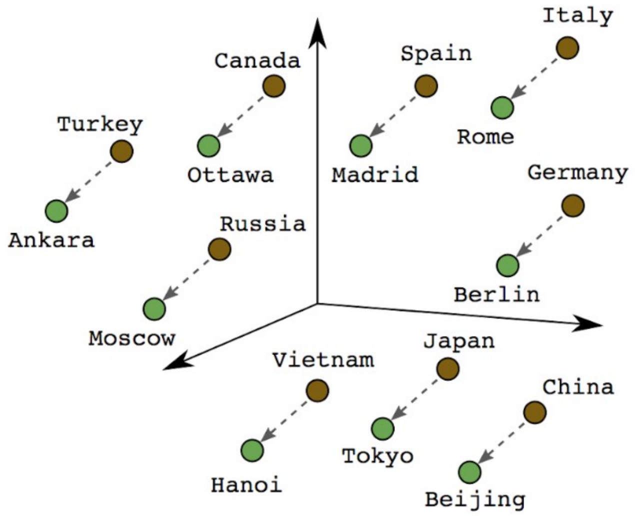
- By designation (e.g., textual descriptions in a dictionary)
- As nodes in a network (e.g., in a knowledge graph in in WordNet)
- Formally (e.g., using an OWL ontology)
- As *vectors* in a *latent semantic space*!



Male-Female



Verb Tense



Country-Capital

These examples only show a few selected axes...

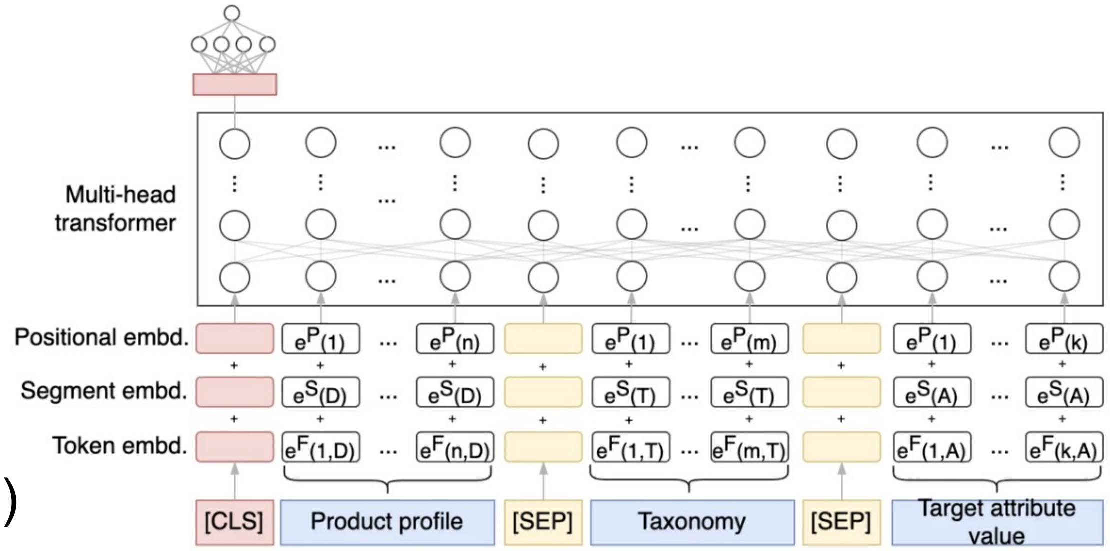
How can we represent the meaning of words?

- By designation (e.g., textual descriptions in a dictionary)
- As nodes in a network (e.g., in a knowledge graph in in WordNet)
- Formally (e.g., using an OWL ontology)
- As *vectors* in a *latent semantic space*!
  - normalised values: [0.01 0.62 0.03 … 0.41]
  - important use: as inputs to deep neural networks that process NL text
  - trained, e.g., so that similar words are close to one another
  - ...so that position differences between words can be systematic
    - [Paris] – [France] + [Italy] ≈ [Rome]
  - ...so that position differences between words can represent relations
    - [J. K. Rowling] + [influenced by] ≈ [J. R. R. Tolkien]

Data imputation module (repeat)

- Identification of terms in product descriptions
  - that may fit the new product and attribute categories
  - but which are not yet represented in the KG
  - the product type is included among the inputs
- *Word embeddings* represent descriptive terms as points in a *vector space*
  - example terms:
    - product type: **Gaming Headset**
    - attributes: **model name**, **connector type**, **weight**, ...
    - attribute values: **BlackShark V2X**, **3.5mm jack**, **240g**, ...
  - the vector space is trained to group together related terms
  - some terms are labelled with product type or attribute:
    - if many terms in a cluster share the same label,
      should all the terms in the cluster have that label too?

Data cleaning module



- Detecting bad attribute values
  - using a *transformer model*
  - inputs:
    - NL product description
    - an attribute (e.g., flavour...)
    - an attribute value (e.g., vanilla...)
  - is the attribute-value pair aligned with the product?
- Trained on
  - positive examples: valid attribute-value pairs that occur across many instances of the product type (e.g., all ice cream types have flavours)
  - negative examples: generated by random replacement of values in valid attribute-value pairs

Synonym finding module

- Analysis of product and attribute sets to find mergeable KG nodes
  - customer interaction data to identify items that were viewed during the same queries
    - their product and attribute descriptions are candidate synonyms
  - a combination of techniques to filter the candidate terms
    - edit distance
    - neural network

Ongoing work

- Open questions:
  - how to handle products with multiple hypernyms (i.e., products that have multiple “parents” in the product hierarchy)?
  - how to clean data before it’s used to train our models?
  - how to use image data + textual data to improve model performance

Bosch's
Line Information System
(LIS)

Industry 4.0

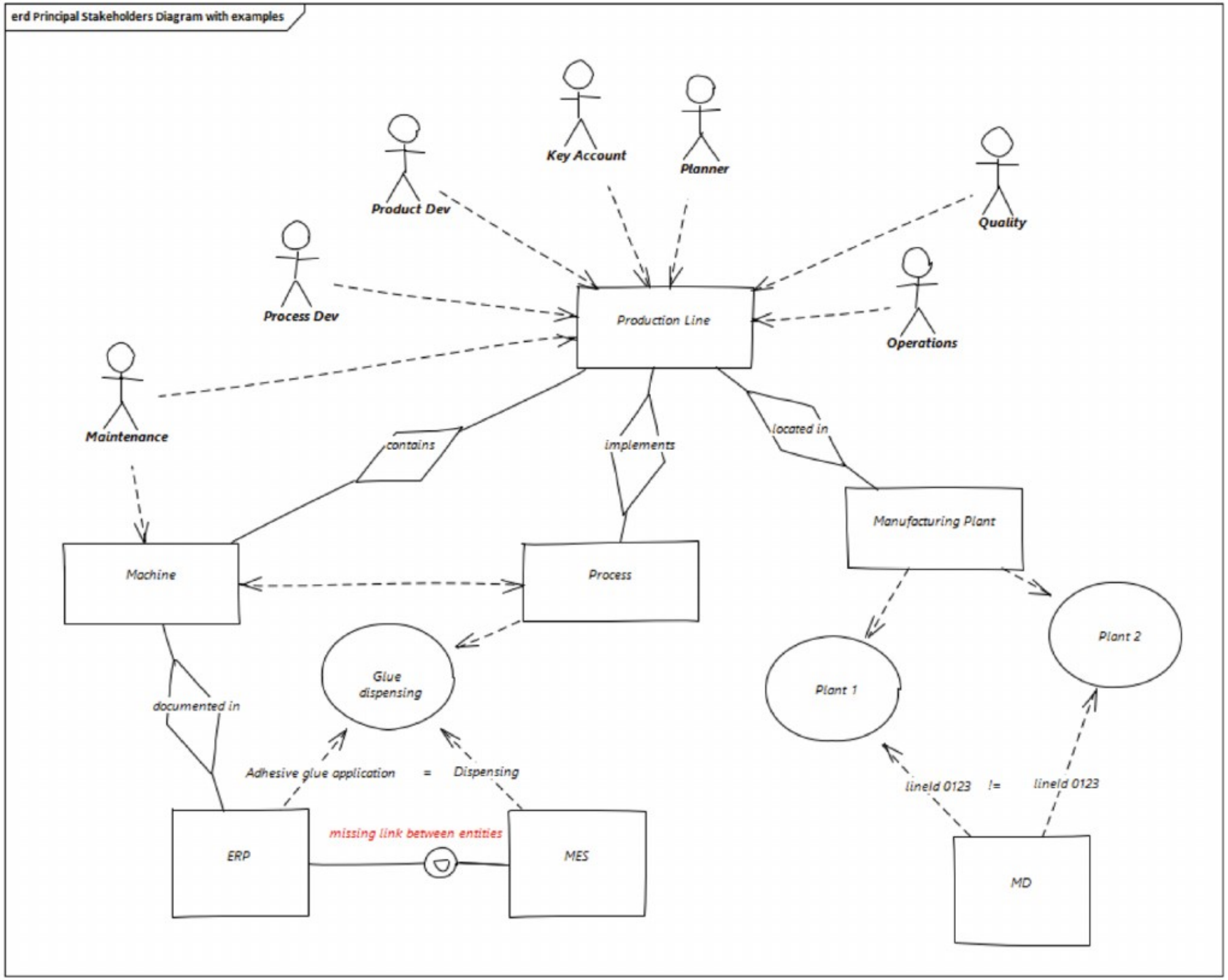
- Also called the “Fourth Industrial Revolution”
  - increasing automation and data exchange in manufacturing technologies
- Key components and technologies:
  - *Internet of Things (IoT)*: connecting devices and machinery to enhance operational efficiency and enable predictive maintenance.
  - *Cloud Computing and Analytics*: flexible cloud infrastructures of vast data amounts for better decision-making and optimized processes
  - *Artificial Intelligence (AI) and Machine Learning*: analysis of data to identify patterns, predict outcomes, and automate decision-making to increase efficiency and innovation further
  - *Smart Factories*: transformation into smart factories that are more efficient, adaptive, and can self-optimize performance across a broader network to automating processes and improve manufacturing operations

Bosch’s Line Information System (LIS)

- Bosch GmbH
  - a German multinational engineering and technology company
  - a *manufacturing* enterprise
    - automotive parts, power tools, security systems, home appliances, engineering, electronics, cloud computing, Internet of Things (IoT)
  - *production lines* are central
    - a defined number and sequence of *production processes* with specified capabilities to manufacture or assemble a *product* until ready for shipment to the *customer*
    - processes are realized by *physical assets* or *machines*
    - in the processes, *value* is added to the product by *materials* and *resource consumption*, e.g., operations personnel, machine wear, maintenance...

Bosch's Line Information System (LIS)

- Central concepts
  - *manufacturing plant*
  - *production lines*
  - *stakeholders*
  - *production processes*
  - *machines*
  - *systems:*
    - *ERP*
    - *MES*
    - *MD*



(c) Andreas L Opdahl, 2024

The demand

- Management tasks require *semantically integrated information*:
  - production planning and operation
  - product and production process development
  - production process optimization
  - purchase
  - quality management
  - traceability of products
- To answer business questions
  - all data must be integrated and semantically harmonized
  - different views must be reconciled into a uniform understanding of the domain

The problem

- Before LIS, answering business questions about production lines involved different *stakeholders* and *systems*, e.g.,
    - Manufacturing Execution Systems (MES)
    - Enterprise Resource Planning (ERP) systems
    - Master Data (MD) systems
  - all have different views on the same *production line*
  - the data reside in isolated *silos*
  - also unstructured data: intranets, word, and pdf documents
  - also non-IT-available data in the head of experts (which ones?)
- Example: long-term production planning
  - needs integrated information about processes/machines
  - the information is disconnected and inconsistently named

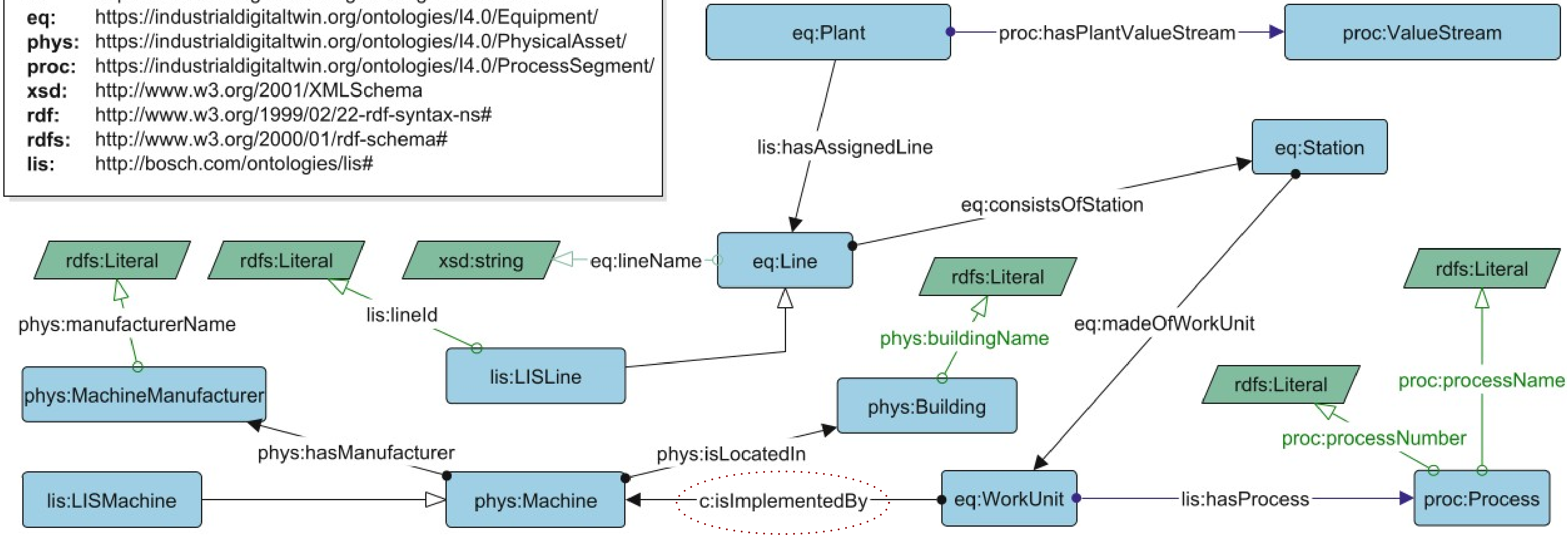
Bosch’s Line Information System (LIS)

- A Knowledge Graph (KG)-based ecosystem
  - enables a 360° view of manufacturing data for all stakeholders
  - allows querying available data in an integrated way
- Central components:
  - LIS ontology (formal and semantic data model)
  - data mappings
  - semantically integrated data:
    - MES, ERP, and MD
    - 12 Bosch plants, > 1100 production lines, > 16 000 physical machines, > 400 manufacturing processes
  - procedure to ensure the quality of the data in the KG
  - procedure to resolve *semantic interoperability conflicts*

LIS ontology *(more about ontologies later!)*

Prefixes

c:	https://industrialdigitaltwin.org/ontologies/I4.0/Core/
eq:	https://industrialdigitaltwin.org/ontologies/I4.0/Equipment/
phys:	https://industrialdigitaltwin.org/ontologies/I4.0/PhysicalAsset/
proc:	https://industrialdigitaltwin.org/ontologies/I4.0/ProcessSegment/
xsd:	http://www.w3.org/2001/XMLSchema
rdf:	http://www.w3.org/1999/02/22-rdf-syntax-ns#
rdfs:	http://www.w3.org/2000/01/rdf-schema#
lis:	http://bosch.com/ontologies/lis#

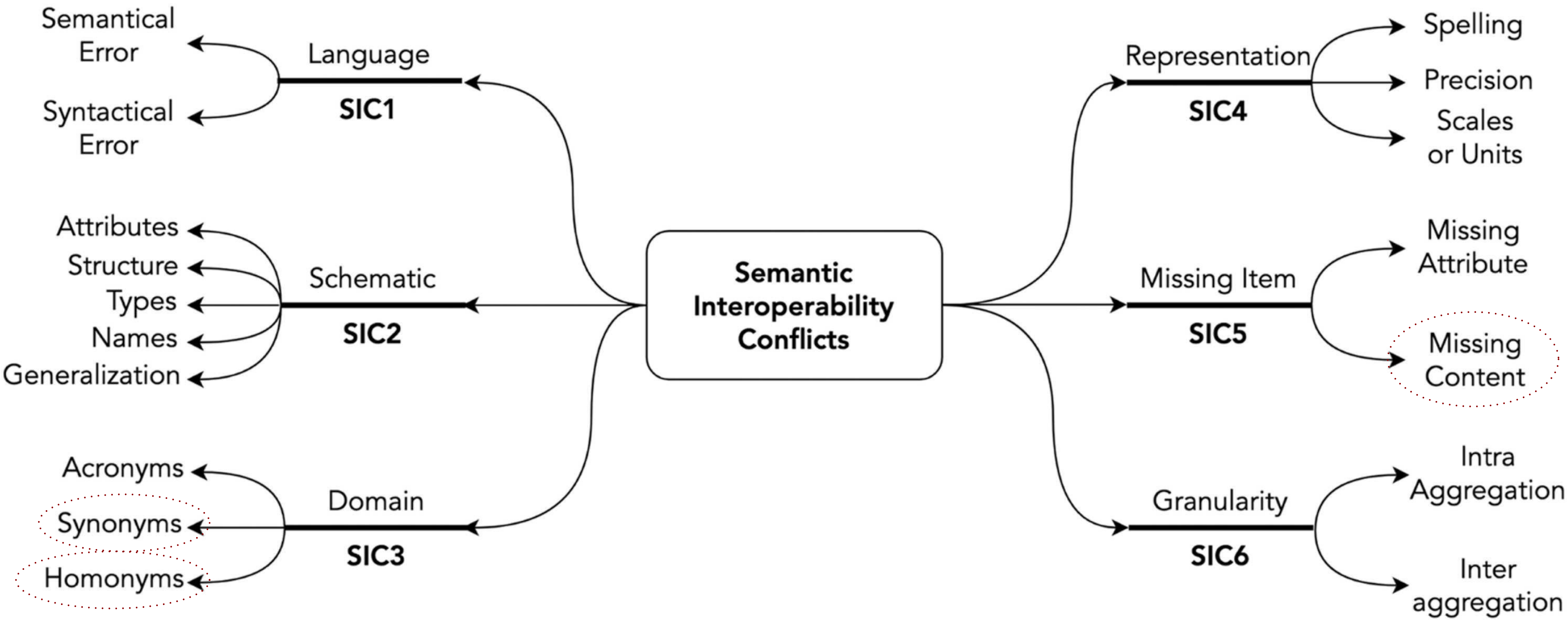


Extends existing standard for Core Information Model for Manufacturing (CIMM) from the Industrial Digital Twin association

Data mapping problems

- Manufacturing planning
  - forecast production requirements for 1-8 years
  - requires a complete overview about available production lines
  - must be correlated with the customers demand of products
  - *production line identifiers are not unique*
- Data integration:
  - data about machines and production processes are distributed in different ERP and MES systems
  - a process (MES) is implemented by a machine (ERP)
  - *no link between physical assets and logical processes*
- Missing standardization rules:
  - *different process names for same process*

Semantic Interoperability Conflicts (SICs)



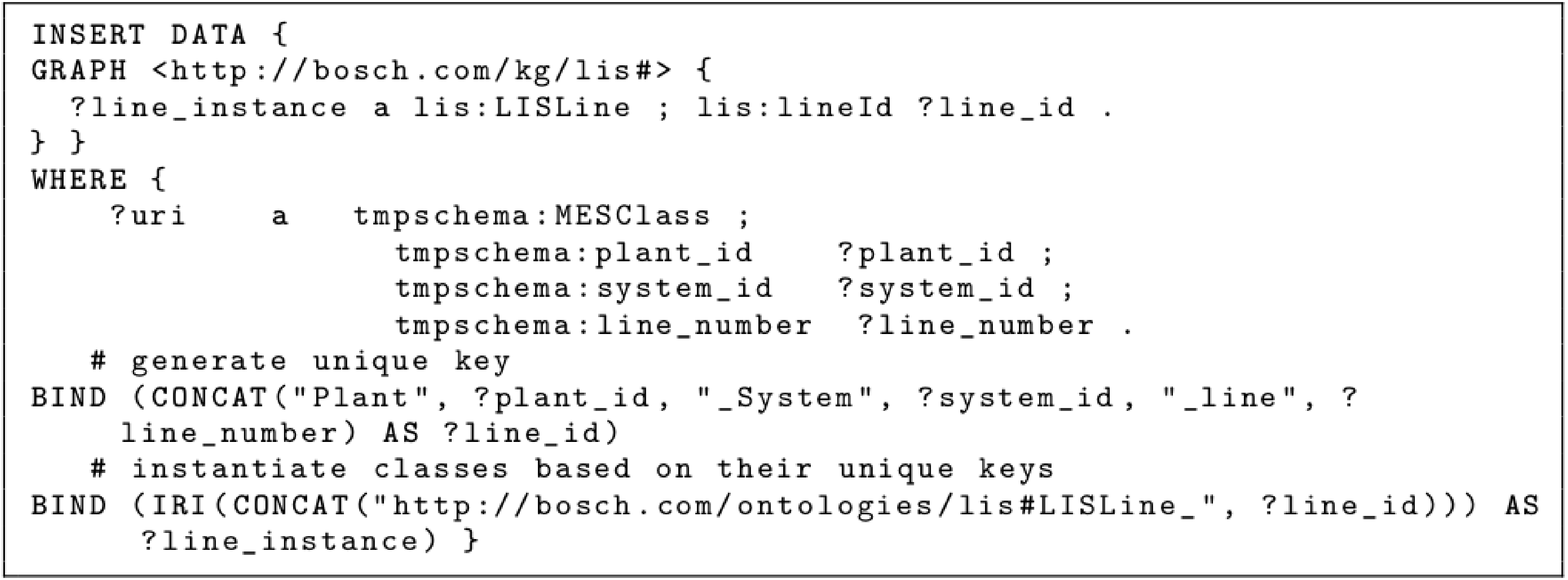
Melluso, N., Grangel-González, I., & Fantoni, G. (2022).
Enhancing industry 4.0 standards interoperability via knowledge graphs with natural language processing. Computers in Industry, 140, 103676.

Semantic Interoperability Conflicts (SICs)

- Domain (SIC1): different interpretations of the same domain are represented
  1. **homonyms**: the same name is used to represent concepts with different meaning
  2. **synonyms**: distinct names are used to model the same concept
  3. **acronyms**: different abbreviations for the same concept are employed
- Schematic (SIC2): sources that are modeled using different schemas
  1. different **attributes** representing the same concept in different sources
  2. the same concept is modelled using different **structures** in the distinct data sources
  3. different **types** that represent the same concept
  4. the same concept is described at different levels of specialization/**generalization**
  5. different **names** that represent the same concept
- Granularity (SIC3): different granularity is given to the same domain
  1. **intra-aggregation**: the same data is divided differently, e.g., full person names against first-middle-last
  2. **inter-aggregation**: appears when there exist sums or counts as added values
- Representation (SIC4): different representations are used to model the same concept
  1. Different **scales or units**
  2. Various values of **precision**
  3. Incorrect **spellings**
- Missing Item (SIC5): different items in distinct data sources are missing
  1. **missing attributes**
  2. **missing content**
- Language (SIC6): different languages are used to represent the data or metadata, i.e., schema
  1. **semantical mis-match**
  2. **syntactical mis-match**

Melluso, N., Grangel-González, I., & Fantoni, G. (2022).
Enhancing industry 4.0 standards interoperability via knowledge graphs
with natural language processing. Computers in Industry, 140, 103676.

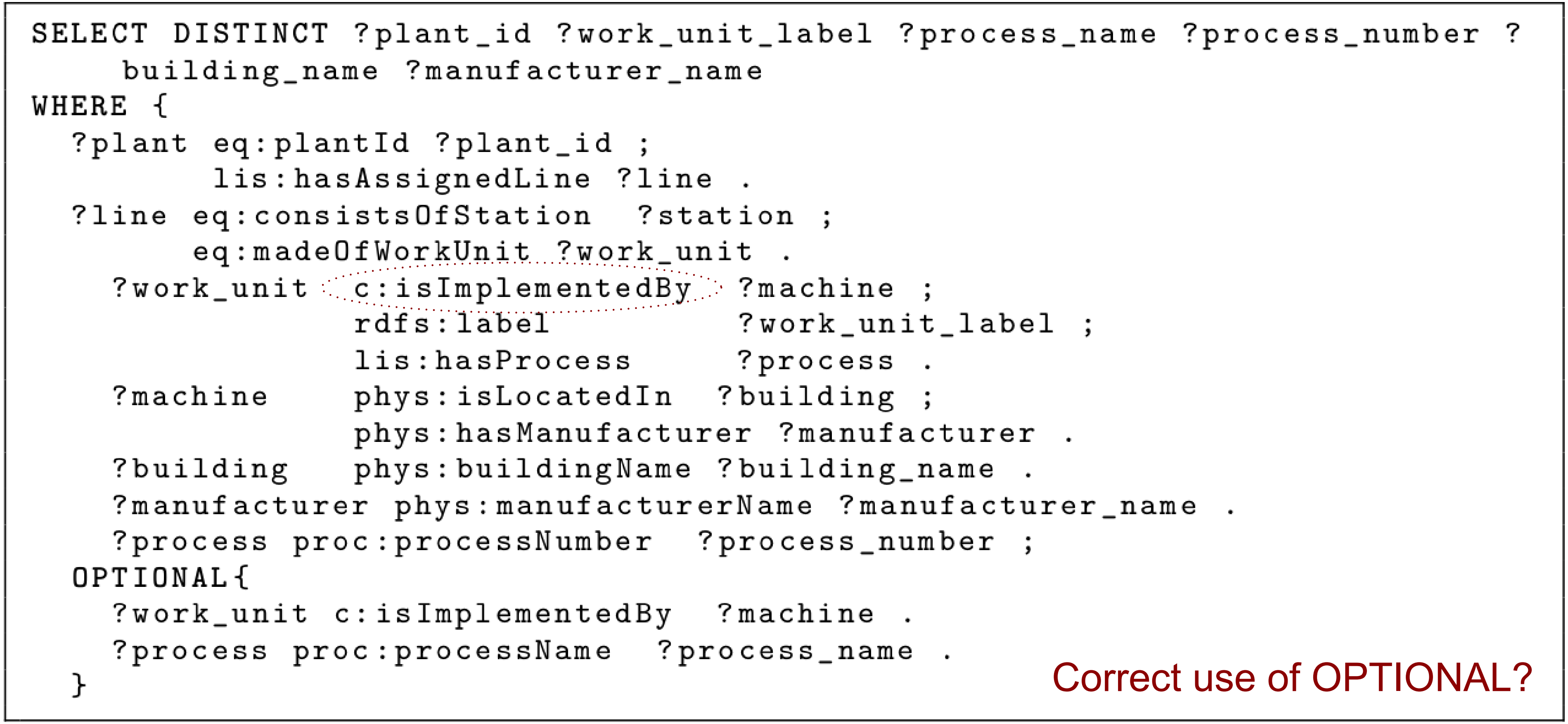
Data mapping



```
INSERT DATA {
GRAPH <http://bosch.com/kg/lis#> {
  ?line_instance a lis:LISLine ; lis:lineId ?line_id .
} }
WHERE {
    ?uri    a   tmpschema:MESClass ;
                  tmpschema:plant_id    ?plant_id ;
                  tmpschema:system_id   ?system_id ;
                  tmpschema:line_number  ?line_number .
   # generate unique key
BIND (CONCAT("Plant", ?plant_id, "_System", ?system_id, "_line", ?
     line_number) AS ?line_id)
   # instantiate classes based on their unique keys
BIND (IRI(CONCAT("http://bosch.com/ontologies/lis#LISLine_", ?line_id)) AS
      ?line_instance) }
```

SPARQL Update to ensure that line identifiers in the KG are unique across manufacturing plans and systems

Example query: integrated process information

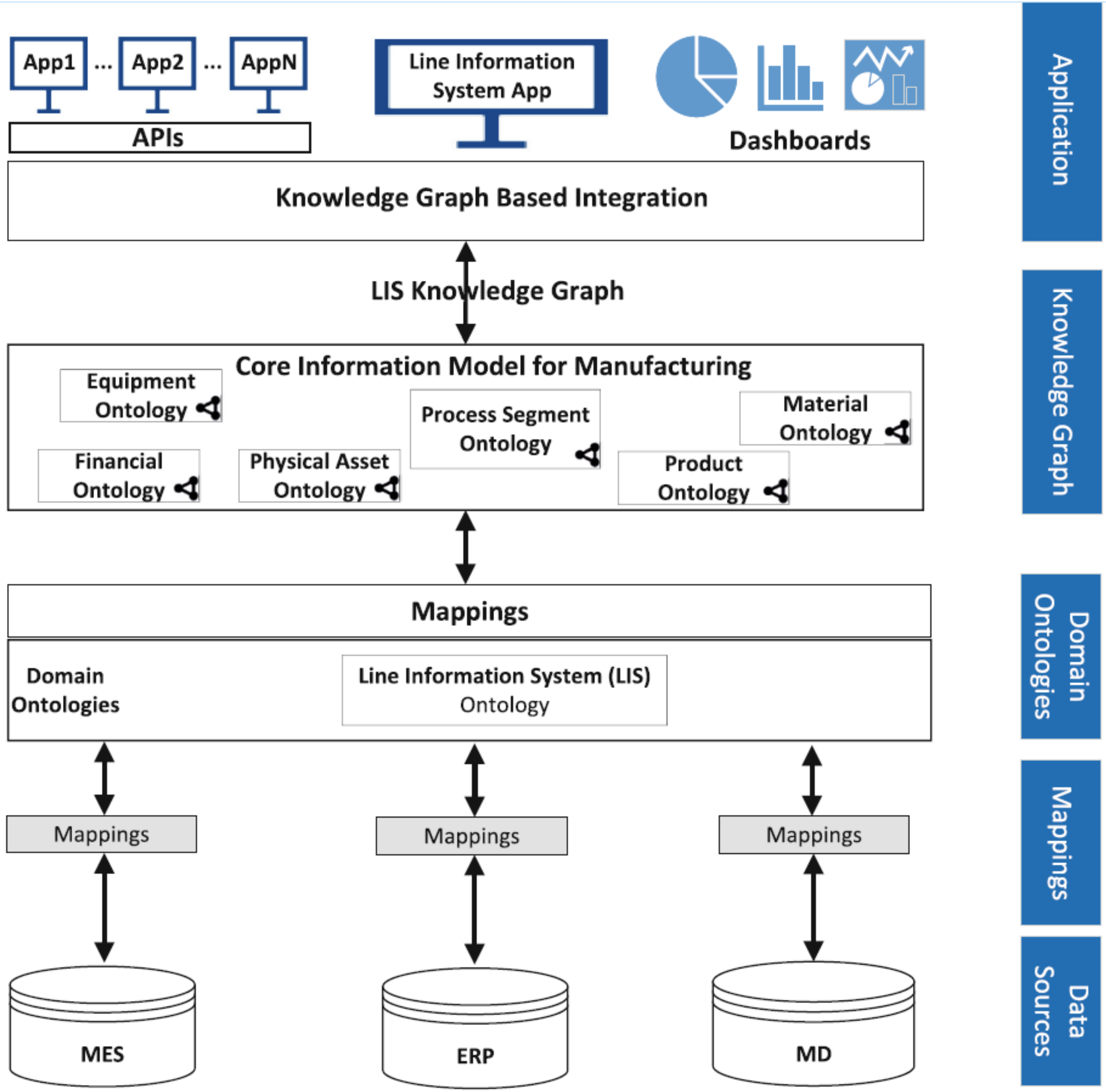


```
SELECT DISTINCT ?plant_id ?work_unit_label ?process_name ?process_number ?
    building_name ?manufacturer_name
WHERE {
  ?plant eq:plantId ?plant_id ;
         lis:hasAssignedLine ?line .
  ?line eq:consistsOfStation  ?station ;
        eq:madeOfWorkUnit ?work_unit .
    ?work_unit  c:isImplementedBy  ?machine ;
                rdfs:label         ?work_unit_label ;
                lis:hasProcess     ?process .
    ?machine    phys:isLocatedIn  ?building ;
                phys:hasManufacturer ?manufacturer .
    ?building   phys:buildingName ?building_name .
    ?manufacturer phys:manufacturerName ?manufacturer_name .
    ?process proc:processNumber  ?process_number ;
  OPTIONAL{
    ?work_unit c:isImplementedBy  ?machine .
    ?process proc:processName  ?process_name .
  }
```

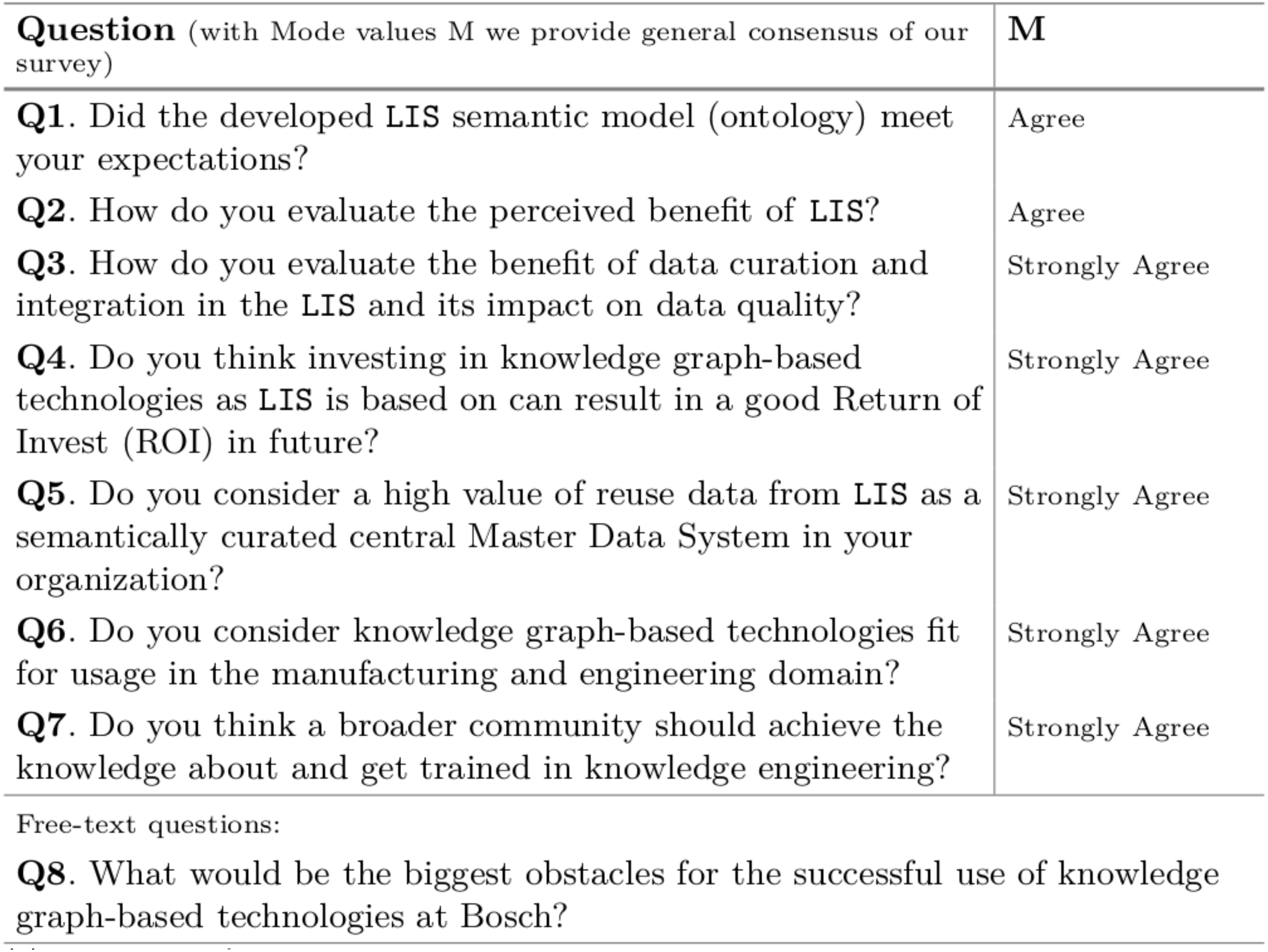
Correct use of OPTIONAL?

LIS ecosystem

- Web application to access semantically reconciled data
- Data sharing service (APIs) on top of the LIS KG
- Dashboards
  - to control data quality
  - interactive queries
- *LIS acts as a master data management system and a sharing procedure as well as a reporting system*



Preliminary evaluation



Question (with Mode values M we provide general consensus of our survey)	**M**
Q1. Did the developed LIS semantic model (ontology) meet your expectations?	Agree
Q2. How do you evaluate the perceived benefit of LIS?	Agree
Q3. How do you evaluate the benefit of data curation and integration in the LIS and its impact on data quality?	Strongly Agree
Q4. Do you think investing in knowledge graph-based technologies as LIS is based on can result in a good Return of Invest (ROI) in future?	Strongly Agree
Q5. Do you consider a high value of reuse data from LIS as a semantically curated central Master Data System in your organization?	Strongly Agree
Q6. Do you consider knowledge graph-based technologies fit for usage in the manufacturing and engineering domain?	Strongly Agree
Q7. Do you think a broader community should achieve the knowledge about and get trained in knowledge engineering?	Strongly Agree
Free-text questions:	
Q8. What would be the biggest obstacles for the successful use of knowledge graph-based technologies at Bosch?	

- Questionnaire
  - 8 questions
  - 5-point Likert-like scale
- 21 respondents from inside Bosch:
  - 7 managers
  - 7 developers
  - 7 users
- *Not a strong set-up...*

Next week:
Rules (SHACL and RDFS)