



Hochschule Karlsruhe – Data Engineering WS 2021/2022 DSCB330 – Vorlesung 2 – Ontologien & RDF



Git Repository

<https://github.com/thomasbierweiler/HKADataEngineering.git>

Vorlesung / Übung

Ab dem 14.10.2021 findet

- + die Vorlesung donnerstags, 08:00 Uhr, im Raum E002,
 - + die Übung donnerstags, 09:45 Uhr, im Raum E003
- statt.

Data Integration

- Data Formats (csv, XML, json)
- Extract, Transform, Load
- Object Relation Mapper (ORM)
- Staging

Data Processing

- Relationale Datenbanken
- nicht-relationale Datenbanken
- **Resource Description Framework (RDF)**
- **Ontologien**
- Data Warehouse

Data Modelling

- Serialisierung
- OPC UA
- MQTT
- Pub/Sub
- Data pipelines
 - Apache Airflow
 - gRPC

Web-Service Architektur

- Front-End
- Backend for Frontend (BFF)
- Micro Services
- Docker Container

Security

- Security ist wichtig in allen Phasen der Softwareentwicklung und Datenbereitstellung.

Übung

- Erstellung eines Daten-Modells einer prozesstechnischen Anlage
- Statische Daten
- Dynamische Daten
- Auswertung der Daten

- + Sprachlich gefasste und formal geordnete Darstellungen einer Menge von Begriffen und der zwischen ihnen bestehenden Beziehungen
- + Ontologien dienen als Mittel der Strukturierung und zum Datenaustausch

Bestandteile

- + Begriffe (concepts): Die Beschreibung gemeinsamer Eigenschaften wird als Begriff definiert. Begriffe werden auch als Klassen bezeichnet.
- + Typen repräsentieren Objekttypen.
- + Instanzen repräsentieren Objekte in der Ontologie.
- + Relationen beschreiben Beziehungen zwischen den Instanzen.
- + Es ist möglich, Relationen und Eigenschaften der Begriffe zu vererben.
- + Axiome sind Aussagen innerhalb der Ontologie, die immer wahr sind.

- + Standardisierung z.B. durch das W3C
- + Web Ontology Language (OWL)
- + ECLASS – Product Data for Engineering
- + ECLASS OWL
- + Asset Administration Shell (Verwaltungsschale)
- + VDI 2770 Digitale Herstellerinformationen
- + Struktur einer Ontologie ähnlich zu objekt-orientiertem Programmieren

Nutzen

- + Wissenstransfer zwischen verschiedenen Systemen

Schwierigkeiten

- + Begriffe sind nicht immer eindeutig definiert.
- + Begriffe sind nicht vorhanden oder können nicht vollständig definiert werden.

Lösungsansatz

- + Systeme handeln KI-gestützt die Ontologie selbst aus.

Quelle: [https://de.wikipedia.org/wiki/Ontologie_\(Informatik\)](https://de.wikipedia.org/wiki/Ontologie_(Informatik))

- + **Lightweight-Ontologien** beinhalten Begriffe, Taxonomien und Beziehungen zwischen Begriffen und Eigenschaften, welche diese beschreiben. Daher sind Lightweight-Ontologien typischerweise für eine bestimmte Anwendungsdomäne entworfen.
- + **Heavyweight-Ontologien** sind eine Erweiterung von Lightweight-Ontologien und fügen diesen Axiome und Einschränkungen hinzu, wodurch die beabsichtigte Bedeutung einzelner Aussagen innerhalb der Ontologie klarer wird.

Ontologie - Resource Description Framework - Anwendung



Ontologien

- + Web Ontology Language (OWL)
- Standardisierung durch W3C
- Nutzung z.B. in ECLASS OWL
- + F-Logic, DAML+OIL

Informationsübertragung

- + intelligent information Request and Delivery Standard (iiRDS)
- Standard für die Auslieferung von digitalen Anwenderinformationen
- Standardisierung in Zusammenarbeit mit dem VDI (Verein Deutscher Ingenieure), VDI 2770

Resource Description Framework (RDF)

- + technische Herangehensweise im Internet zur Formulierung logischer Aussagen über beliebige Dinge (Ressourcen)
- + Syntax für gemeinsamen Datenaustausch
- + W3C-Empfehlung



Quelle: <https://de.wikipedia.org/wiki/iiRDS/>; https://de.wikipedia.org/wiki/Resource_Description_Framework;
https://www.plattform-i40.de/IP/Redaktion/EN/Downloads/Publikation/vws-in-detail-presentation.pdf?__blob=publicationFile&v=7

Hochschule Karlsruhe | Data Engineering | DSCB330 | VL 2 | WS 2021/2022 | Dipl.-Phys. Thomas Bierweiler

07.10.2021

6

Resource Description Framework-Schema (RDFS)

- + RDF + gemeinsames Vokabular (Dublin Core)
- + W3C-Empfehlung

Asset Administration Shell

- + implements the Digital Twin
- + Definition of Digital Twin (Industrial Internet Consortium (IIC))
 - Digital representation, sufficient to meet the requirements of a set of use cases
- + Digital representation = information that represents characteristics and behaviors of an entity (asset) i.e. the Asset Administration Shell is the implementation of the Digital Twin for Industrie 4.0
- + Paket PyI40AAS, weiterführende Informationen im Ordner Literatur/81p_Heppner.pdf bzw. <https://git.rwth-aachen.de/acplt/pyi40aas>



Beispiel für die Nutzung eines standardisierten Datenaustauschs



NE 100 (NAMUR-Empfehlung NE 100)

Nutzung von Merkmalleisten im PLT-Engineering-Workflow

- + z.B. automatisierte Angebotsanfrage eines Kunden an Lieferanten durch Austausch von XML-Dateien.

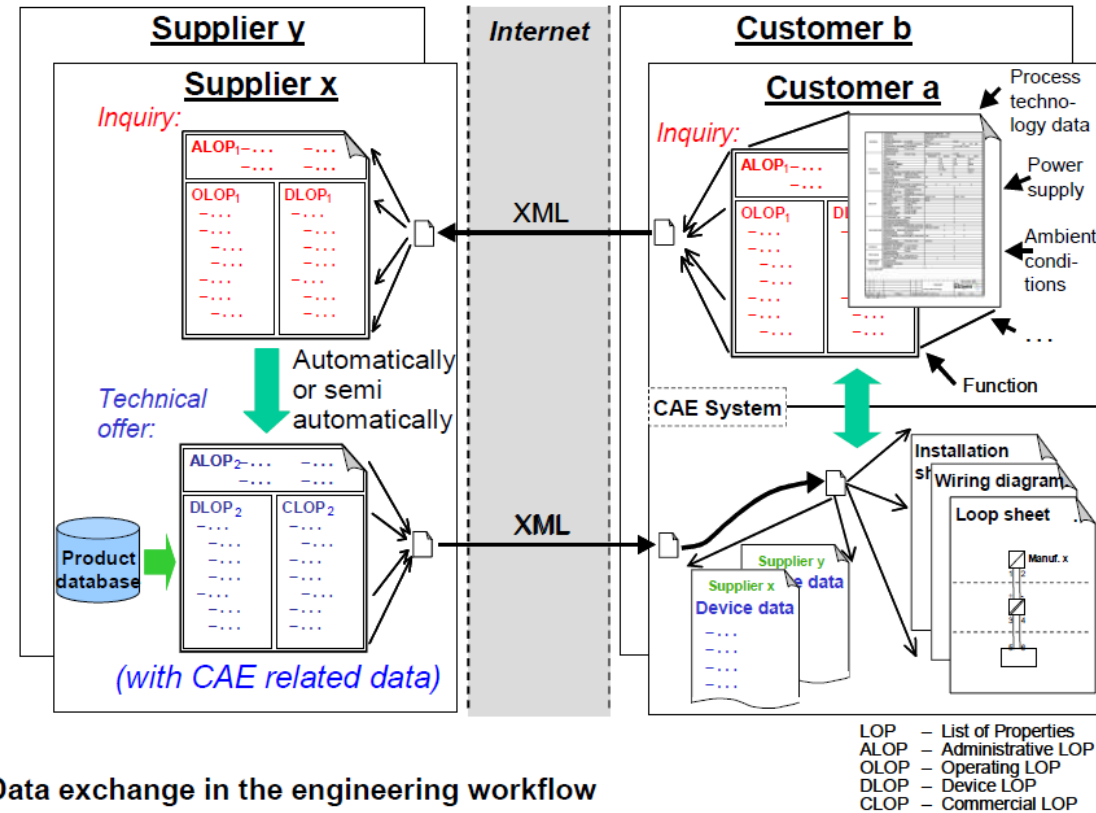


Fig. 9: Data exchange in the engineering workflow

Quelle: NAMUR-Empfehlung NE100, www.namur.net

Hochschule Karlsruhe | Data Engineering | DSCB330 | VL 2 | WS 2021/2022 | Dipl.-Phys. Thomas Bierweiler

07.10.2021

Struktur

Composition of components	
Product identifier	
Number of components	1
Component	
Application Class IRDI	
Product identifier	
Manufacturer	
Link to product description	
Number of relations	1
Relation	
Type of relation	
Port A of relation	
Product identifier	
Port identifier	
Port B of relation	
Product identifier	
Port identifier	

Number of components

Component

Number of relations

Relation between product and component

Beispiel für eine zusammengesetzte Komponente



ID MP

27-37-04-24 Wiring set for circuit breaker (Advanced AC : 0173-1#01-ADO384#008)
 27-37-04-01 Motor protection circuit-breaker (Advanced AC : 0173-1#01-ADN710#008)
 27-37-10-03 Power contactor, AC switching (Advanced AC : 0173-1#01-ADO120#008)
 27-37-09-05 Motor starter combination (Advanced AC : 0173-1#01-ADN843#008)

Template:	
LOAD FEEDER .DS S00 0,14-0,2A, DC24V	
BMEcat attributes	
Composition of components	
Product identifier	MP
Number of components	3
Component_1	
Application Class IRD1	0173-1---ADVANCED_1_1#01-ADN710#008
Product identifier	CB
Manufacturer	
Link to product description	../
Component_2	
Application Class IRD1	0173-1---ADVANCED_1_1#01-ADO384#008
Product identifier	LM
Manufacturer	
Link to product description	../
Component_3	
Application Class IRD1	0173-1---ADVANCED_1_1#01-ADO120#008
Product identifier	PC
Manufacturer	
Link to product description	../
Number of relations	20
Relation_1	
Relation_2	

Parts ID



CB



LM



PC

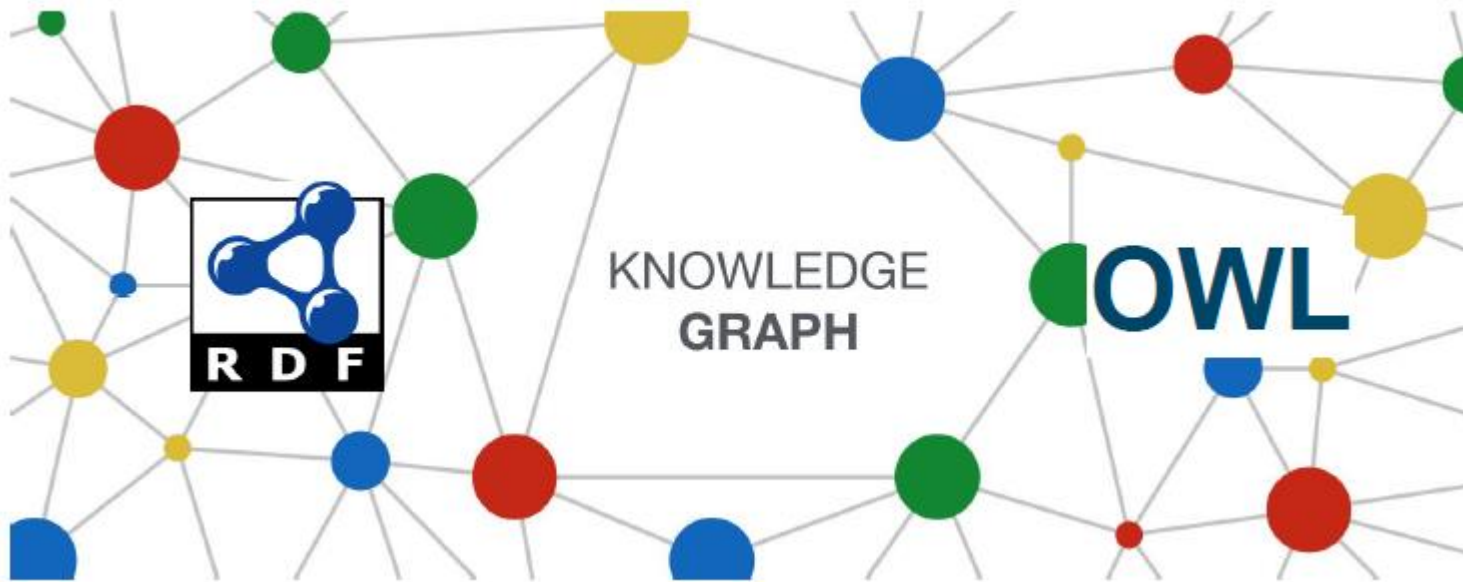
Number of relations	20
Relation_1	
Type of relation	Connection of connectors
Port A of relation	
Product identifier	CB
Port identifier	2
Port B of relation	
Product identifier	LM
Port identifier	1
Relation_2	
Relation_3	
Relation_4	
Relation_5	
Relation_6	

Semantic representation of ECLASS



Challenges of Product Data

- + Historically grown data bases, created for a specific use case
- + Data sources are static, adding new requirements is difficult
- + Not all requirements referring to product description are known yet
- + Dynamically adding and linking new data is essential



Semantic representation of ECLASS



ECLASS model and exchange format

- + OntoML
 - ECLASS model
 - XML-based exchange structure
 - Standard ISO 13584-32
-
- + BMEcat
 - standardisiertes Austauschformat für Katalogdaten
 - ECLASS instances (catalogue data)
 - XML-Schema
 - Reference to IRDIs of OntoML model
 - Last version 2005

eClassOWL - The Web Ontology for Products and Services

- + Research prototype, not an official release
- + Available formats: RDF/XML/HTML



Quelle: <https://www.eclass.eu/>; <http://www.heppnetz.de/projects/eclassowl/>

Hochschule Karlsruhe | Data Engineering | DSCB330 | VL 2 | WS 2021/2022 | Dipl.-Phys. Thomas Bierweiler

07.10.2021

11



Aufbau

- + siehe Datei 07-eClassOWL/eclass_514en.zip
- HTML-Version: eclass_514en.html
- Classes/Properties: eclassClassesProperties_514en.owl



```
<owl:Class rdf:ID="C_AAA186004">
  <eco:recommendedProperty rdf:resource="#P_BAA001001"/>
  <eco:recommendedProperty rdf:resource="#P_BAA059001"/>
  <eco:recommendedProperty rdf:resource="#P_BAA271001"/>
  <eco:recommendedProperty rdf:resource="#P_BAA316001"/>
```

- Properties/Values: eclassPropertiesValues_514en.owl

```
<owl:ObjectProperty rdf:ID="P_BAA007001">
  <eco:recommendedValue rdf:resource="#V_BAB122001"/>
  <eco:recommendedValue rdf:resource="#V_BAB168001"/>
  <eco:recommendedValue rdf:resource="#V_WAA090001"/>
</owl:ObjectProperty>
```

- Begriffe: eclass_514en.owl

```
<owl:Class rdf:ID="C_BAA812001-gen">
  <rdfs:label xml:lang="en">Flow meter (magnetic inductive probe) [Generic Concept: This type of goods]</rdfs:label>
  <rdfs:comment xml:lang="en">This class subsumes all actual instances of the following type of goods and true species</rdfs:comment>
  <rdfs:subClassOf rdf:resource="#C_BAA812001-tax"/>
  <rdfs:subClassOf rdf:resource="#&gr;#ProductOrService"/>
</owl:Class>
```

Quelle: <https://www.eclass.eu/>; <http://www.heppnetz.de/projects/eclassowl/>

Hochschule Karlsruhe | Data Engineering | DSCB330 | VL 2 | WS 2021/2022 | Dipl.-Phys. Thomas Bierweiler

07.10.2021

12

SPARQL - Abfrage von Informationen aus RDF



SPARQL Query Language for RDF

- + <https://www.w3.org/TR/sparql11-overview/>

Abfrage von Inhalten aus einem Triple-Store (RDF)

foaf: friend of a friend

link to FOAF Vocabulary Specification

PREFIX foaf: <<http://xmlns.com/foaf/0.1/>>

SELECT ?name

WHERE {

 ?p rdf:type foaf:Person .

 ?p foaf:name ?name .

}

Tutorial

- + <https://cambridgesemantics.com/blog/semantic-university/learn-sparql/sparql-by-example/>

Übung 8 SPARQL

- + <https://www.dbpedia.org/resources/sparql/>
- + <https://dbpedia.org/snorql/>

Wikipedia als RDF: Wikidata

- + https://www.wikidata.org/wiki/Wikidata:Main_Page
- + Zugangspunkt für SPARQL auf wikidata
<https://query.wikidata.org>

Quelle: <https://rdflib.readthedocs.io/en/stable/gettingstarted.html>

Hochschule Karlsruhe | Data Engineering | DSCB330 | VL 2 | WS 2021/2022 | Dipl.-Phys. Thomas Bierweiler

07.10.2021

13

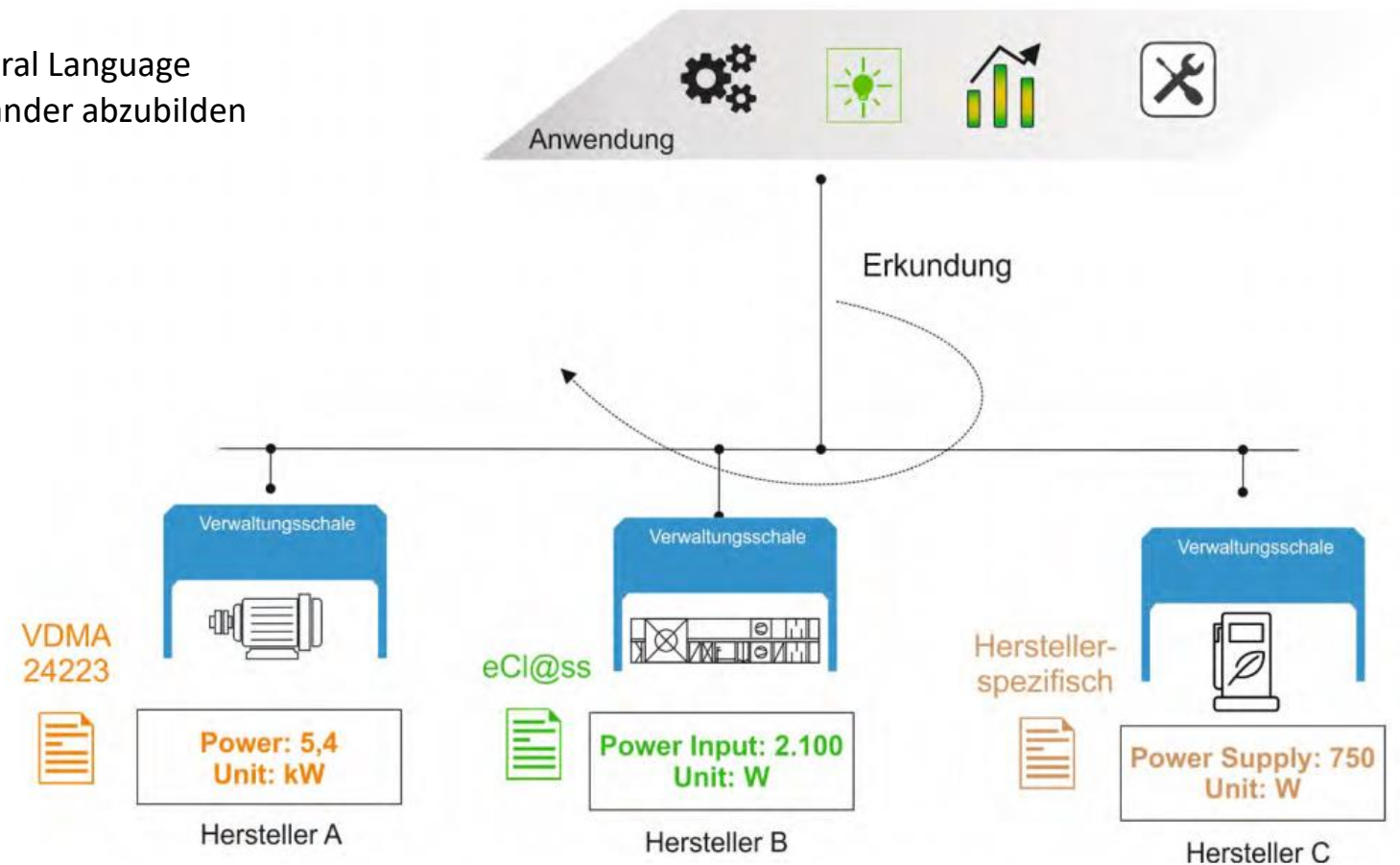


Ontologien – Semantische Interoperabilität durch Natural Language Processing



Beispiel heterogener Semantik

- + Einsatz von Methoden aus dem Bereich Natural Language Processing, um heterogene Semantik aufeinander abzubilden



Quelle: Maximilian Both, Dr. Jochen Müller, Semantische Interoperabilität durch Natural Language Processing als Basis für Self-X-Fähigkeiten von Verwaltungsschalen in semantisch heterogenen Asset-Netzwerken, Kongress Automation 2021; Datei Literatur/53p_Both.pdf

Hochschule Karlsruhe | Data Engineering | DSCB330 | VL 2 | WS 2021/2022 | Dipl.-Phys. Thomas Bierweiler

07.10.2021

14

Übungsaufgabe 3



Rubens Lehrer

- + Erstellen Sie in python ein SPARQL-Query, das die Lehrer des flämischen Malers Peter Paul Rubens abfragt.
Nutzen Sie dbpedia ("<http://dbpedia.org/sparql>") als Endpoint.



Übungsaufgabe 4



Graph Database

- + Erstellen Sie z.B. mit SQL Server Management Studio eine Graphdatenbank, die den rötlich hinterlegten Teil des Fließbilds abbildet.
- + Führen Sie dazu die SQL-Skripte in Uebung\04-Schema-der-Anlage-Graphdatabase aus.
- + Die Flüssigkeit fließt aus dem Behälter VE1000 durch das Ventil YC10001 durch die beiden Pumpen PL1100 und PL1200 und kann dann bei geöffnetem Ventil YS14005 durch die Leitung R10 in den Behälter VE1000 zurückfließen.
- + Erstellen Sie eine Abfrage (in SQL oder als Kombination von python und SQL), die als Ergebnis die beiden möglichen Wege vom Behälter VE1000 zum T-Stück T5 zurückgibt.

