



①
**Erstellung einer Graph-
Datenbank zur Verwaltung der
② identifizierbaren Konzept-
beschreibungen für eindeutige ③
Semantik von Datenmerkmalen
in der Automatisierungstechnik**

Abschlussvortrag

Zhewen Xiao

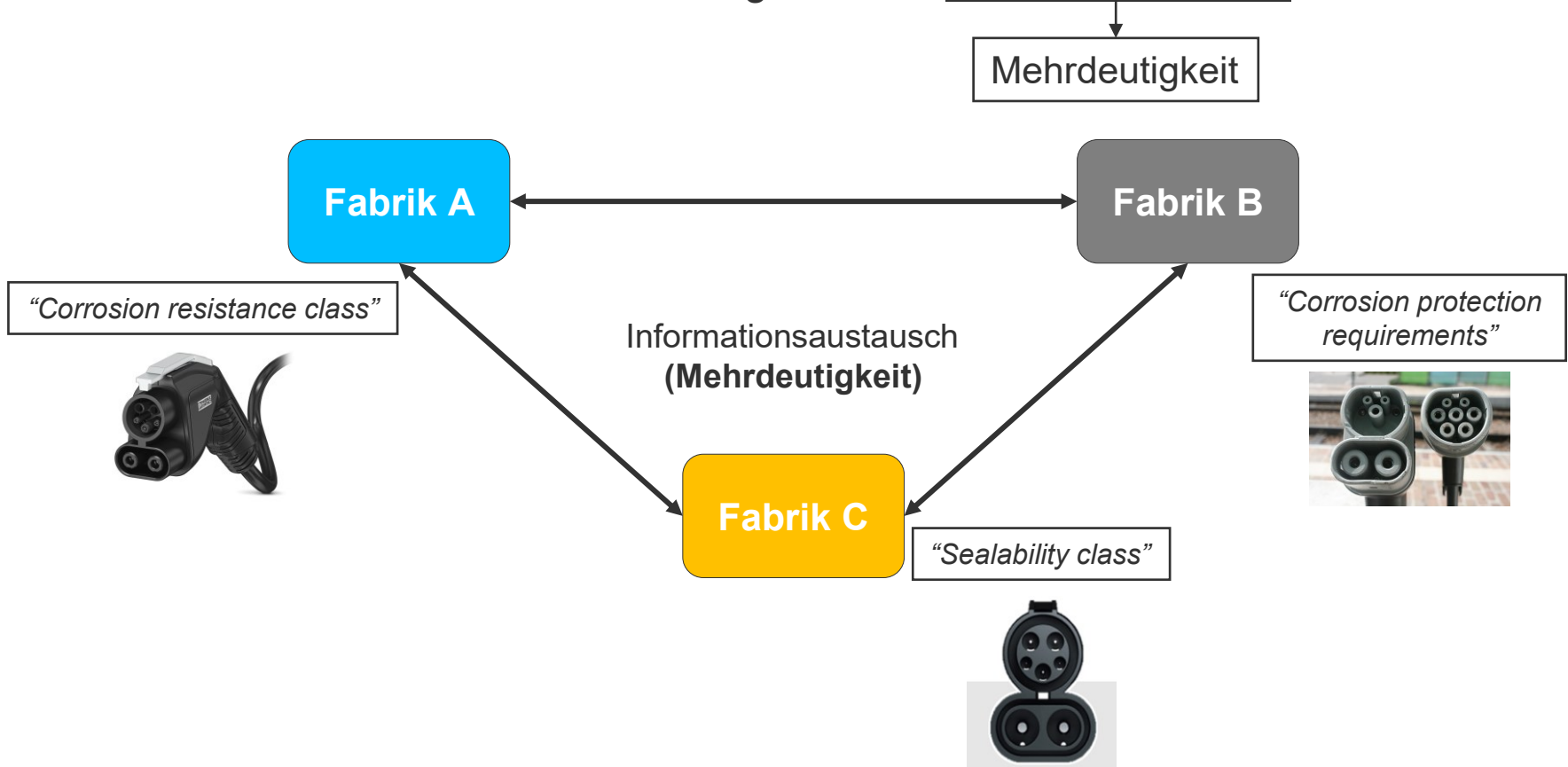
Studiengang: EIT

Supervisor: Yuchen Xia

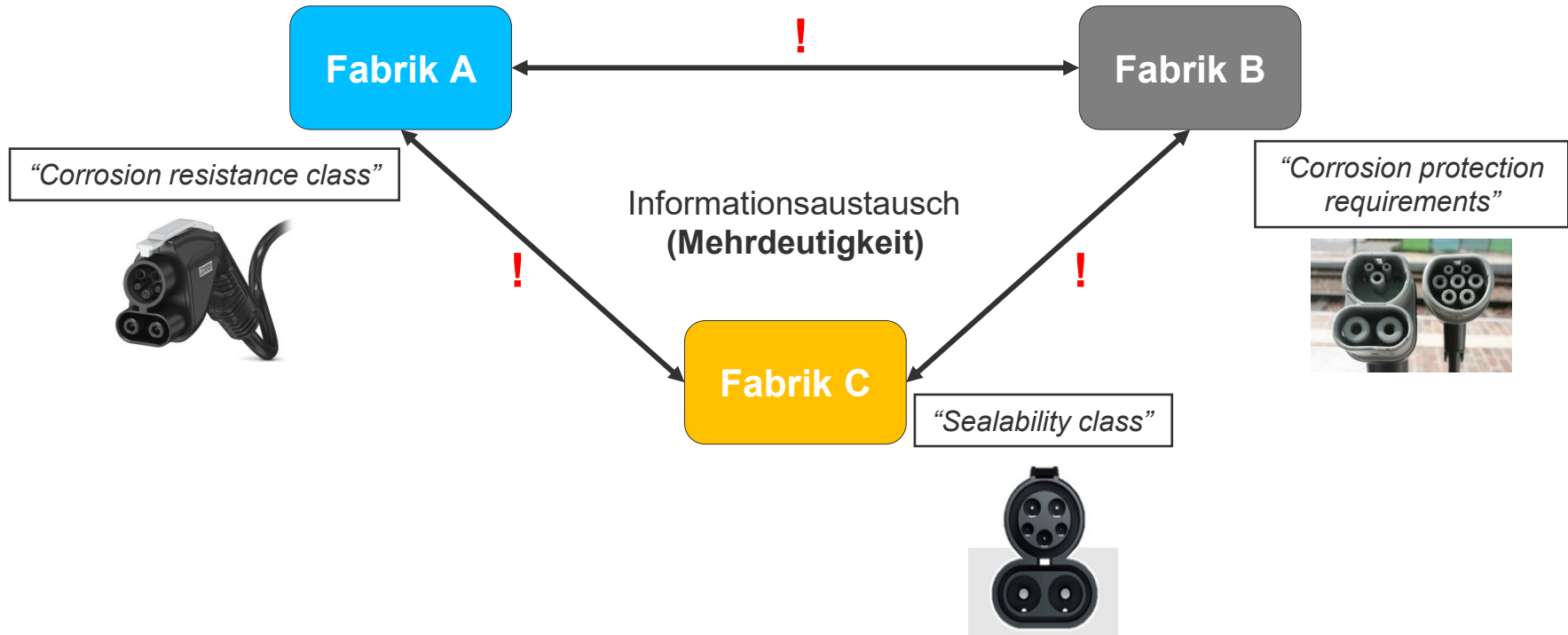


Motivation

Austauschbarkeit der Information hängt von der Qualität der Daten ab.



Problem



Wegen der **Mehrdeutigkeit der Informationen** ist der Produktinformations- bzw. Datenaustausch zwischen Unternehmen nicht effizient.

Anforderungen

- Datenmerkmalen sollten global eindeutig definiert werden.
- Bedeutung von Datenmerkmalen sollten mit maschinenlesbaren expliziten Meta-Informationen beschrieben werden. (Annotation)
- Sammlung einer großen Menge von Metadaten.
- Datenverwaltung in einer Graph-Datenbank.

Grundlagen und Ressourcen

Semantische Meta-Daten

Anforderungen

- Datenmerkmalen sollten global eindeutig definiert werden.
- Bedeutung von Datenmerkmalen sollten mit maschinenlesbaren expliziten Meta-Informationen beschrieben werden. (Annotation)
- Sammlung einer großen Menge von Metadaten.
- Datenverwaltung in einer Graph-Datenbank.

- Existierende Standards



- Wiederverwendbare Ontologie-Modelle

MASON

SOSA

AML

SMO

SSN

.....

Datenbereinigung und Import

Graph-Datenbank



MASON (MAnufacturing's Semantics ONtology)
SOSA (Sensor, Observation, Sample, and Actuator)
AML (AutomationML)
SMO (Smart-Manufacturing-Ontology)
SSN (Semantic Sensor Network Ontology)

Grundlagen und Ressourcen

Semantische Annotation der Daten



Im Allgemeinen:



Vergleich semantischer Meta-Daten

Evaluierung gesammelter Daten in vier Dimensionen:

Quelle	Inhalt	Menge	Wiederverwendbarkeit	Identifizier
ECLASS	+	+	+	+
OPC UA	+	+	+	-
STO	+	+	+	+
MASON	+	-	+	+
AML	o	o	+	+
I40KG	+	+	+	+
RAMI	+	o	+	+
SOSA	o	-	+	+
SMO	+	o	o	+
Production_Systems	+	-	o	+
Production_Systems_ssn	+	-	+	+
+ : erfüllt o : bedingt erfüllt - : nicht erfüllt				

Datenimport

neo4j\$ MATCH (n) RETURN n LIMIT 100

Node Properties

- Resource
- owl_Class
- <id> 85
- rdfs_isID DIN 8550
- efinedBy
- rdfs_lab Manufacturing Process
- uri http://ontologies.ift.at/production_systems.ttl/0.4#ManufacturingProcess

Joining $\xrightarrow{\text{subClassOf}}$ Manufacturing Process

Node Properties

- Resource
- owl_Class
- <id> 71
- rdfs_isID DIN 8550
- efinedBy
- rdfs_lab Joining
- el
- uri http://ontologies.ift.at/production_systems.ttl/0.4#Joining

Datenbereinigung

```
1 MATCH (n) WHERE EXISTS(n.rdfs_comment)
2 RETURN DISTINCT n.rdfs_label as rdfs_label, n.rdfs_comment AS rdfs_comment
```

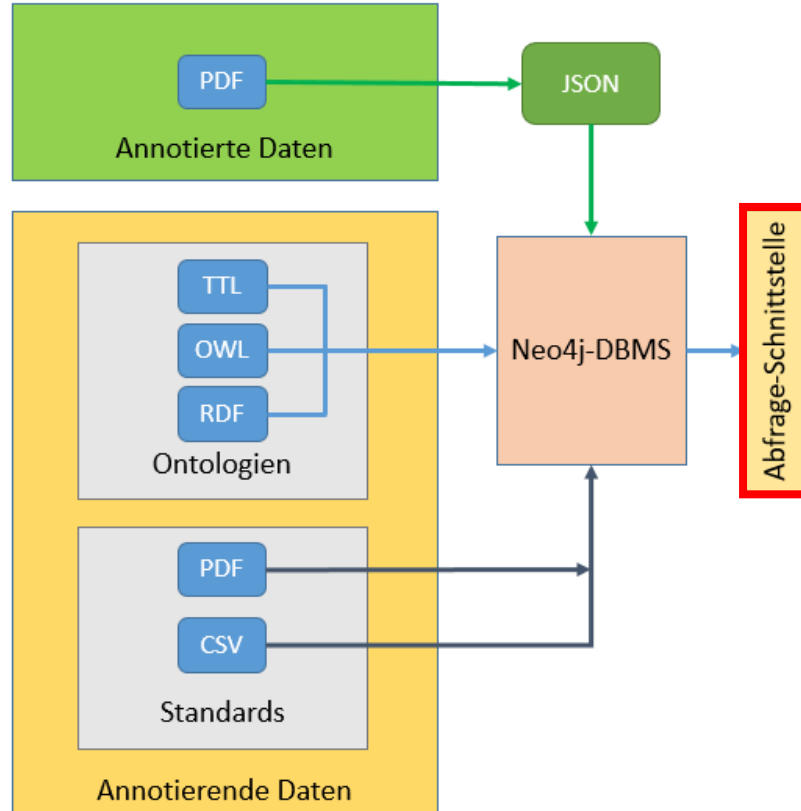
	rdfs_label	rdfs_comment
1	"has Implementation Format"	"Describes the name of the implementation or serialization format of the Publication, e.g., XML, Document, binary, batch"
2	"DataRecovery"	""
3	"FIWARE Cloud Hosting"	"The Cloud Chapter offers Generic Enablers (GEs) that comprise the foundation for designing a modern cloud hosting infrastructure that can be used to de The IaaS Data Center Resource Management (DCRM) GE is offering provisioning and life cycle management of virtualized resources (compute, storage, ..."
4	"Description"	"An account of the resource."
5	"Open Digital Rights Language"	"The Open Digital Rights Language (ODRL) is a policy expression language that provides a flexible and interoperable information model, vocabulary, and ... Policies are used to represent permitted and prohibited actions over a certain asset, as well as the obligations required to be meet by stakeholders. In add
6	"Reactive Streams"	"A standard for asynchronous stream processing with non-blocking back pressure on the JVM. (source: https://www.postscapes.com/internet-of-things-pro)"

Started streaming 333 records after 1 ms and completed after 3 ms.

Die Daten werden zunächst als CSV-Datei aus der Datenbank exportiert und dann bereinigt. (Theoretisch würde die Bereinigung über Cypher-Anweisungen in der Graph-Datenbank auch funktionieren, allerdings mit höherem Arbeitsaufwand und den gleichen Ergebnissen.)

Daten-Pipeline

Nach Verarbeitung der Daten (Sammlung, Bereinigung und Import) ergibt sich folgende Daten-Pipeline:



Datenabfragen

Abfrage durch *Cypher*-Anweisungen:

Bsp.1: Abfrage nach Namen des Attributes:

```
MATCH (n) WHERE EXISTS(n.rdfs__comment)
RETURN DISTINCT n.rdfs__label AS rdfs_label, n.rdfs__comment AS rdfs__comment
```



rdfs_label	rdfs__comment
"has Implementation Format"	"Describes the name of the implementation or serialization format of the Publication, e.g., XML, Document, binary, batch"
"FIWARE Cloud Hosting"	"The Cloud Chapter offers Generic Enablers (GEs) that comprise the foundation for designing a modern cloud hosting infrastructure that can be used to de The IaaS Data Center Resource Management (DCRM) GE is offering provisioning and life cycle management of virtualized resources (compute, storage,
"Description"	"An account of the resource."
"Open Digital Rights Language"	"The Open Digital Rights Language (ODRL) is a policy expression language that provides a flexible and interoperable information model, vocabulary, and Policies are used to represent permitted and prohibited actions over a certain asset, as well as the obligations required to be met by stakeholders. In add
"Reactive Streams"	"A standard for asynchronous stream processing with non-blocking back pressure on the JVM. (source: https://www.postscapecs.com/internet-of-things-pro)"

Bsp.2: „Fuzzy-Search“:

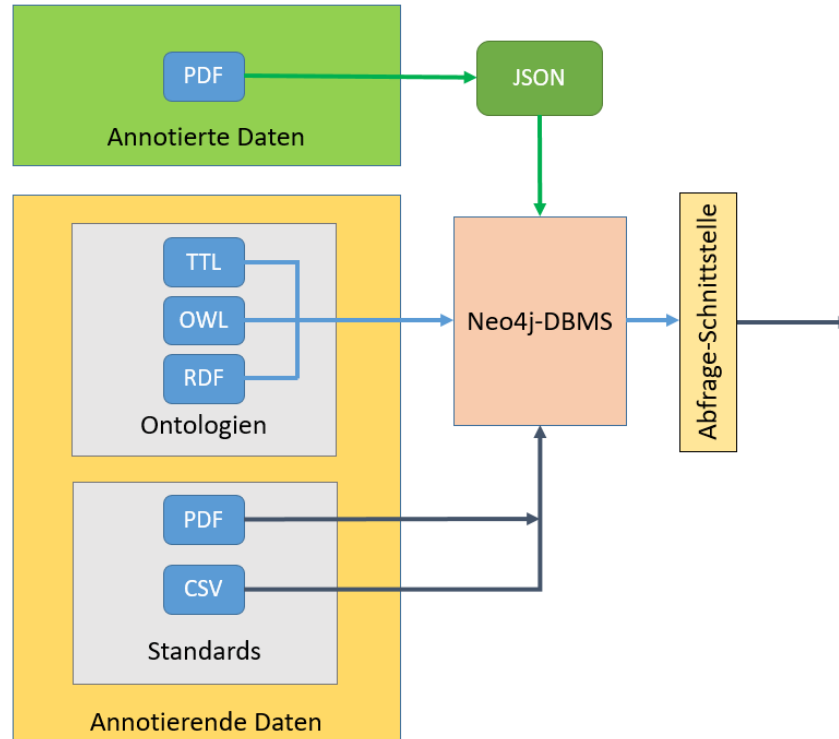
```
MATCH (n) WHERE n.Definition =~ '*Voltage.*'
RETURN n
```

oder

```
MATCH (n) WHERE n.Label STARTS WITH 'Electrical'
RETURN n
```

Hilfsfunktionen der Datenbank

Nachdem die Graph-Datenbank erstellt wurde, können die gespeicherten Meta-Daten verwendet und die entsprechenden Hilfsfunktionen entwickelt werden.



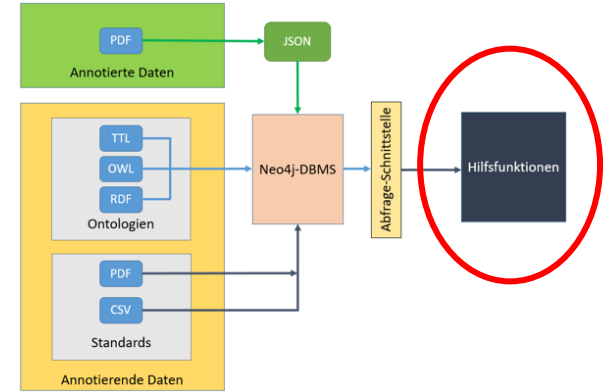
Hilfsfunktionen der Datenbank

Zur besseren Verwaltung der Datenbank:

Wie kann die Abfrage von semantischen Daten

effizienter gestaltet werden?

Wie kann die Abfrage von semantischen Daten intuitiver gestaltet werden?

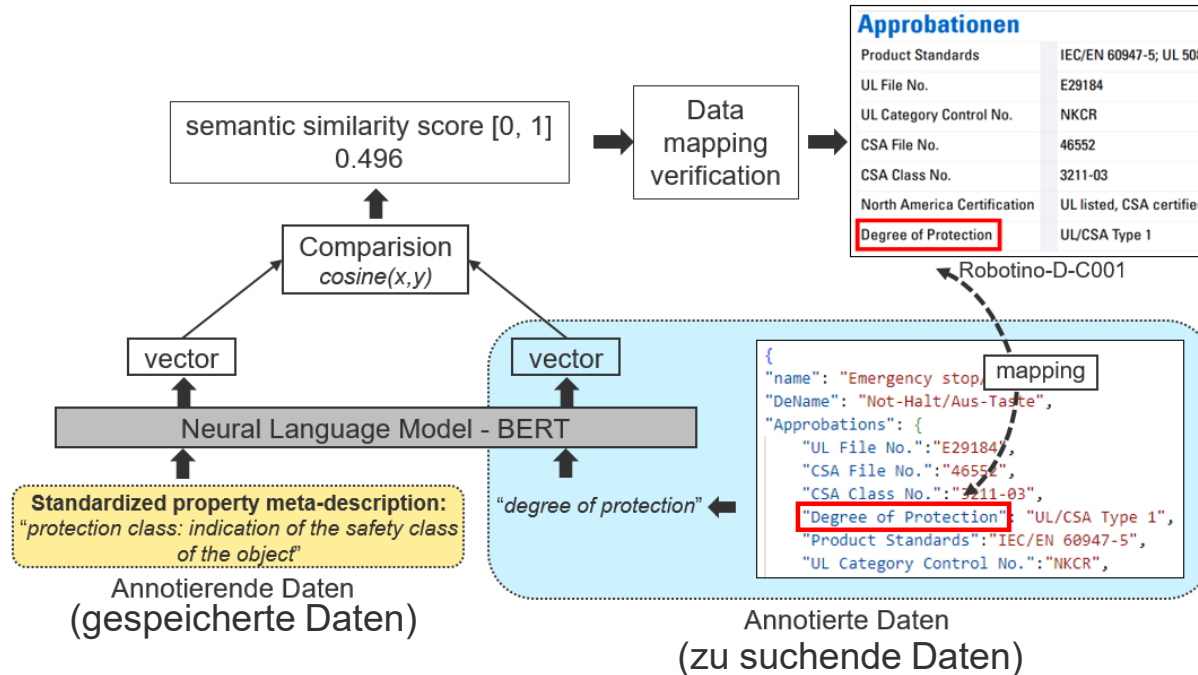


1. Semantische Suche
2. Visualisierung der Suchergebnisse
3. K-Means Clustering

Hilfsfunktionen der Datenbank (1/3)

1. Funktion „Semantische Suche“

Für die Implementierung dieser Hilfsfunktion werden **Indexing** benötigt.



Hilfsfunktionen der Datenbank (1/3)

1. Funktion „Semantische Suche“: Query vs. Semantische Suche

Vorher: Abfrage durch *Cypher*-Anweisungen:

```
MATCH (n) WHERE n.Definition=~'.*Corrosion.*'
RETURN N
```



Über **200** Datenelemente, die das Schlüsselwort „Corrosion“ in der Definition enthalten anstatt semantischer Suchergebnisse.

Nachher: Semantische Suche mit „Suchmaschine“, Output: (Corrosion)

Top 1	0.6519359946250916	<u>Corrosion resistance</u>	Protected against the dec
Top 2	0.5660135746002197	<u>Sealing</u>	Additional organic protection layer to
...	0.5135456919670105	<u>corrosion protection</u>	protection from the react
	0.5109236836433411	<u>Coating metal</u>	Material which is applied as a c
	0.4966057538986206	<u>Corrosive effect on copper</u>	reaction of the pro
	0.49139726161956787	<u>Surface protection</u>	Type of measures for achie
	0.47575104236602783	<u>Grease packing against corrosion present</u>	whet
	0.4741849899291992	<u>corrosion protection requirements present</u>	spec
	0.4672285318374634	<u>auxiliary device fault - corrosion</u>	auxiliary d

Hilfsfunktionen der Datenbank (1/3)

1. Funktion „Semantische Suche“: Evaluierung der Genauigkeit

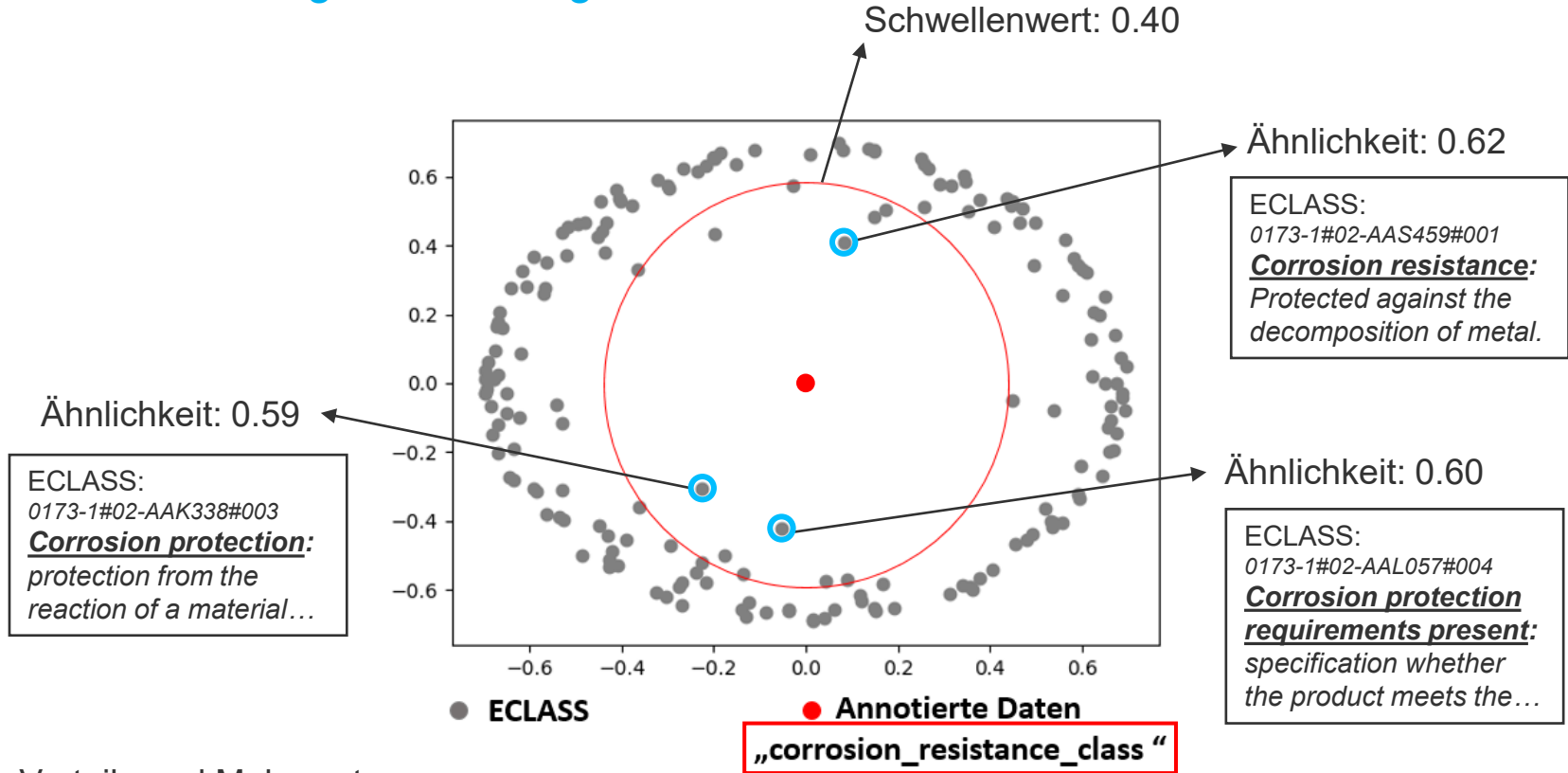
Annotierte Daten	Anzahl <u>relevanter</u> Einträge von Top-20 Output	Genauigkeit
serial_number	16	80.0%
measurement_method	17	85.0%
corrosion_resistance_class	16	80.0%
degree_of_protection	19	95.0%
test_conditions	15	75.0%
ambient_temperature	19	95.0%
degree_of_pollution	15	75.0%
max_output_current	16	80.0%
product_weight	18	90.0%
max_switching_frequency	17	85.0%

$$\frac{16}{20} = 80.0\%$$

Durchschnittliche Genauigkeit: **84.0%**

Hilfsfunktionen der Datenbank (2/3)

2. Visualisierung der Suchergebnisse



Vorteile und Mehrwert:

Intuitiver zu schauen, wie viele Punkte und welche Punkte relevant sind.

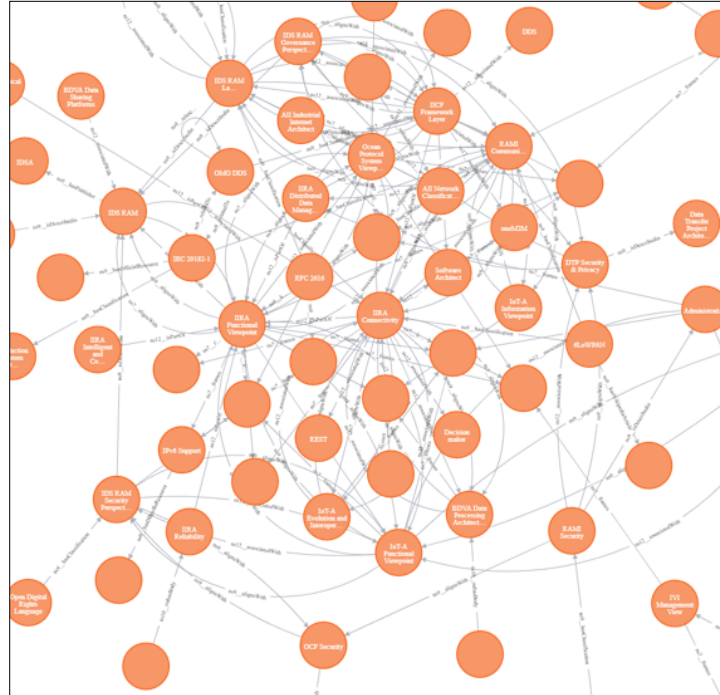
Hilfsfunktionen der Datenbank (3/3)

3. K-Means Clustering

Die in der Graph-Datenbank gespeicherten Meta-Daten: ungeordnet, zahlreichen Korrelationen.

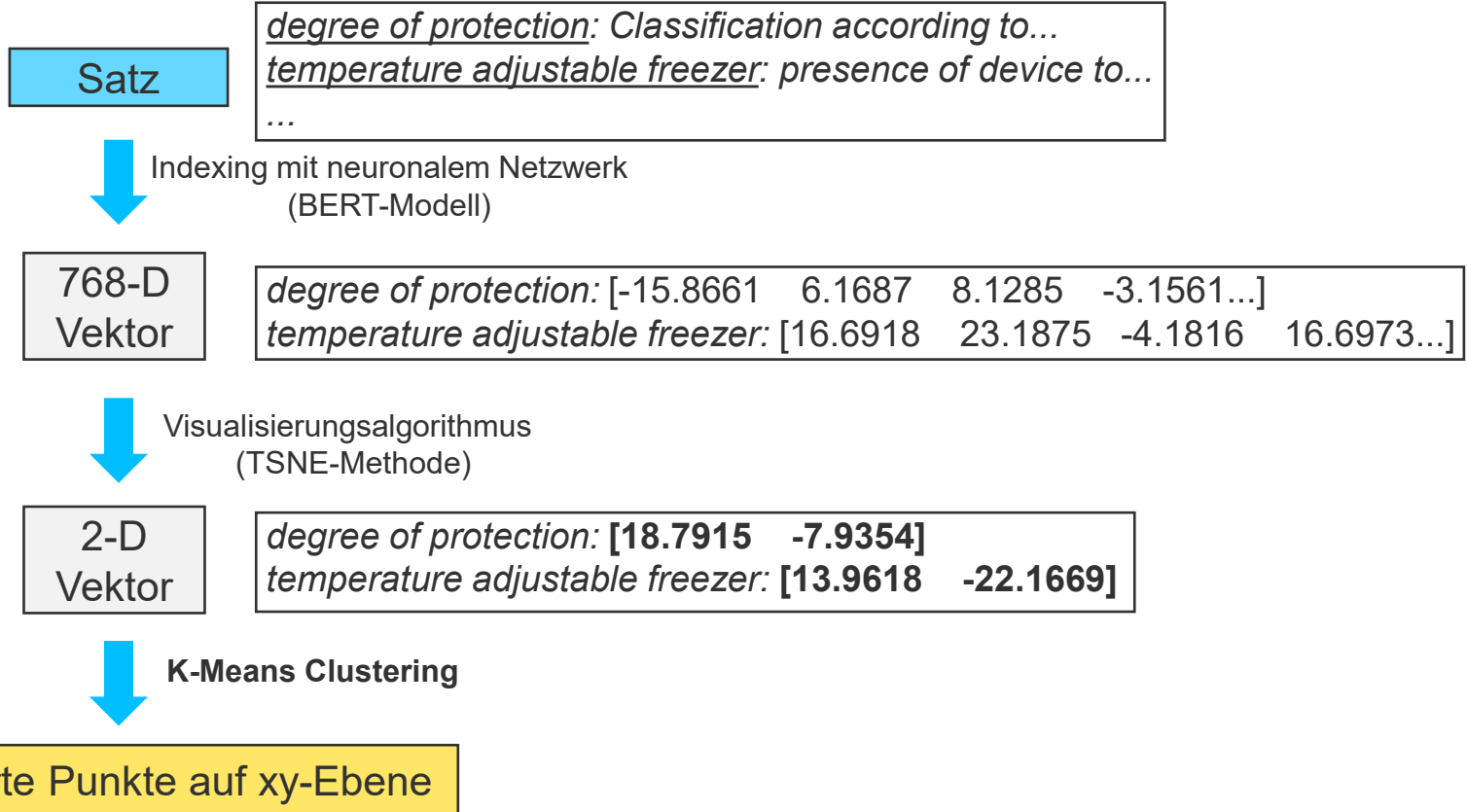
Anforderung: Gruppierung von Daten nach ihren semantischen Ähnlichkeiten.

Ziel: Globalen Überblick von allen gespeicherte Daten und ihre Gruppierung zu erhalten.



Hilfsfunktionen der Datenbank (3/3)

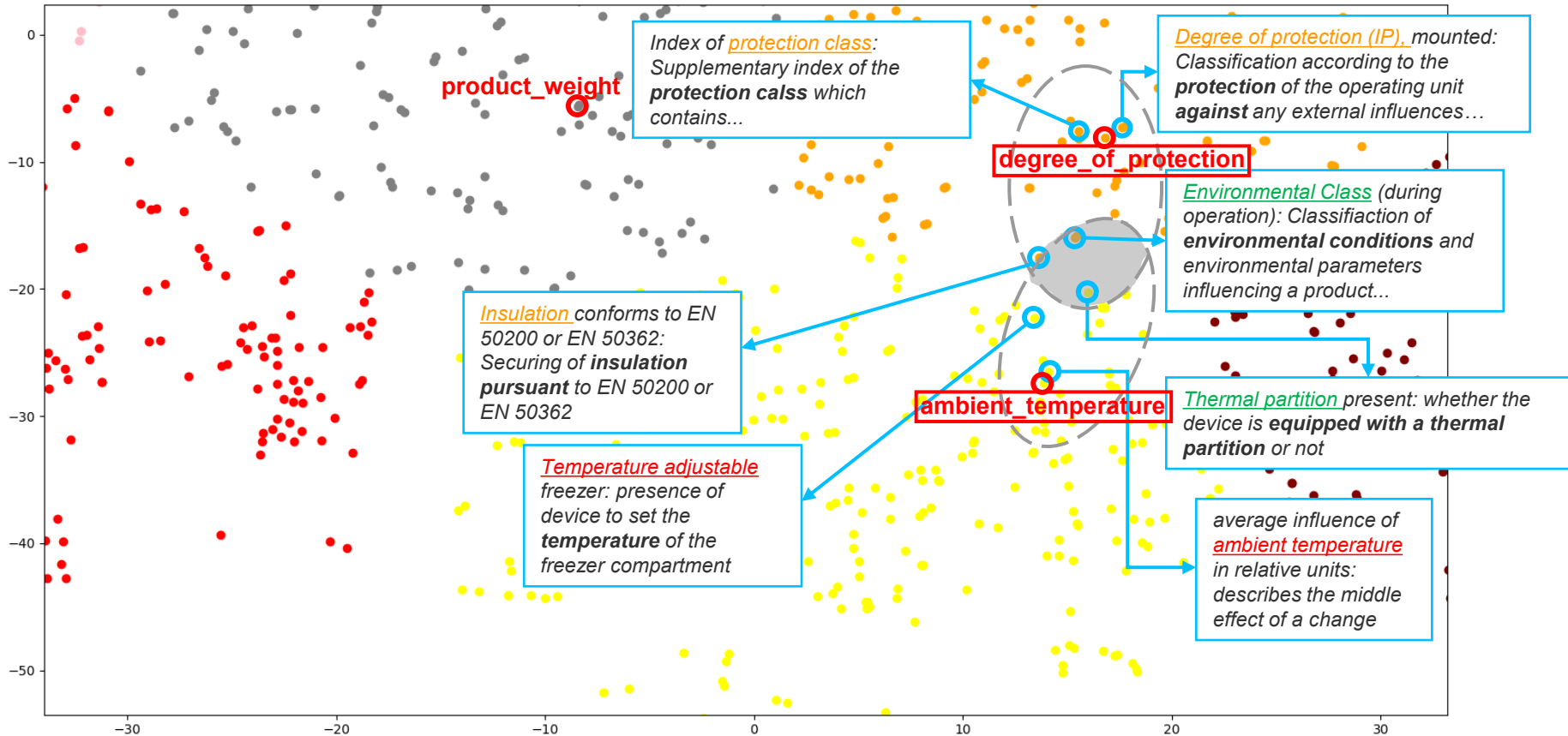
3. K-Means Clustering



(TSNE: T-distributed Stochastic Neighbor Embedding)

Hilfsfunktionen der Datenbank (3/3)

3. K-Means Clustering



Zusammenfassung und Ausblick

- Folgende Ergebnisse wurden durch die Studie erzielt:
 - **23.172** annotierende Daten aus *ECLASS*, *OPC UA* und Ontologiemodellen wie *I40KG*, *AML*, *SOSA* usw. sowie **131** annotierte Daten aus herstellerspezifischen Datenblättern wurden gesammelt und bereinigt.
 - Eine **Neo4j-Graph-Datenbank** wurde aufgebaut und die gesammelten Daten in der Datenbank wurden gespeichert und verwaltet.
 - **3 Hilfsfunktionen** der Datenbank wurden entwickelt, um die Verwaltung von semantischen Meta-Daten zu verbessern und vereinfachen:
 - **Funktion „Semantische Suche“**
 - **Visualisierung der Suchergebnisse**
 - **K-Means Clustering**
 - Diese Forschungsarbeit bildet eine Grundlage für die künftige semantische Annotation der Daten im Bereich der Automatisierungstechnik.



Universität Stuttgart

Vielen Dank!



Zhewen Xiao

E-Mail st165303@stud.uni-stuttgart.de

Telefon +49 (0) 711 685-

Fax +49 (0) 711 685-

Universität Stuttgart

