

TIB

Requirements Analysis for an Open Research Knowledge Graph

Arthur Brack¹, Anett Hoppe¹, Markus Stocker¹
Sören Auer^{1,2}, and Ralph Ewerth^{1,2}

¹ TIB Leibniz Information Centre for Science and Technology, Hannover, Germany

² L3S Research Center, Leibniz University Hannover, Germany

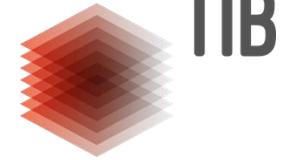
26th August 2020, TPDFL 2020 (online)

Motivation

Work of scientists with more than 2 mio. publications per year [12] ...



Open Research Knowledge Graph (ORKG)

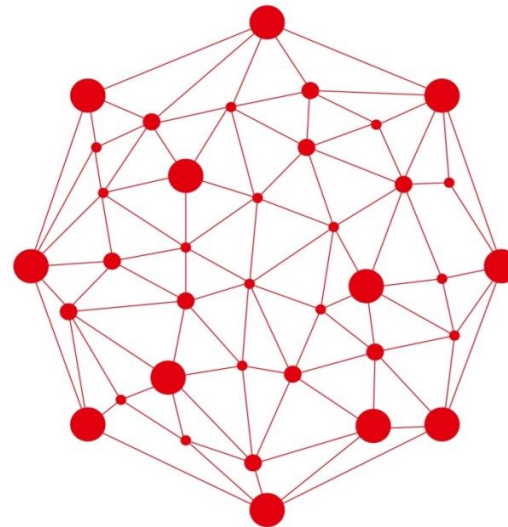


Overarching Concepts

- Research problems
- Definitions
- Research approaches
- Methods

Artefacts

- Publications
- Data
- Software
- Image/Audio/Video
- Knowledge Graphs / Ontologies

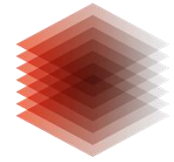


Open Research Knowledge Graph makes **comprehensive** and **subject-specific concepts** clearly identifiable and links them semantically (with **clearly described relations**) with each other and with relevant further artefacts.

Domain specific Concepts

Mathematics	Physics	Chemistry	Computer Science	Technology	Architecture
<ul style="list-style-type: none">• Definitions• Theorems• Proofs• Methods• ...	<ul style="list-style-type: none">• Experiments• Data• Models• ...	<ul style="list-style-type: none">• Substances• Structures• Reactions• ...	<ul style="list-style-type: none">• Tasks• Concepts• Implementations• Evaluations• ...	<ul style="list-style-type: none">• Standards• Processes• Elements• Units, Sensor data	<ul style="list-style-type: none">• Regulations• Elements• Models• ...

Problem Statement

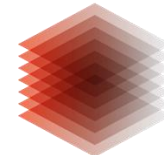


How to populate such a KG?

We need a **comprehensive** KG with **domain-specific** concepts, **high granularity**, **high coverage** and **high quality!**

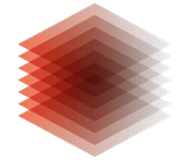


Research Questions



- Approach: Requirements Analysis following Design Science Research [33]
- Research questions in this paper:
 - RQ1: Which **functionalities** should be provided by the ORKG system?
 - RQ2: What kind of data (**ontologies**) is necessary for these functionalities?
 - a) Degree of **domain specialisation**
 - b) **Granularity** of information representation
 - RQ3: Which **coverage** and **quality** of **instance data** is sufficient?
 - RQ4: Which **approaches** are suitable **to populate** the ORKG?

Outline



Related work

- Existing research KGs
- Ontologies representing scholarly knowledge
- Approaches for KG construction (automatic and manual)

Elicit requirements

- Interviewing ORKG group members
- Literature review about systematic literature reviews

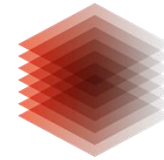
Define requirements

- Functional requirements (use cases)
- KG requirements (ontology and instance data) for the use cases

Elaborate approaches

- Manual and automatic approaches for KG construction

Related Work



Existing Research KGs

- Citation graphs with metadata (e.g. Microsoft Academic KG [24])
- Interlink papers with artefacts (e.g. Research Objects [7])
- Domain-specific KGs (e.g. www.paperswithcode.com)

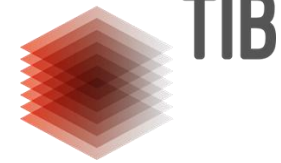
Scientific Ontologies

- Focus on structure (rhetorical [66], argumentative [63], activities [45])
- Focus on domain-specific [40,61] and domain-independent [5,13] entities and relations

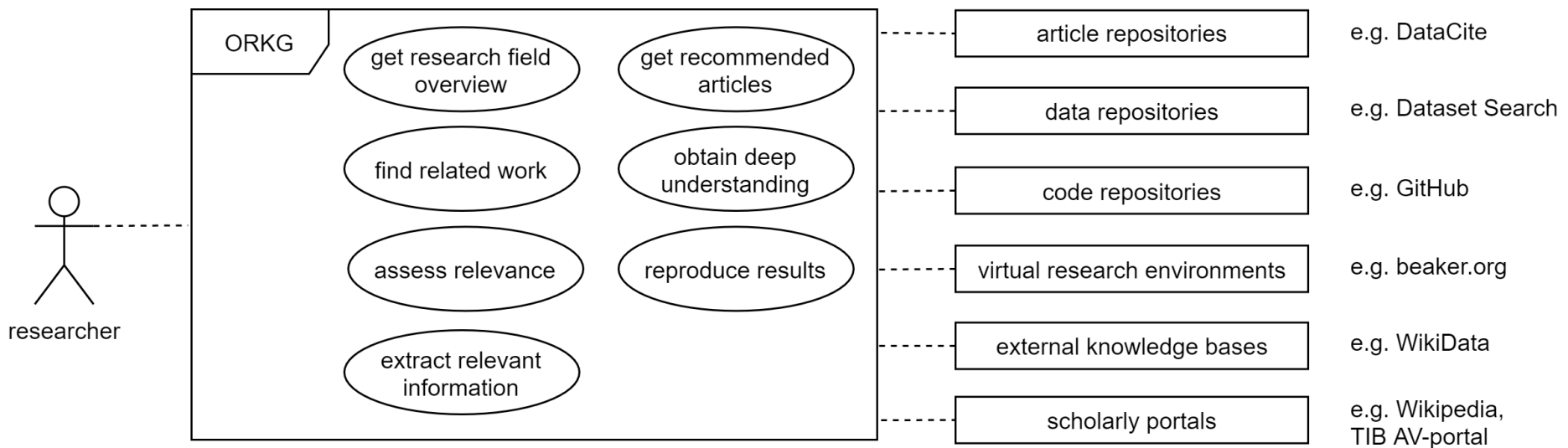
KG Construction

- Manual approaches (e.g. WikiData [65])
 - Automatic approaches (e.g. Information Extraction [46], Graph Construction [56])
- Only moderate accuracy on scientific text with simple ontologies

Use Cases

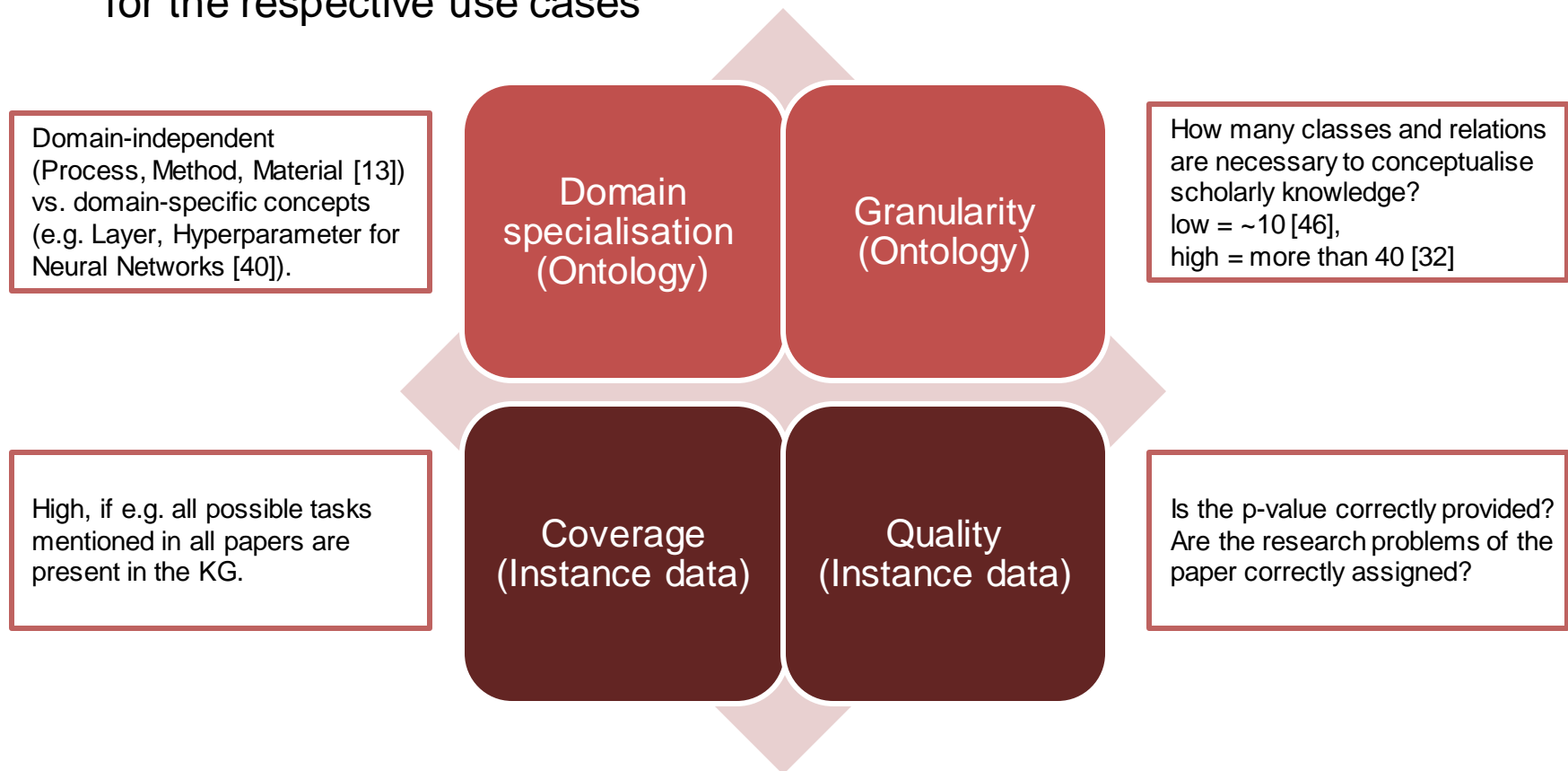


- Main use cases between a researcher, ORKG, and external systems

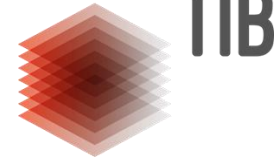


Requirements for the ORKG

- Dimensions for the requirements on the ontology design and instance data for the respective use cases



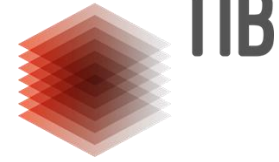
Requirements for the ORKG



- Minimum requirements for the ontology and instance data for the use cases

		Extract relevant info	Research field overview	Deep understanding	Reproduce results	Find related work	Recommend articles	Assess relevance
Ontology	Domain specialisation	high	high	medium	medium	low	low	medium
	Granularity	high	high	medium	medium	low	low	low
Instance data	Coverage	low	low	low	medium	high	high	medium
	Quality	high	high	high	high	low	low	medium

Requirements for the ORKG

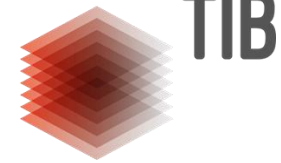


- Minimum requirements for the ontology and instance data for the use cases

		Extract relevant info	Research field overview	Deep understanding	Reproduce results	Find related work	Recommend articles	Assess relevance
Ontology	Domain specialisation	high	high	medium	medium	low	low	medium
	Granularity	high	high	medium	medium	low	low	low
Instance data	Coverage	low	low	low	medium	high	high	medium
	Quality	high	high	high	high	low	low	medium

Similar requirements and high overlap in the ontology concepts and instance data

Approaches for ORKG Construction

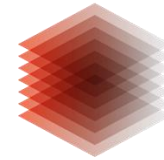


		Extract relevant info	Research field overview	Deep understanding	Reproduce results	Find related work	Recommend articles	Assess relevance
Ontology	Domain specialisation	high	high	medium	medium	low	low	medium
	Granularity	high	high	medium	medium	low	low	low
Instance data	Coverage	low	low	low	medium	high	high	medium
	Quality	high	high	high	high	low	low	medium

Suitable for manual / semi-automatic approaches

Suitable for automatic approaches

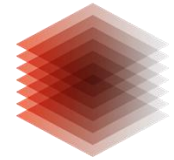
Conclusions



ORKG represents the content of papers in a semantic way

- Several literature-related tasks can be supported by an ORKG
 - Population of a comprehensive ORKG is very challenging
 - Use cases have different requirements for
 - ontology design
 - instance data
- require different approaches for KG construction (automatic vs. manual)

Future Work



TIB

- Refine, implement and evaluate the suggested approaches incrementally

- Current progress of ORKG:

<https://www.orkg.org>

Try it and get involved!



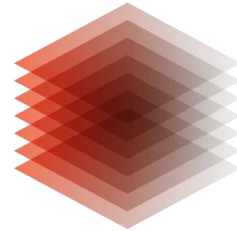
ORKG

- Survey on ORKG requirements:

<http://tib.eu/survey-orkg-requirements>



LEIBNIZ INFORMATION CENTRE
FOR SCIENCE AND TECHNOLOGY
UNIVERSITY LIBRARY



TIB

Thank you for your attention

Contact

Arthur Brack

Email: arthur.brack@tib.eu



Creative Commons Attribution 3.0 Germany
<https://creativecommons.org/licenses/by/3.0/de/deed.en>