Ontology Design for Pharmaceutical Research Outcomes

AUTHORS:

- Zeynep Say,
- Said Fathalla,
- Sahar Vahdati,
- Jens Lehmann,
- Sören Auer.

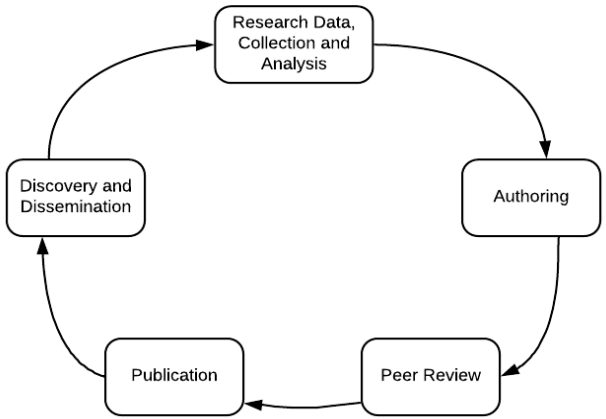
Contents

Introduction

Lifecycle of Scholarly Communication:

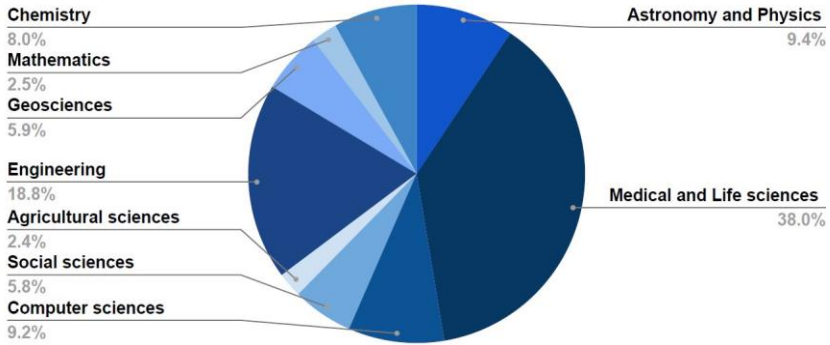
“ The system that scholarly and research writings are generated, assessed, disseminated to the scholarly community, and maintained for future use."

The Publication Lifecycle



National Science Foundation (NSF) Science and Engineering Statistics:

- Publications output: 2.4 million articles in 2017.
- Medical science and life sciences have the highest percentage %38.0.



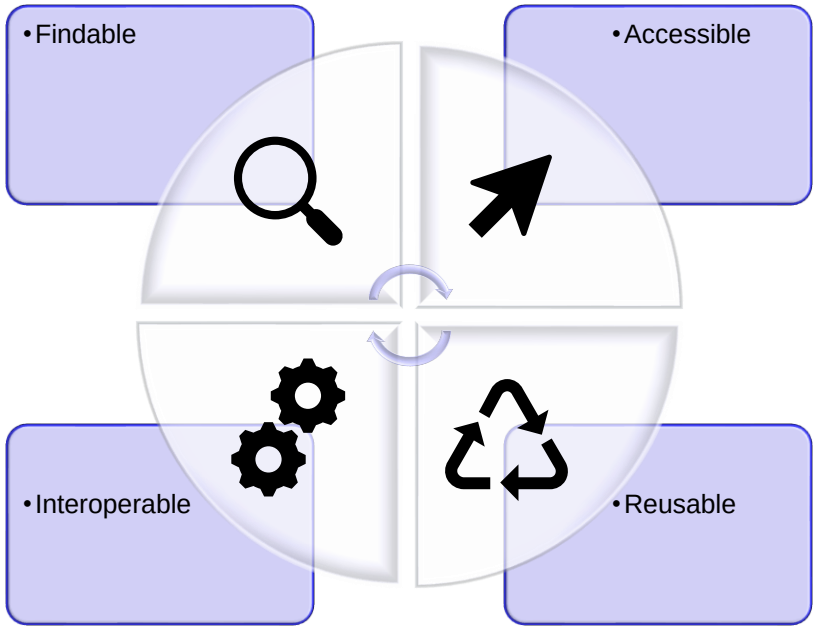
Scientific publication output percentages by field in the world for the year 2017

Reference of figure: Guenther, L. and Joubert, M., 2017. Science communication as a field of research: identifying trends, challenges and gaps by analysing research papers.

Introduction

Main Problems:

- Lack of fully Findable, Accessible, Interoperable, and Reusable (FAIR) data resources,
- Paper documents and their electronic versions,
- Interpreting meaning from unstructured data,
- Research products are scattered across several repositories, journals, or search engines (e.g., Google Scholar , Microsoft Academic , Nature).



FAIR Data Priniples

Google Scholar:https://scholar.google.com/
Microsoft Academic: https://academic.microsoft.com/home '
Nature: https://www.nature.com/

Research Questions

i. How can the scholarly pharmaceutical knowledge be supported with a machine-readable and interoperable domain model?

ii. How can we increase the reusability and accessibility of pharmaceutical research data more effectively?

Goal and Objective

A model for pharmaceutical research, PharmSci Ontology:

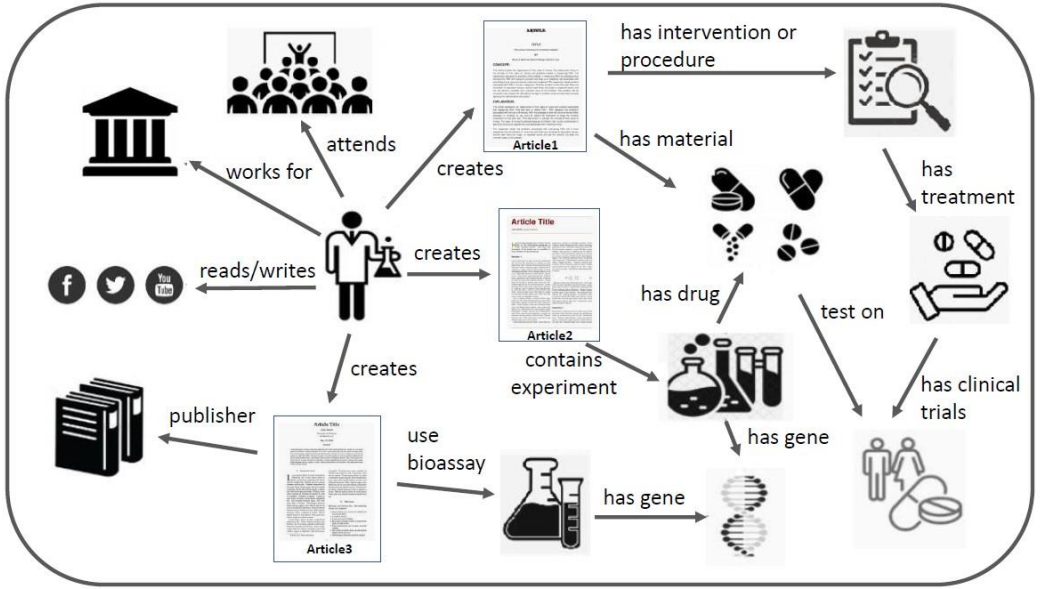
- Facilitating knowledge discovery and management,
- Increasing the reproducibility and reusability of pharmaceutical research,
- Acquire, represent, curate, and integrate knowledge from unstructured web,
- Find out reliable reference materials, sufficient details of experiments or procedures, and re-investigate experiment results.

Methodology

Structure of Knowledge Graphs:

Google introduced Knowledge Graphs as " Things, not strings ".

- Ontologies are employed to create KGs
- Entities are the nodes of the graph
- Relations are the edges of the graph



A knowledge graph of the pharmaceutical research process

Knowledge Graphs: https://www.blog.google/products/search/introducing-knowledge-graph-things-not/

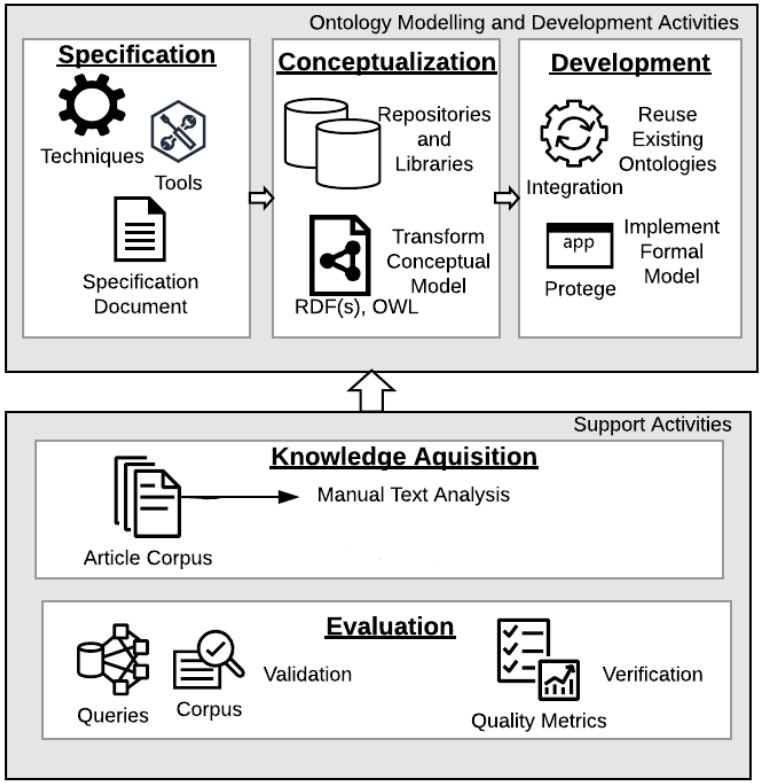
Methodology

Ontological Engineering Aspects

- Specification
- Conceptualization
- Development
- Knowledge Acquisition
- Evaluation

Specification:

- Domain: Pharmaceutical Research
- Purpose: Relevant research results
- Data coverage: Pharmaceutical Research Publications
- Tools: Graffoo, Protégé,...



Ontology Development Activities

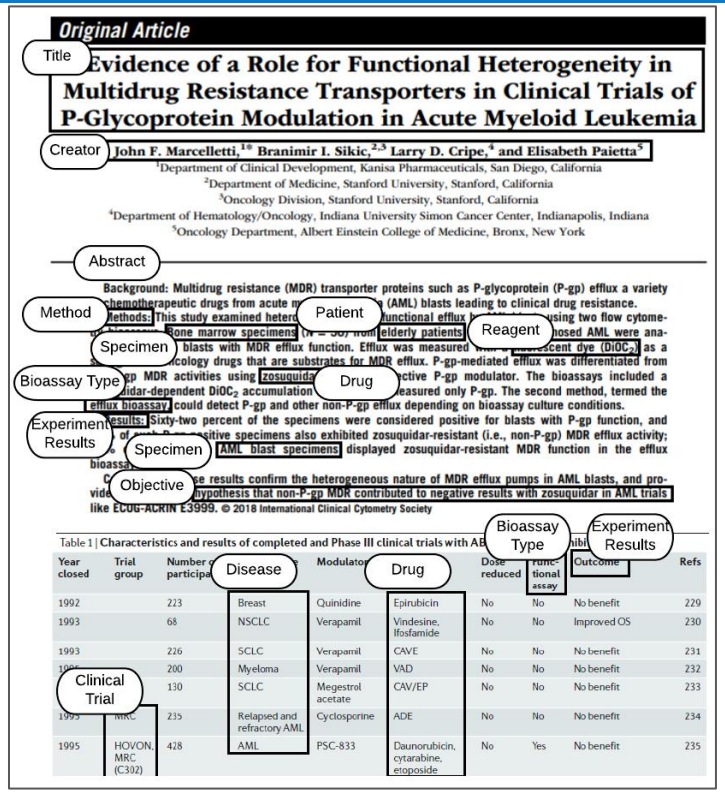
Methodology

Knowledge Acquisition

- Text analysis as a knowledge acquisition technique.
- Corpus: `multidrug resistance and ABC transporters in cancer
- 25 articles are chosen with a systematic review from pharmaceutical journals in Google Scholar and ScienceDirect

Conceptualization

- Informal view of a domain into a semiformal representation
- Complete Glossary of Terms (GT)
- Concept-classification trees
- Subject-predicate-object expressions
- Repositories and open libraries



Captured Entities of PharmSci Ontology from publications

https://scholar.google.com
https://www.sciencedirect.com

Development

PharmSci Ontology Overview:

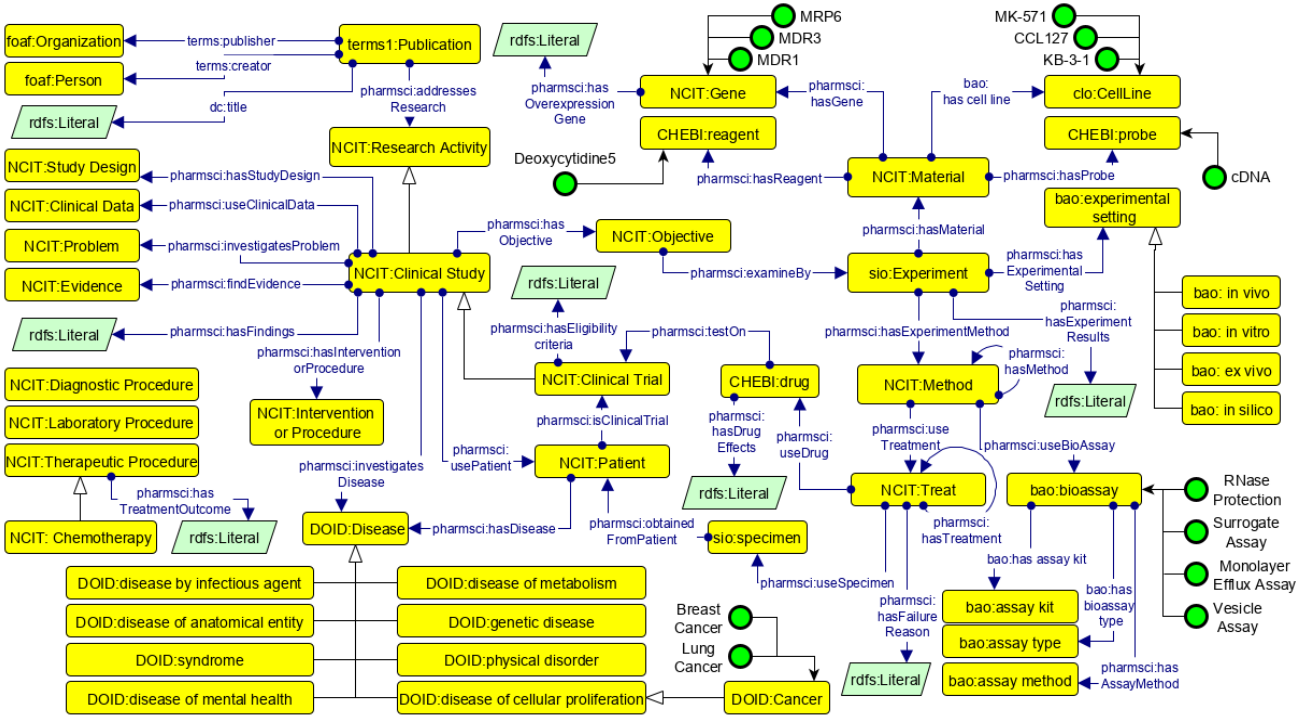
- Formalised by using OWL.
- Formalised ontology is drawn by Graffoo.
- Developed by Protégé v5.5.0

Integrating Reusable Domain Ontologies:

Bioportal - https://bioportal.bioontology.org/obofoundry.org
OntoBee - http://www.ontobee.org/
OBOFoundry - http://www.obofoundry.org/
Linked Open Vocabularies (LOV) - https://lov.linkeddata.es/dataset/lov/

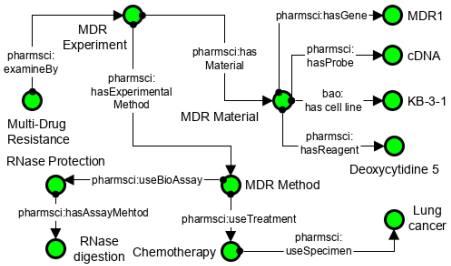
Vocabulary Name:	prefixIRI	URL
The National Cancer Institute's Thesaurus and Ontology	NCIT	http://purl.obolibrary.org/obo/ncit.owl
Human Disease Ontology	DOID	http://purl.obolibrary.org/obo/doid.owl
BioAssay Ontology	bao	http://www.bioassayontology.org/bao
Chemical Entities of Biological Interest	CHEBI	http://purl.obolibrary.org/obo/chebi.owl
Cell Line Ontology	CLO	http://www.ebi.ac.uk/cellline/
Nature Publishing Group Ontologies	terms1	http://ns.nature.com/terms/
FOAF Vocabulary	foaf	http://xmlns.com/foaf/0.1/
The Semanticscience Integrated Ontology	sio	http://semanticscience.org/resource/
The Dublin Core Metadata Initiative	terms	https://www.dublincore.org/

Development of PharmSci Ontology



The integration between the existing ontological entities, subclass hierarchies, object/data properties, and instances in PharmSci Ontology.

Main entities:
Publication, Research Activity, Clinical Study, Experiment, Material, Method, and Intervention or Procedure.



Instances and their relations

PharmSci Ontological entities

Development - Reasoning and Inference

- Reasoning support with SWRL
- SWRL rules to infer new relations
- The rules have been applied with Drools reasoner in Protégé.

Rule 1: ClinicalStudy(?x)∧hasObjective(?x,?z)∧examinedBy(?z,?y)→hasExperiment(?x,?y)

Rule 2: ClinicalStudy(?x)∧hasPatient(?x,?z)∧hasDisease(?z,?y)→investigatesDisease(?x,?y)

Rule 3 : Experiment(?x)∧hasMethod(?x,?y)∧useDrug(?y,?z)→hasMaterial(?x,?z)

Rule 4: hasMethod(?x,?z)∧hasMethod(?z,?y)→hasMethod(?x,?y)

Rule 5: hasTreatment(?x,?z)∧hasTreatment(?z,?y)→hasTreatment(?x,?y)

Evaluation - Validation of Ontology

Competency Questions:

- Knowledge base should be able to answer
- Determine the coverage of the model
- 25 competency questions
- Single SPARQL query for each question

Query	Text
Q1	Which Objective examined by Experiment Y for Clinical Study Z?
Q2	Clinical Study use the Experiment Method Y for Experimental Material X by using Gene as material?
Q3	Which Cancer type X is studied by the Clinical Study Y?
Q4	Which Drugs are used in Therapeutic Procedure X that is used in Clinical Study Y for Disease Z?
Q5	What is title of the Publications that use the BioAssay Y as an Experiment Method?
Q6	Which Cell Lines, Genes, Drugs, Probes are used in the Research Activity X?
Q7	Give Publications that uses Chemotherapy X with drug Y for cancer type Z?
Q8	Give Publication with Experiment Setting In vitro for experiment material Y and Clinical Study X?
Q9	Which Drugs are used in Experiment Y of Clinical Study X?
Q10	What kind of Drugs are used in Clinical Study Y for the Treatment Z?

Query Execution of Competency Questions:

Q5: ""What is the title of the Publications use the BioAssay 'Efflux Bioassay' as experiment method?"

Query answer: "Different Efflux Transporter Affinity and Metabolism of 99mTc-2-Methoxyisobutylisonitrile and 99mTc-Tetrofosmin for Multidrug Resistance Monitoring in Cancer".

```
SELECT  DISTINCT ?title
WHERE {
?publication    pharmsci:addressesResearch  ?study.
?publication    terms:title                 ?title.
?publication    terms:creator               ?creator.
?study          pharmsci:hasExperiment      ?experiment.
?experiment     pharmsci:hasMethod          ?method.
?method         pharmsci:useBisoassay pharmsci:Efflux_Bioassay.
}
```

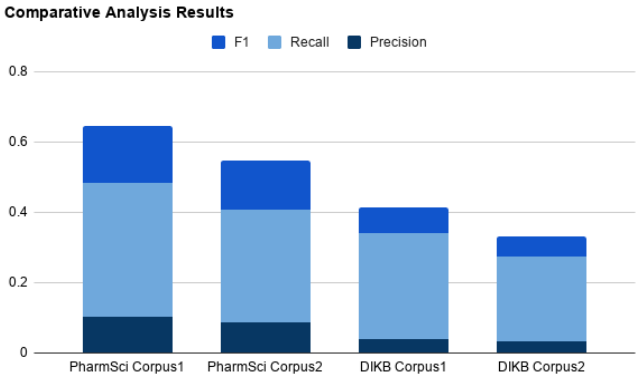
Evaluation - Validation of Ontology

Comparative Analysis:

- Compared with Drug Interaction Knowledge Base (DIKB)
- CORPUS 1: "multidrug resistance in cancer"
- CORPUS 2: "in vitro evaluation in drug delivery“

Latent Semantic Analysis:

- Semi-automatic detection of data
- TF-IDF weight calculation



Precision, Recall, F1 values for PharmSci Ontology and DKIB Ontology

Corpus	Ontology	Class	Keywords	Hits	Precision	Recall	F1
Corpus-1	PharmSci	181	50	19	0.10	0.38	0.16
	DIKB	360	50	15	0.04	0.3	0.07
Corpus-2	PharmSci	181	50	16	0.09	0.32	0.14
	DIKB	360	50	12	0.03	0.24	0.06

Evaluation - Verification of Ontology

FOCA Methodology:
Ontology Type and Questions Verification:

- Goal, metrics, and questions (GQM) approach,

Goal	Question	Metric
1. Check if the ontology complies with Substitute.	Q1. Were the competency questions defined? Q2. Were the competency questions answered? Q3. Did the ontology reuse other ontologies?	1. Completeness. 1. Completeness. 2. Adaptability
2. Check if the ontology complies Ontological Commitments.	Q4. Did the ontology impose a minimal ontological commitment? Q5. Did the ontology impose a maximum ontological commitment? Q6. Are the ontology properties coherent with the domain?	3. Conciseness. 3. Conciseness. 4. Consistency.
3. Check if the ontology complies with Intelligent Reasoning.	Q7. Are there contradictory axioms? Q8. Are there redundant axioms?	4. Consistency. 3. Conciseness.
4. Check if the ontology complies Efficient Computation.	Q9. Did the reasoner bring modelling errors? Q10. Did the reasoner perform quickly?	5. Computational efficiency. 5. Computational efficiency.
5. Check if the ontology complies with Human Expression.	Q11. Is the documentation consistent with modelling? Q12. Were the concepts well written? Q13. Are there annotations in the ontology that show the definitions of the concepts?	6. Clarity. 6. Clarity. 6. Clarity.

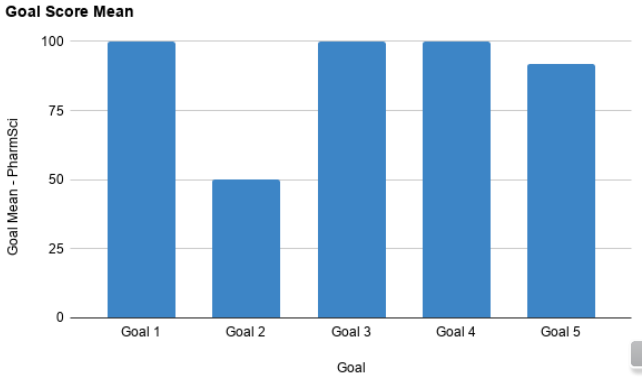
Quality Verification:

- Calculated by the beta regression model. Result: 0.99423.

$$f(x) = \exp\{-0.44 + 0.03(Cov_S * Sb)_i + 0.02(Cov_C * Co)_i + 0.01(Cov_R * Re)_i + 0.02(Cov_{Cp} * Cp)_i - 0.66LExp_i - 25(0.1 * Nl)_i\}$$

$$q(x) = 1 + \exp\{-0.44 + 0.03(Cov_S * Sb)_i + 0.02(Cov_C * Co)_i + 0.01(Cov_R * Re)_i + 0.02(Cov_{Cp} * Cp)_i - 0.66LExp_i - 25(0.1 * Nl)_i\}$$

$$\hat{\mu}_i = \frac{f(x)}{q(x)}$$



Reference: Foca Methodology, Bandeira, J., Bittencourt, I.I., Espinheira, P., Isotani, S.: Foca: A methodology for ontology evaluation. arXiv preprint arXiv:1612.03353 (2016)

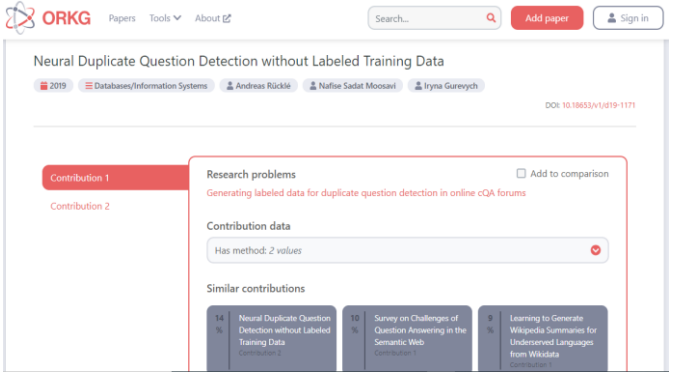
Related Work - Vocabularies and Platforms

Scholarly Domain:

- The Open Research Knowledge Graph (ORKG)
- The Semantic Survey Ontology (Semsur)
- SN SciGraph
- SPAR (Semantic Publishing and Referencing)
- CSO Classifier

Life Science Domain:

- The Open Biomedical Ontologies (OBO) Foundry
- Medical Subject Headings (MeSH)
- The pharmaceutical research domain ontologies (DIO, DINTO, DIDEO, DIKB, and DDI.)



Open Research Knowledge Graph UI

Vocabulary	URL
Cell Ontology (CL)	http://www.obofoundry.org/ ontology/cl.html
Gene Ontology (GO)	http://www.geneontology.org
Protein Ontology (PRO)	http://pir.georgetown.edu/ pro
RNA Ontology (RnaO)	http://obofoundry.org/ ontology/rnao.html
Disease Ontology (DO)	http://diseaseontology.sf. net

OBO Foundry Ontologies

OBO foundry:http://www.obofoundry.org/
ORKG website:https://projects.tib.eu/orkg/

Conclusion

Statement of Result:

- A domain model by using Semantic Web-based solutions,
- Represents rich metadata and machine-interpretable information,
- PharmSci Ontology is one of the Science Knowledge Graph Ontologies (SKGO) Suite ontologies[1].
- As a future work, ontology will be implemented to ORKG and other scientific fields will be covered.
- The documentation of PharmSci Ontology can be found on https://w3id.org/skgo/pharmsci#,
- Prefix(pharmsci) is registered in https://prefix.cc/, a name-space lookup service.

PharmSci Ontology Documentation:

PharmSci Ontology Github Repository:

SKGO Github Repository:

[1]Towards the semantic formalization of science, Fathalla, Said and Auer, Sören and Lange, Christoph, Proceedings of the 35th Annual ACM Symposium on Applied Computing, 2020

Thanks.