

Machine Learning on Semantic Scientific Knowledge

Gollam Rabby

Student (Doctor of Philosophy)

Advisor: doc. Ing. Tomáš Kliegr, Ph.D.

Consultant: prof. Ing. Vojtěch Svátek, Dr.

Department of Information and Knowledge Engineering (KIZI)

University of Economics, Prague, Czech Republic

- ▶ Introduction.
- ▶ Related works.
- ▶ Goals and objectives.
- ▶ Anticipated practical and scientific contributions.
- ▶ Research methodology.
- ▶ Expected outcomes.
- ▶ Current status and achievements.
- ▶ Planning.
- ▶ Conclusion.
- ▶ Publications.

"The goal of creating an efficient machine learning base knowledge graph completion method to the scientific knowledge."

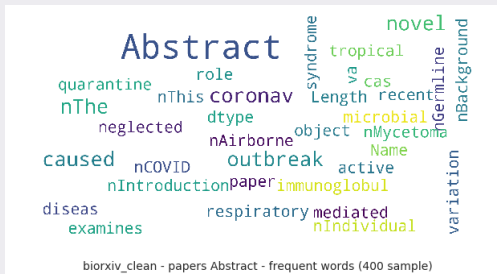


Figure: Most frequent word from papers abstract.

- ▶ Description of different types of Knowledge Graphs.
 - ▶ knowledge graphs.
 - ▶ Research knowledge graphs.
- ▶ Research Ontologies.
- ▶ Model Training and Algorithms.
 - ▶ Knowledge Graph specific modelling.
- ▶ Knowledge Graph completion methods.
 - ▶ Type Completion.
 - ▶ Relation Prediction Methods.

Goals and Objectives



- ▶ To design a concept representation method for scientific knowledge using the existing entity extraction techniques and apply supervised machine learning methods for predicting the impact of scientific publications.
- ▶ To enhance knowledge for a scientific knowledge graph (such as find missing triples) from the different Knowledge base (such as DBpedia, Wikidata, ConceptNet) using an impactful scientific publication and predict the relation in the newly added triples using different machine learning techniques.
- ▶ To apply rule learning tools/algorithms for knowledge graph completion on scientific knowledge KGs.
- ▶ To develop a tool that can assist the scientific knowledge publishers.

- ▶ Create a machine learning method that involves the enhancement of semantic scientific knowledge.
- ▶ A potentially better and validated approach on working with heterogeneous, structured, and unstructured data sources.
- ▶ A method or best practice on how to transform data and use them in the end-user application in a specific domain.

► RDF Graph Sub-Structures Extraction.

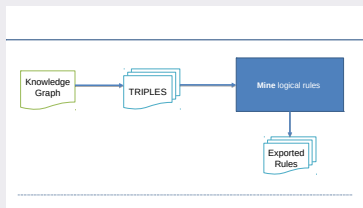


Figure: Procedure for mining rules from KG.

- Association Rule Mining Approach - AMIE+.
- Rule Patterns.
 - *Coronavirus* \Rightarrow *causes* \Rightarrow *covid – 19*
 - *Coronavirus* \Rightarrow *causes* \Rightarrow *flu*
 - *causes* (*coronavirus*, *covid – 19*)
 - *causes* (*coronavirus*, *flu*)
 - *isA*(*X*, *Y*) \Rightarrow *causes*(*Z*, *X*), *causes*(*Z*, *Y*)
 - *Covid-19* \Rightarrow *is a* \Rightarrow *flu*.

- ▶ Constraints
 - ▶ *OnlyPredicates*(x): rules must contain only predicates defined in the set x .
 - ▶ *WithoutPredicates*(x): rules must not contain predicates defined in the set x .
 - ▶ *WithInstances*: enable to mine rules with constants at the subject or object position.
 - ▶ *WithObjectInstances*: enable to mine rules with constants only at the object position.
 - ▶ *WithoutDuplicatPredicates*: rules that contain one predicate more than once will be removed.
- ▶ To create training data depends on the true triples that we found from the rule base machine learning system, train a model and send the false triples for post evaluation.

- ▶ Neural Language Models – C2V.
 - ▶ Bag-of-Concepts Model.
- ▶ Classify important research papers or documents using OpenCitations ontology.
 - ▶ [0].
 - ▶ [0+ to 10].
 - ▶ [10+ to 100].
 - ▶ [100+].
- ▶ Develop a tool that can assist the scientific knowledge publishers.
 - ▶ Focused Categorization Power (FCP).

- ▶ Knowledge artifacts.
 - ▶ Which is represented by introducing a concept called *bag – of – concepts* and a *machine learning method* for enhancement knowledge of the semantic scientific knowledge.
- ▶ Data artifacts.
 - ▶ Which include information about *research articles, citations, the impact of the research, research output*, etc. Also, its represented by the enhancement of the semantic scientific knowledge (KGs) that includes find out the *missing triples, ontologies, vocabularies*, etc.
- ▶ Software artifacts.
 - ▶ Which include *API* and a *tool that can assist the scientific knowledge base and publishers*.

- ▶ Academic Duties.
 - ▶ Subjects (Research Methods, Knowledge Acquisition Methods, KDD – principles, examples and trends).
 - ▶ Webinars (CORD 19 Semantic Annotation, Coronawhy).
- ▶ IGA project.
 - ▶ Knowledge Engineering of Researcher Data (KNERD) (March 2020 - Principal Investigator).
 - ▶ Knowledge Base Enhancement of Researcher Data (KBERD) (Will be submit -March 2021 - Principal Investigator).
- ▶ Collaboration.
 - ▶ Prof. Dr. Sören Auer from Leibniz Universität Hannover.
 - ▶ Prof. Dr. Frank van Harmelen from Vrije Universiteit Amsterdam.
 - ▶ Dr. Ondřej Zamazal from University of Economics, Prague.
- ▶ Industrial Collaboration.
 - ▶ UNICO.AI (September 2020).

- ▶ The time plan for the rest of the author's study program until Summer, 2022.
- ▶ The rest of the 2nd semester and during the 3rd semester.
 - ▶ More requirements analysis, qualitative analysis and comprehensive literature review on machine learning on semantic scientific knowledge.
 - ▶ Internship - Leibniz Universität Hannover, Germany (Open Research Knowledge Graph (ORKG) Group).
- ▶ During the 4th and 5th semesters.
 - ▶ Apply machine learning techniques on semantic scientific knowledge.
 - ▶ Develop a API and deploy models in a server.
 - ▶ Comprehensive evaluation.
- ▶ 6th semesters
 - ▶ Thesis write-up.
 - ▶ Thesis defense (End of this semester).

- ▶ Introduced the context problem to his dissertation and provided a descriptive insight into the proposed project.
- ▶ Mentioned the current status of literature review.
- ▶ Mentioned the anticipated practical and scientific contributions.
- ▶ Basic research methods were introduced.
- ▶ Mentioned the current status of project and several preliminary up-to-date results.
- ▶ Time plan for the next 2 years of his study program.

- ▶ Viet Bach Nguyen, Vojtěch Svátek, Gollam Rabby, and Oscar Corch (Already Submitted - 2020). Ontologies Supporting Research-related Information Foraging Using Knowledge Graphs: Literature Survey and Holistic Model Mapping. International Conference on Knowledge Engineering and Knowledge Management.
- ▶ Gollam Rabby, Tomáš Kliegr (Will be submit - 15 July 2020). Entity-based document classification. Cognitive Computation, (Q1-4.3).

