

Knowledge Graph-based Conceptual Models Search

Syed Juned Ali¹

¹*Business Informatics Group, TU Wien, Austria*

Abstract

Conceptual modeling enables domain understanding and communication among stakeholders and is integral to realizing system requirements. Designing models from scratch can be tedious, even unnecessary and the knowledge from the existing models needs to be utilized to support the conceptual modelers and encourage collaborative learning. An efficient search engine is required to access and search information of interest by applying reasoning to the conceptual models. Conceptual models' structural and semantic information is not naturally applicable to reasoning and AI-enhanced semantic processing. Knowledge Graphs (KG) are an inference-based data structure for storing and reasoning knowledge. Knowledge Graphs benefit from AI-based methods, but such benefits are not directly applicable to conceptual models. This thesis aims to synergize conceptual modeling with the benefits of AI methods on Knowledge Graphs and asserts the validity and value of such a synergy by proposing a modeling language-agnostic KG-based search engine for conceptual models using graph-based machine learning.

Keywords

Conceptual Modeling, Search Engine, Artificial Intelligence, Knowledge Graphs, Graph-based machine learning

1. Introduction

Conceptual modeling is a means to understand and communicate relevant aspects of a system or domain by graphically mapping the domain semantics to the semantics of a modeling language. Therefore, well-designed conceptual models and their encapsulated modeling practices are essential to realizing system requirements. Designing models from scratch can be tedious or even unnecessary. The availability of shared model knowledge provides opportunities for reusing, adapting, and learning from already available models of high quality (F.A.I.R. principles [1]). Publicly available sources of models help create and connect communities of modelers and provide a platform for collaborative learning and empirical research. Despite the discussed benefits and the presence of some language- and domain-specific model repositories¹ [2], we cannot profit from this plethora of knowledge without an efficient way of accessing and searching on and within these models. A model search engine must aggregate data from various sources and provide users with the most relevant models.

Search engines in current modeling tools only provide narrow value because they are limited to individual modeling languages, i.e., each search space is restricted to models represented in one specific language. Moreover, current model similarity metrics are limited to the similarity

ER'22: International conference on conceptual modeling, 17-20 October 2022, Hyderabad, India

 syed.juned.ali@tuwien.ac.at (S. J. Ali)

 0000-0003-1221-0278 (S. J. Ali)

 © 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

¹<https://www.genmymodel.com/>

between search terms and the labels used for model elements (e.g., find all process models having an activity entitled ‘submit application’). Search approaches with OCL-like queries [3, 4] only select exact matches. Existing approaches are not useful in practice since they do not provide access to a high number of diverse and public models. A further fundamental problem when developing enterprise-wide knowledge management based on conceptual models, is the scattered and heterogeneous nature of the relevant knowledge assets (i.e., models). Therefore the knowledge encoded by the conceptual models is not fully utilized and aggregated in the existing search approaches.

Reasoning on the knowledge encoded in conceptual models is necessary to support search based on user requirements; however, such knowledge is not naturally applicable to reasoning. Ontologies are applied to improve the semantics of conceptual models, creating a more sophisticated representation of the domain being modeled and a higher level of domain understanding by its modeler users. This enables machine processing and reasoning based on formal axioms. However, ontology-based reasoning lacks the utilization of the structural knowledge of conceptual models, e.g., graph-based properties.

Google introduced Knowledge Graphs (KG), which hugely improved google search capabilities by aggregating information from various sources (e.g., World Factbook, Wikipedia, and Wikidata) and utilizing the contextual information (cf. Latent Semantic Indexing[5]) of the query by searching on entities, the meaning encoded by the entities and the relationships between them². Using KG based search, a dissimilar appearing result might be relevant because of the incorporation of the contextual information in the search criteria. E.g., a search result can be highly relevant even if the result does not contain the query text but contains concepts associated with the entities in the query. A summary of search engines for models by [6] shows that the existing tools only focus on structure-based search. Therefore, a lack of conceptual models search that utilizes contextual information of models and their elements and aggregates data from various sources to search for relevant models (and not only structurally similar) motivated us to represent conceptual models as a knowledge graph.

Knowledge Graphs represent a collection of interlinked descriptions of entities – objects, events, and concepts. KGs provide a foundation for data integration, fusion, analytics, and sharing based on linked data and semantic metadata. Ontologies represent the formal semantics or data schema of a KG and ensure a shared understanding of the data and its meanings [7]. KGs can effectively organize and represent knowledge to be efficiently utilized by applying different kinds of reasoning (e.g., rule-based and machine learning-based) [8]. The semantically structured information in KGs provides essential solutions for many tasks, including question answering, recommendation, and information retrieval [9]. Therefore KG-based representation acts as an ideal intermediary data structure for capturing contextual information of the models and its elements as well as allows data aggregation.

Consequently, this thesis, on a high level, explores the benefits of applying AI-based methods like graph-based machine learning to the KG representation of conceptual models. In order to achieve this exploration, this thesis attempts to fill the gap of a conceptual model search by, firstly, designing a generic KG transformation tool to create a representation that virtually integrates the heterogeneous models using the domain, core, and foundational ontologies as a

²Google Knowledge Graph url: <https://blog.google/products/search/introducing-knowledge-graph-things-not/>

semantic backbone. This representation semantically links heterogeneous knowledge assets and, during model search, acts as the context for the model and its elements to provide highly relevant results. With the structural and semantic knowledge captured in the KG representation of a conceptual model, this representation is then used as an input for the search engine workflow to index, store, and search conceptual models based on different search criteria. The details of the search engine are described in Section 4

The remainder of this paper is structured as follows - Section 2 presents the related works. Section 3 presents the research definition, laying out the objectives and hypotheses. Section 4 presents the approach for designing the semantic search engine with concrete steps involved and Section 5 present the challenges involved. We conclude this paper with Section 6.

2. Related Work

In this section, first, we describe the works related to our KG-based conceptual model search approach. Further, we describe the relationship between ontology and knowledge graphs with conceptual modeling and how they relate to our model search proposal.

Existing approaches - Basciani et al. [10] show that the search facilities provided by existing model repositories like GitHub, Gitlab, and GenMyModel³ are typically keyword-based, tag-based or there is no search facility at all. There have been efforts in achieving conceptual model search, but several challenges remain. José and Jesús [6] introduce an approach of storing graphs in the form of the possible paths between the nodes present in the graph. This approach focuses on finding structurally similar results and does not consider the contextual knowledge of the searched models. Search engines for individual languages like WebML, UML, BPMN exist [11, 12, 13] however they are not efficient or scalable. MoScript [3] is proposed as a model-independent tool for querying model repositories using OCL-like queries that only retrieve exact models, and Kalnina et al. [14] present a tool for finding model fragments in models. MOOGLE [15] is a generic search engine that performs text search, and the user can specify the type of the desired model element to be returned. Further, these approaches lack the usage of contextual knowledge associated with the models and their elements during the search. In this thesis, the integration of ontologies acts as the semantic backbone of our approach and provides the context for models during search.

Ontology-based semantic enrichment - Ontology-driven conceptual modeling (ODCM) extends or supports conceptual modeling techniques by ontological theories that further articulate and formalize the conceptual modeling grammars of these modeling languages [16]. Foundational ontologies like Bunge-Wand-Weber (BWW) and the Unified Foundational Ontology (UFO) define a range of top-level domain-independent ontological categories, which form a general foundation for more elaborated domain-specific ontologies [17]. Other approaches define language profiles to extend existing conceptual modeling languages and thereby improve the domain semantics [18]. We use the existing techniques from ODCM to incorporate the external semantics to enrich conceptual models.

Knowledge Graphs and Conceptual models - In order to use the benefits of KGs, we present a generic KG transformation of conceptual models. However, several works exist that focus on

³<https://www.genmymodel.com/>

creating KGs from conceptual models to utilize the benefits of KGs as the data model. Sun et al. [19] propose a model-driven approach to automatically constructing KGs from relational databases based on the Model Driven Architecture (MDA). Smajevic et al. [20] propose a CM2KG framework to transform a generic model into a KG, however, lack ontological integration. Huang et al. [21] propose methods to map data, information, and knowledge between class diagrams and KGs bidirectionally and finally generate abstracted class diagrams automatically.

KG embeddings are generated from KG using graph-based machine learning approaches, and further, the similarity metrics are applied to the embeddings for downstream tasks like link prediction, graph clustering, graph classification [8, 9].

3. Thesis Scope Specification

Conceptual models search attempts to mitigate the research gaps mentioned above by integrating ontologies and AI reasoning on the KG representation of conceptual models. To conduct this research, we will use Design Science Research [22]

3.1. Research Objectives

The objective of developing a search engine for conceptual models is divided into the following clear research sub-objectives in order to streamline the research and efficiently develop and evaluate the components of the search engine -

RO-1 Design a generic transformation plugin that maps conceptual models of different metamodels like Ecore, Archimate to a KG by mapping the model and metamodel elements of a conceptual model (entities and relationships) to the metamodel of a Knowledge Graph, thereby creating a **Conceptual Knowledge Graph (CKG)**. The plugin should support the integration of semantics using external domain or foundational ontologies as well as from the results of the graph and model analysis techniques. **RO-1** focuses on creating a modeling language-agnostic representation, i.e., CKG, to incorporate the structural and semantic properties that form the contextual information for model and its elements during search.

RO-2 Design a graph-based machine learning workflow for generating node, edge, and graph embeddings for the CKGs such that the CKG is transformed into a vector space. The embedding needs to capture structural and semantic information of the original CKG. **RO-2** focuses on encoding the CKG into feature vectors that are further used by similarity metrics to efficiently compare vectors and thereby comparing graph components or entire graphs.

RO-3 Design a query language to support the modelers to search within the models based on their requirements. For e.g., search all models with a specific EA smell like cyclic dependency⁴. **RO-3** provides the accessible interface to end-users to search models based on different structural and semantic properties of models.

RO-4 Design similarity metrics to support the different user-specific search criteria that produce highly relevant results. **RO-4** provides the means to achieve search based on different properties by supporting suitable similarity metrics for particular search criteria, for e.g., SimRank [24] for models with elements having neighbours having specific properties.

⁴Hacks et al. introduce the concept of EA Smells [23] analogous to Code Smells

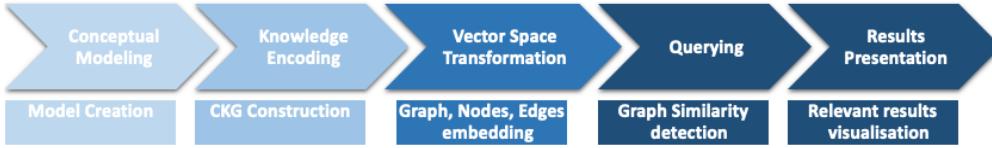


Figure 1: Semantic Conceptual Models Search Process Flow

RO-5 Finally, design a web-based tool to provide a user-friendly interface for the modelers to use the advanced query language and search for models with specific requirements.

3.2. Hypotheses

In order to assess the feasibility, accuracy, and performance of the research objectives, we define the following hypotheses -

H-1 The semantic and structural knowledge stored in a conceptual model can be represented as a CKG, and external sources like foundational or domain ontologies or the results from graph or model analysis can enrich the knowledge captured by the CKG.

This hypothesis deals with the evaluation of **RO-1**. The evaluation will be achieved by developing SPARQL queries that (i) validate the transformation of the original model's graphical structure to CKG (e.g., checking the number of nodes, edges, degree of nodes), (ii) validate model and metamodel level properties on nodes and edges of CKG and, (iii) validate the domain, or foundational ontology constructs integration into the CKG (e.g., checking that a model element from health domain model is linked to an appropriate health ontology construct).

H-2 Machine learning trained CKG embeddings consisting of nodes, edges, and graph embeddings reflect the structure and semantics of the CKG.

This hypothesis deals with evaluating **RO-2**. In order to show search results based on the context of the models and their consisting elements, this information needs to be reflected in the embeddings. Vector space embeddings are feature vectors that capture the properties of the elements of the CKG. Embeddings generation methods can be based on random walk [25] or deep learning [26]. The embeddings will be evaluated based on the idea that embeddings of nodes with similar properties should produce high similarity scores.

H-3 The query language covers the search criteria to search for properties covering structural and semantics knowledge within models.

This hypothesis deals with the evaluation of **RO-3** and focuses on evaluating and validating the expressiveness of the query language. The evaluation needs to check if the requirements in the query are translated to the search engine such that the search engine initiates a search on the required parameters. The search query interface will support textual and graphical search queries (see Fig. 3), and the query needs to be processed to extract the requirements or search parameters.

H-4 The search results reflect the search query requirements with good accuracy

Finally, this hypothesis deals with evaluating **RO-4**. The search requirements trigger specific similarity metrics based on search requirements/parameters to retrieve results based on high similarity scores. The relevant results should have structurally similar models and models that share elements with similar contexts, e.g., models with similar UML design patterns [27]. The

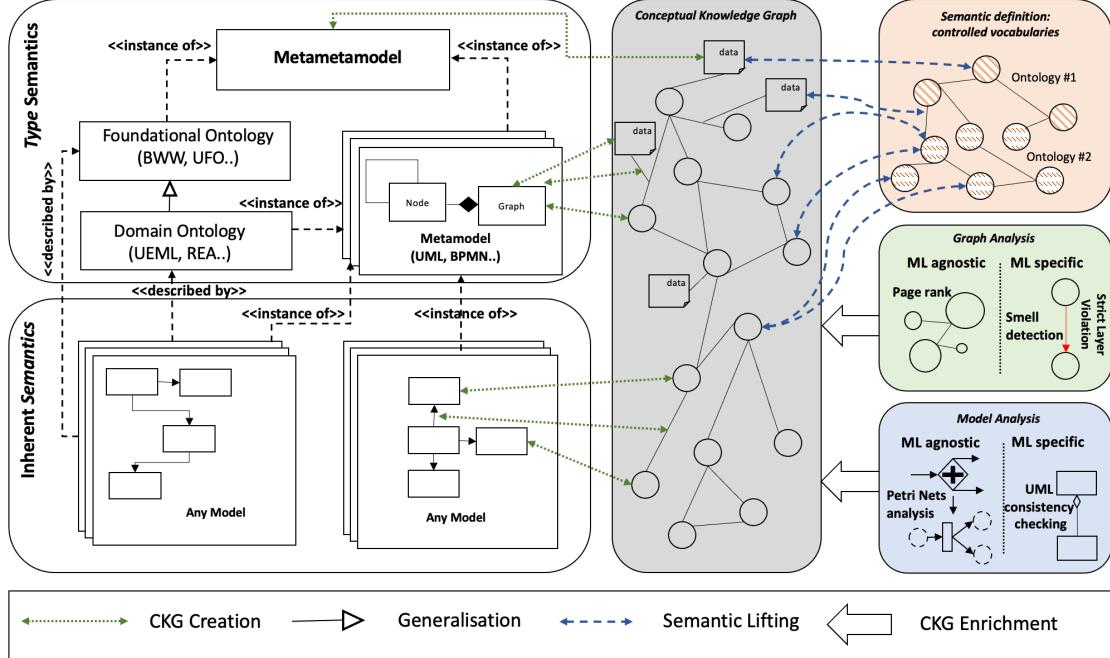


Figure 2: CKG Transformation Framework

results are validated against a validation set of models for each search criteria, and the quality of the results is verified using **recall** and **precision** on the dataset of models.

4. Proposed Search Solution

In this section, we propose the framework that enables AI-based methods for the semantic processing of conceptual models and an end-to-end conceptual models search workflow.

4.1. CKG Construction Framework

During the CKG construction phase, the naming and classification of objects, relations, and properties found in conceptual models are mapped to nodes, edges, and properties in the resulting CKG. Fig. 2 shows this mapping with links from the meta-metamodel, the metamodels or ontologies, and the models to the CKG. The meta-metamodel and metamodel to CKG mapping provide the initial schema for the CKG by forming its nodes and edges. The data attribute associated with edges and nodes of the CKG holds necessary further knowledge. After encoding the explicit semantics of conceptual models in the first step, the second step of the CKG construction process enriches the initial CKG with external knowledge using ontologies (referred to as *Semantic Lifting* [28]) or deriving latent knowledge from existing *graph analysis* or *model analysis* techniques. SPARQL queries and graph algorithms agnostic to the modeling language can be applied to CKGs for analyzing and extracting knowledge. For example, Cypher queries have been applied to EA graphs to find EA smells [20] to enrich the graph with language-dependent knowledge.

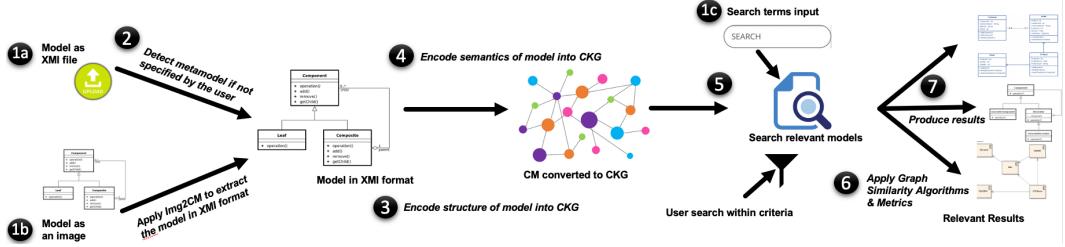


Figure 3: Search Workflow

4.2. Conceptual Models Search Workflow

We describe the end-to-end flow of the proposed search workflow. For querying conceptual models, three means of defining a query seem promising: example-based (alternative 1a in Fig. 1), image-based (alternative 1b in Fig. 3), and keyword-based (alternative 1c in Fig. 3). The knowledge about the metamodel is necessary before searching for a model. The metamodel stores the metadata about the structural and semantic information of the knowledge representation in a model on the type level. In alternatives 1b and 1c mentioned, the metamodel is not defined in advance. [29] propose a machine learning-based approach for automatic classification of the metamodel repositories [[30], p. 2] present an img2uml approach that extracts the Unified Modeling Language (UML) diagram from an image. We propose to extend their work and create an Img2CM framework for converting the image of several widely used conceptual models into an XMI file.

Once we have the XMI file, the conceptual model is transformed into a CKG enriched with structural and semantic information from the conceptual model and integrated with external sources of knowledge. The CKG is then transformed into a vector space by applying GNN algorithms to generate node, edge, and graph embeddings. The GNN models are trained to capture the CKG knowledge into the embeddings. The query language provides the modelers with the interface to search models based on different user requirements. Similarity metrics that combines structural and semantic information (e.g.,[31]) contained by the nodes in KGs will provide relevant results. Therefore, advanced graph similarity metrics can identify graphs similar to the input CKG by incorporating the contextual information of models and its elements in the similarity criteria. Finally, the relevant results are shown in the web interface developed to host the search engine.

5. Challenges

Several challenges need to be handled in order to achieve the proposed objectives of this thesis - **Ontology enrichment for diverse modeling languages** - The CKG transformation is a generic, language-agnostic solution which means the solution needs to be robust towards different modeling languages, and the implementation of the plugins should not be biased towards any modeling language. Our transformation is based on a generic CM2KG platform [20], and we extend the platform towards enrichment. Inferring the applicable ontology to the CKG is a challenge that must be tackled before ontology enrichment. The model's domain can support finding the applicable ontology, or using a unified foundational ontology with manual

intervention might be required for knowledge enrichment through ontology mapping.

Indexing Conceptual Models - Conceptual models contain relationships between entities with semantics depending on the domain, the metamodel of the modeling language as well as the modeler. [6] apply some Natural Language Processing (NLP) techniques (for e.g., stemming, stop words removal, tokenization) on the labels of the graph in their search process, but the graph in the search query can have semantically similar predicates. If the search only captures structural aspects of the graph, the search engine will not fetch graphs with semantically similar relationships. The search engine should tackle a graph's semantic similarity and elements by incorporating the contextual information of element and graphs in the similarity metrics. E.g., graph embeddings should learn modeling patterns like ontology patterns [32], UML patterns [27], and domain-specific patterns as part of the feature vector, which can be further used to find models with similar patterns.

Embeddings of search query CKG - The conceptual model in the search query is first transformed into a CKG and then into the graph embeddings, but it is not easy to directly get the vector embeddings of this new graph because GNN was not trained on this new graph. [33] presents an on-the-fly learning word embeddings method. In this research, we need to develop an analogous method for graphs to generate graph embeddings on the fly and then apply query-specific similarity metrics to retrieve relevant results.

6. Conclusion

In this thesis proposal, we have proposed an AI-enhanced KG-based search engine for conceptual modeling. The main objective of the thesis is to improve the existing conceptual model search by incorporating contextual information about the model and its elements search. We propose a Knowledge Graph-based approach to construct CKGs starting from conventional (ontology-driven) conceptual models further enriched with external knowledge. The generated CKGs will be trained using graph-based machine learning, and then the query language will be used to query the huge corpus of models based on user requirements search criteria.

The novelty of our research lies not only in the research questions we address – which are yet uncovered but also in the innovative idea of facilitating the strengths of KGs and intertwining them with the strengths of conceptual modeling. KGs initially virtually integrate heterogeneous data sources. In this project, we enhance this concept by considering conceptual models as the data to be integrated and linked. KGs further allow semantic similarity calculation (e.g., using the entities linked to the model elements) and structural similarity. We proposed the research objectives of the thesis and presented the steps involved in realizing the solution based on design science methodology.

Acknowledgements

This research has been partly funded by the Austrian Research Promotion Agency (FFG) via the Austrian Competence Center for Digital Production (CDP) under the contract number 854187.

References

- [1] A. Jacobsen, R. de Miranda Azevedo, N. Juty, D. Batista, S. Coles, R. Cornet, M. Courtot, M. Crosas, M. Dumontier, C. T. Evelo, et al., Fair principles: interpretations and implementation considerations, 2020.
- [2] J. Di Rocco, D. Di Ruscio, L. Iovino, A. Pierantonio, Collaborative repositories in model-driven engineering [software technology], *IEEE Software* 32 (2015) 28–34.
- [3] W. Kling, F. Jouault, D. Wagelaar, M. Brambilla, J. Cabot, Moscript: A dsl for querying and manipulating model repositories, in: International conference on software language engineering, Springer, 2011, pp. 180–200.
- [4] K. Barmpis, D. Kolovos, Hawk: Towards a scalable model indexing architecture, in: Proceedings of the Workshop on Scalability in Model Driven Engineering, 2013, pp. 1–9.
- [5] C. H. Papadimitriou, P. Raghavan, H. Tamaki, S. Vempala, Latent semantic indexing: A probabilistic analysis, *Journal of Computer and System Sciences* 61 (2000) 217–235.
- [6] J. A. H. López, J. S. Cuadrado, An efficient and scalable search engine for models, *Software and Systems Modeling* (2021) 1–23.
- [7] J. Sequeda, O. Lassila, Designing and building enterprise knowledge graphs, *Synthesis Lectures on Data, Semantics, and Knowledge* 11 (2021) 1–165.
- [8] X. Chen, S. Jia, Y. Xiang, A review: Knowledge reasoning over knowledge graph, *Expert Systems with Applications* 141 (2020) 112948.
- [9] X. Zou, A survey on application of knowledge graph, in: *Journal of Physics: Conference Series*, volume 1487, IOP Publishing, 2020, p. 012016.
- [10] F. Basciani, J. Di Rocco, D. Di Ruscio, L. Iovino, A. Pierantonio, Exploring model repositories by means of megamodel-aware search operators., in: MoDELS (Workshops), 2018, pp. 793–798.
- [11] B. Bislimovska, A. Bozzon, M. Brambilla, P. Fraternali, Textual and content-based search in repositories of web application models, *ACM Transactions on the Web (TWEB)* 8 (2014) 1–47.
- [12] P. Gomes, F. C. Pereira, P. Paiva, N. Seco, P. Carreiro, J. L. Ferreira, C. Bento, Using wordnet for case-based retrieval of uml models, *AI Communications* 17 (2004) 13–23.
- [13] R. Dijkman, M. Dumas, L. García-Bañuelos, Graph matching algorithms for business process model similarity search, in: International conference on business process management, Springer, 2009, pp. 48–63.
- [14] E. Kalnina, A. Sostaks, Towards concrete syntax based find for graphical domain specific languages, in: 2019 ACM/IEEE 22nd International Conference on Model Driven Engineering Languages and Systems Companion (MODELS-C), IEEE, 2019, pp. 236–242.
- [15] D. Lucrédio, R. P. de M Fortes, J. Whittle, Moogle: a metamodel-based model search engine, *Software & Systems Modeling* 11 (2012) 183–208.
- [16] M. Verdonck, F. Gailly, R. Pergl, G. Guizzardi, B. Martins, O. Pastor, Comparing traditional conceptual modeling with ontology-driven conceptual modeling: An empirical study, *Information Systems* 81 (2019) 92–103.
- [17] G. Guizzardi, G. Wagner, A unified foundational ontology and some applications of it in business modeling., in: CAiSE Workshops (3), 2004, pp. 129–143.
- [18] F. Gailly, G. Poels, Conceptual modeling using domain ontologies: Improving the domain-

- specific quality of conceptual schemas, in: Proceedings of the 10th Workshop on Domain-Specific Modeling, 2010, pp. 1–6.
- [19] S. Sun, F. Meng, D. Chu, A model driven approach to constructing knowledge graph from relational database, in: Journal of Physics: Conference Series, volume 1584, IOP Publishing, 2020, p. 012073.
 - [20] M. Smajevic, D. Bork, From conceptual models to knowledge graphs: a generic model transformation platform, in: International Conference on Model Driven Engineering Languages and Systems Companion, IEEE, 2021, pp. 610–614.
 - [21] L. Huang, Y. Duan, X. Sun, Z. Lin, C. Zhu, Enhancing uml class diagram abstraction with knowledge graph, in: International Conference on Intelligent Data Engineering and Automated Learning, Springer, 2016, pp. 606–616.
 - [22] A. Hevner, S. Chatterjee, Design science research in information systems, in: Design research in information systems, Springer, 2010, pp. 9–22.
 - [23] S. Hacks, H. Höfert, J. Salentin, Y. C. Yeong, H. Licher, Towards the definition of enterprise architecture debts, in: 2019 IEEE 23rd International Enterprise Distributed Object Computing Workshop (EDOCW), IEEE, 2019, pp. 9–16.
 - [24] M. Kusumoto, T. Maehara, K.-i. Kawarabayashi, Scalable similarity search for simrank, in: Proceedings of the 2014 ACM SIGMOD international conference on Management of data, 2014, pp. 325–336.
 - [25] B. Perozzi, R. Al-Rfou, S. Skiena, Deepwalk: Online learning of social representations, in: Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, 2014, pp. 701–710.
 - [26] S. Cao, W. Lu, Q. Xu, Deep neural networks for learning graph representations, in: Proceedings of the AAAI Conference on Artificial Intelligence, volume 30, 2016.
 - [27] R. B. France, D.-K. Kim, S. Ghosh, E. Song, A uml-based pattern specification technique, IEEE transactions on Software Engineering 30 (2004) 193–206.
 - [28] D. Karagiannis, D. Bork, W. Utz, Metamodels as a conceptual structure: some semantical and syntactical operations, in: The Art of Structuring, Springer, 2019, pp. 75–86.
 - [29] P. T. Nguyen, J. Di Rocco, D. Di Ruscio, A. Pierantonio, L. Iovino, Automated classification of metamodel repositories: A machine learning approach, in: 2019 ACM/IEEE 22nd International Conference on Model Driven Engineering Languages and Systems (MODELS), IEEE, 2019, pp. 272–282.
 - [30] B. Karasneh, M. R. Chaudron, Online img2uml repository: An online repository for uml models., in: EESSMOD@ MoDELS, Citeseer, 2013, pp. 61–66.
 - [31] M. A. Alkhamees, M. A. Alnuem, S. M. Al-Saleem, A. M. Al-Ssulami, A semantic metric for concepts similarity in knowledge graphs, Journal of Information Science (2021) 01655515211020580.
 - [32] R. d. A. Falbo, G. Guizzardi, A. Gangemi, V. Presutti, Ontology patterns: clarifying concepts and terminology, in: Proceedings of the 4th Workshop on Ontology and Semantic Web Patterns, volume 1188, 2013.
 - [33] D. Bahdanau, T. Bosc, S. Jastrzebski, E. Grefenstette, P. Vincent, Y. Bengio, Learning to compute word embeddings on the fly, arXiv preprint arXiv:1706.00286 (2017).