

# Building Contextual Knowledge Graphs for Personalized Learning Recommendations using Text Mining and Semantic Graph Completion

Hasan Abu-Rasheed  
Knowledge-based Systems and  
Knowledge Management Inst.  
University of Siegen  
Siegen, Germany  
ORCID: 0000-0002-2921-4809

Mareike Dornhöfer  
Knowledge-based Systems and  
Knowledge Management Inst.  
University of Siegen  
Siegen, Germany  
ORCID: 0000-0002-7069-8534

Christian Weber  
Knowledge-based Systems and  
Knowledge Management Inst.  
University of Siegen  
Siegen, Germany  
ORCID: 0000-0001-6606-5577

Gábor Kismihók  
Leibniz Information Centre for  
Science and Technology  
Leibniz University Hannover  
Hannover, Germany  
ORCID: 0000-0003-3758-5455

Ulrike Buchmann  
Vocational Education Science  
Inst.  
University of Siegen  
Siegen, Germany

Madjid Fathi  
Knowledge-based Systems and  
Knowledge Management Inst.  
University of Siegen  
Siegen, Germany  
ORCID: 0000-0002-7602-9593

**Abstract**— *Modelling learning objects (LO) within their context enables the learner to advance from a basic, remembering-level, learning objective to a higher-order one, i.e., a level with an application- and analysis objective. While hierarchical data models are commonly used in digital learning platforms, using graph-based models enables representing the context of LOs in those platforms. This leads to a foundation for personalized recommendations of learning paths. In this paper, the transformation of hierarchical data models into knowledge graph (KG) models of LOs using text mining is introduced and evaluated. We utilize custom text mining pipelines to mine semantic relations between elements of an expert-curated hierarchical model. We evaluate the KG structure and relation extraction using graph quality-control metrics and the comparison of algorithmic semantic-similarities to expert-defined ones. The results show that the relations in the KG are semantically comparable to those defined by domain experts, and that the proposed KG improves representing and linking the contexts of LOs through increasing graph communities and betweenness centrality.*

**Keywords**— *Knowledge graphs, Graph-based database models, Learning context, Personalized learning, Text mining*

## I. INTRODUCTION

The rapid progress of digitalization in education increased the reachability of learners to online, open educational resources (OER), e.g., through massive open online courses (MOOCs). While this provides more learning opportunities, an approach of one-size-fits-all on online learning platforms may not work for the wide spectrum of learners, their learning goals, and learning contexts. Personalized learning is proven as an essential requirement for better educational development and learning performance. Moreover, personalization becomes more essential in situations like self-learning, on-the-job training, or vocational education and training (VET), where initial training that learners get, e.g., in formal education, is no longer adequate to the job requirements [1], [2]. This is due to learning taking

place within a context, where not only multiple learning objectives may overlap based on the job's requirements, but also the learner's background and current state of knowledge might differ considerably. Modern digital learning platforms need to be more contextualized and personalized environments to realize an educational setting that is meaningful to learners, related to their domain, and tailored towards their preferences and needs, to enhance their capacity-development [1], [2]. To account for the learning context, personalization approaches, such as recommender systems (RS), have been developed as context-aware algorithms. Their efficiency, however, is also bound to the contextual information of learning materials that are being recommended, not only to the learner profiles. A personalized, context-aware recommendation is offered when RS algorithms can match a learner's context to a context of learning objects (LO) and materials. Digital learning platforms still often organize learning content into groups of materials that deploy a hierarchical nature, e.g., as curricula, which are organized into courses, lectures, topics, and finally materials. While this organization is suited to represent the details of a learning objective, it offers no strategy for learning individualization, or model a cross-domain contextualization, which may indicate a learning recommendation, e.g., of courses from different curricula that are required in one application area or one learning context. To consider this context, data models need to be designed in a flexible and semantically enhanced way, which can be realized as a knowledge graph (KG) structure [3]. Unlike hierarchical data structures, KG structures offer the potential to consider relations among LOs. Transferring a hierarchical data model into a KG is possible by identifying relations between elements, based on their semantic relatedness and independent of hierarchical limitations. This identification can be done by text mining methods. In KGs, LOs are defined as nodes, and the relations among them are defined as edges. Relations in the KG define one or more types of relevancies

Funded by INVITE-WBsmart project (grant ID: 21INV121), and ERASMUS+ Key Action 204 OSCAR project (grant ID: 2020-1-DE01- KA203-005713).

between LOs, thus creating a contextual semantic relation based on e.g., the domain of the LOs or their textual similarity. In this paper, we extract information about the shared context of LOs by analyzing their textual semantics. Contextual information is then integrated into the KG as additional relations. Groups of contextually related LOs in the KG may be utilized by a RS, to generate contextual recommendations [4], [5]. We construct the KG utilizing the relations of a hierarchical data model and the semantic relations between LOs, representing the context of those objects in the graph. Semantic relations are extracted with a customized text mining pipeline (TMP), which is designed to analyze and capture the similarities between LOs based on their textual descriptions. A language-dependent similarity algorithm is utilized to mine the semantic similarities between the two description languages of the input data, namely English and German. Our contributions in this paper are: 1) Designing a semantic approach for transferring hierarchical data models in digital learning platforms to graph models. 2) Developing an approach for capturing and modelling the context of LOs to realize semantic KG completion based on a semantic TMP. With the later contribution, present hierarchical structures for LOs can be transformed into contextualizing KGs. In the following sections, an overview of the research background is discussed before we describe the KG construction and semantic relation extraction (RE) among LOs. Then, the evaluation process and results are presented before concluding the paper.

## II. BACKGROUND

Historically, learning is founded in “didactic- and lecture-based methods” which oftentimes are focused only on memorizing content instead of supporting “learning transfer” and engendering “problem-solving” [6]. Wilcke, Bloem and de Boer identified a lack of general data models allowing the representation of more than one domain simultaneously and are thus unsuitable to incorporate heterogeneous data, therefore proposing KGs as a default data model [3]. KGs have been utilized in technology-enhanced learning (TEL) [7]–[9] in recent years. However, the role of the data model in representing the context of LOs in TEL is still less visited in literature, in comparison to e.g. user profiling. Here, we define the context based on situative and subject-oriented learning theories [2], where context is defined as the situation in which learning happens. This includes the learner’s situation (location, previous knowledge, etc.) and the LO’s situation (type, length, implementation scenario, etc.). Context is described by factors that can relate two or more LOs to each other if they serve the same learning situation [8], [10]. Ontologies and KGs have been identified as effective methods for embedding LO context into knowledge representation [7], [11]. Furthermore, research shows that the embedding approaches of contextual information in the KG model and algorithms enable capturing the complex relations that learners follow to solve problems [6], [12], [13]. Several approaches have been followed to construct graph models from LOs [14]. Based on a job-skill hierarchy, authors in [15], [16] utilize the job titles as textual content to create relations between job representations in the KG. To link jobs and their corresponding skills, Dave et al. [17] create three graphs, connecting jobs and skills, using the textual content and existing hierarchical levels. Similarly, de Groot et al. [18] create a text-based skill matching, with the support of the ESCO framework.

The previous approaches, however, do not explicitly address the role of LOs’ context in the KG and do not go far enough to mine further semantic and contextualizing relations between LOs.

## III. CONTEXTUAL KNOWLEDGE GRAPH CONSTRUCTION

We develop a KG as a contextual model that represents the semantic relations among LOs. We adopt the taxonomical structure of the digital learning system in [19], eDoer. This 5-level taxonomy is sufficiently simple and semantically distinctive for the construction of a solid graph data model. Its levels are 1) *Journey*: which represents a learning goal, 2) *Course*: a container that represents one aspect of the learning goal, 3) *Topic*: a concrete knowledge element of a course, 4) *Educational Package*: a group of related and ordered OERs, and 5) *Educational Content*: a concrete OER that can be studied. Semantic relations in the KG connect LOs within one or more taxonomical levels. Hereafter, we use the term LO to represent all levels except the Journey. The resulting KG enables modelling the context of a LO through its connections to other objects, which appear in the same learning scenario.

### A. Text Mining for Semantic Relation Extraction

When creating new Journeys and their content, content-creators provide detailed textual descriptions of each object. These descriptions offer the potential to mine semantic similarities between the objects [5]. Capturing textual semantics requires text-mining algorithms to consider word meanings and sentence structures. We designed a customized TPM, shown in Fig.1, to maximize the use of LO’s textual metadata. The TMP handles objects’ titles (short texts) and descriptions (long texts), written in two languages (English and German). To detect the language, we use a pre-trained model for language detection [20], with a precision of 99% over 53 languages. A text-cleaning step then removes special characters and strings that can affect the text similarity algorithm. The structure of the sentences is kept intact, since text similarities are calculated on a semantic level, not as a bag-of-words (BOW). Titles and descriptions follow two processing paths in the TMP, due to the higher volume of description texts, which may include multiple topics. Such topics provide information about potential relevancies to other LOs. Therefore, the next step for titles is creating title embeddings, while descriptions have an additional step of extracting the integrated topics before the embeddings are created for each extracted topic. To generate title embeddings that are accurate in English and German, we use the Sentence-BERT model [21] and the Spacy library [22] for natural language processing (NLP) in Python. Title embeddings are used to calculate semantic textual similarities between titles using a cosine similarity algorithm [6]. Titles with different languages have embeddings created in different corpus spaces. Therefore, we translate German titles before calculating the cosine similarity. The translation is done automatically using DeepL API translation service [23]. After extracting the main topics from the description using the KeyBERT model [24], text embeddings are created. To compare two descriptions, an intersection matrix is created to determine which topic-pairs from the two topic sets are compared. The final similarity average is then calculated across all topics of both LOs. The resulting scores from titles and descriptions are used to determine if a semantic relation is created between the two LOs,

based on similarity thresholds, which are defined experimentally and then fine-tuned through expert validation.

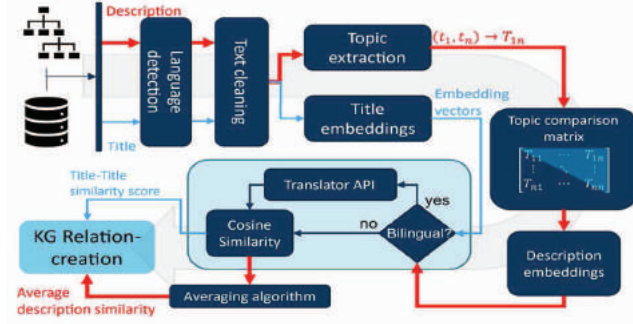


Fig. 1 Text mining pipeline (TMP) and semantic similarity calculation for title and description properties of the learning objects

### B. Knowledge Graph Construction

The structure of our KG includes the original hierarchical model and extends it with semantic relations calculated by the TMP, see Fig.2. We preserve the hierarchical relations since they are created by the content creators, and we add the “has\_semantic\_relation\_to” type to them. In the KG, node types represent taxonomy levels. The new relation type creates direct and indirect connections between learning goals (Journeys) and the LOs. For example, although the Journey “Dementia in elderly care” is not directly connected with the Journey “Communication in elderly care”, several indirect paths were semantically found through connected Courses and Topics that appear in a similar learning context. This context appears in Fig.2 in the form of a densely connected partial network of Topics (area A), whose textual descriptions revealed that they serve the same scenario. As discussed in [4], such a network can enable a learner to achieve a higher knowledge level that enables comprehension, inferencing, and problem-solving. This is accomplished by learning LOs within their context, in contrast to learning isolated LOs that may only enable remembering.

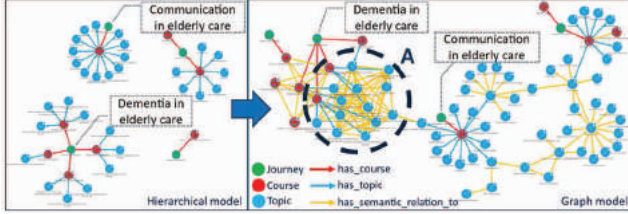


Fig. 2 First three levels of the KG structure, representing the role of semantic relations in creating the KG from the hierarchical one.

## IV. EVALUATION AND RESULTS

The evaluation of KG’s quality utilizes a wide range of metrics, which may have different indications in different domains. In TEL and VET, it is important to consider the metrics, whose indications can reflect the role of the KG in the learning process. KG evaluation should also address the quality of the extracted relations, as a foundation for evaluating the quality of the KG structure. Therefore, our evaluation strategy is implemented quantitatively and qualitatively. Qualitative evaluation is conducted with domain experts to validate the relation-extraction process and results, as well as the KG’s role

in contextualizing the learning recommendations. Quantitative evaluation is designed two-fold: The first part evaluates the quality of extracted relations in the KG. It focuses on the TMP and evaluates its ability to connect LOs semantically in comparison to expert-created links. The second part considers network metrics [25] and quality-control methods, as surveyed in [26], where those are classified in the domain of linked open data (LOD) into six dimensions: representation, accessibility, intrinsic, dataset dynamicity, contextual, and trust. To select the metrics that are relevant and meaningful in the scope of this research, we utilize [6], [27], and [13]. They show that effective evaluation of KG’s ability to represent learning context and the learner’s mental model can be accomplished through metrics that correspond to contextual dimension (relevancy and completeness) and intrinsic dimension (interlinking, density, average degree). We build on those findings and use the metrics of *average degree centrality*, *clustering coefficient*, *weakly connected components*, and *betweenness centrality*.

### A. Evaluation Dataset

For the evaluation, we use a dataset of expert-curated OERs. The data set is provided by Tavakoli et al. [19]. Expert-curated content provides a high level of credibility to the baseline model, which is essential for a fair evaluation. After data exploratory analysis (DEA), we filter out LOs that do not have educational content associated with them and remove the duplicated and isolated ones. The resulting hierarchical model we use for creating the KG features 122 Journeys, 432 Courses, 767 Topics, 2565 Educational packages, and 7358 OERs.

### B. Evaluation of the Semantic Relation Extraction

The goal of semantic relation extraction is to connect LOs that appear in the same or a similar learning context. This goal aligns with the best practices of content creators, who connect an LO to another one if both are needed in the same learning situation. We utilize this logic to evaluate the semantic similarity scores that our TMP calculates among Courses and Topics. We first calculate the average of text cosine similarity scores among the LOs in each Journey ( $i$ ):  $J_i^{sim}$ , by comparing each LO to all other LOs in Journey  $i$ . Then, we compare the similarity score of each semantic relation ( $SR_{i,j}^{sim}$ ) to the average similarity of

the journeys  $i, j$  that it connects  $J_{i,j}^{avg sim} = \frac{J_i^{sim} + J_j^{sim}}{2}$ . If the TMP is able to extract semantic relations with similarity scores that are comparable to the expert-curated relations, we assume that the new semantic relation is meaningful. We define comparability here as being within the same range of both Journeys’ similarity  $J^{sim}$ . This means that  $SR_{i,j}^{sim} \geq J_{i,j}^{avg sim}$ .

Fig.3 shows the scores of a random sample of 240 semantic relations in the KG, alongside the semantic similarity of each Journey-pair they connect. From the results, we can conclude: 1) the similarity scores calculated within each Journey range between 86% and around 90%. This confirms the basic assumption that expert-curated LOs in Journeys are semantically similar. 2) 79% of our semantic similarity scores are either equal or higher than score averages calculated within Journeys, while the remaining 21% is slightly below the average. This indicates that the semantic relations between different Journeys are as meaningful as the relations that experts created within each Journey.



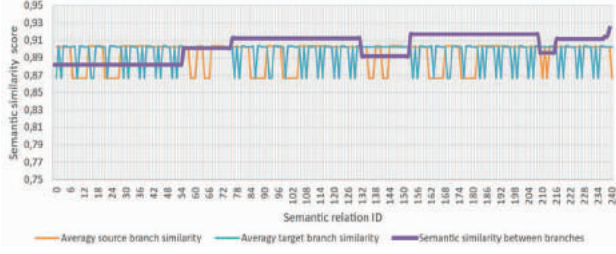


Fig. 3 Semantic-relation textual-similarity scores (Purple), compared to the Journeys (Blue and Orange) connected by those relations through the TMP.

### C. KG Evaluation Metrics

From the wide range of graph quality metrics, our selection is based on the meaning of each metric in the scope of TEL and VET. Selected metrics reflect certain aspects of our KG, when interpreted from a pedagogical point of view in TEL and VET. The comparison here is conducted between the hierarchical structure and the proposed KG, in the light of the value-added that KG structure offers based on the interpretation of each graph quality metric. In the following, we elaborate on the selected metrics and their domain interpretation. Table 1 shows our KG results in comparison to the original hierarchical model.

#### 1) Average Degree Centrality (ADC):

Degree centrality (DC) measures a node's popularity in a graph [28]. It evaluates the connectedness of a node through its incoming and outgoing relations. In an educational use-case, ADC reflects LO's connectedness to other objects that take place in the learning scenario. The reason is that LOs are either connected by experts, or through a semantic relation. While DC in a hierarchical model represents the LO's relation to other LOs from the same Journey, the increase of that object's DC in the graph model shows that new relations have been found to other Journeys, meaning that the LO may appear in more learning contexts. After calculating the DC for all graph nodes, the average (ADC) is determined, which has doubled for our KG.

#### 2) Clustering Coefficient (CC):

CC reflects how strongly a node belongs to a community in a graph. Communities are clusters of nodes that are densely inter-connected. A community is considered more modular if it is easy to separate from other communities. Communities of LOs in our KG represent their context. Therefore, the higher the community count is, the more contexts are represented. Modularity, on the other hand, reflects the potential to separate a learning context from another. Therefore, a lower modularity score is preferred, since it means that learning contexts are well connected to each other. Evaluation results show that our KG increased the detected communities, due to the semantic relations, and decreased graph modularity.

#### 3) Weakly Connected Components (WCC):

WCC detects sets of nodes in the graph that are loosely connected to other graph parts. Therefore, a high WCC score reflects poor connectedness of LO groups in the data model. Using Monge and Elkan's algorithm [29], we find an effective reduction of the loosely connected node groups in our KG.

#### 4) Betweenness Centrality (BC):

BC describes the impact of a node on the flow of information within that graph. It is especially relevant to VET and TEL, since LOs with high BC work as bridges that link individual learning goals to each other, to solve a more complex problem [4]. Using Brandes and Pich's algorithm [30], our results show that the KG has an average BC value of 15.1, which is about 10 times the BC score of the original hierarchical model.

TABLE I. KG QUALITY EVALUATION AGAINST HIERARCHICAL MODEL. PREFERRED VALUE-TRENDS IN VET AND TEL USE-CASE ARE EXPLAINED

Evaluation Metric	Hierarchical data model	KG	Preferred value trends
Average Degree Centrality	1.079	2.262	increasing
Clustering Coefficient (Number of communities)	253	541	increasing
Clustering Coefficient (Average modularity score)	0.779	0.636	decreasing
Weakly Connected Components	63	35	decreasing
Betweenness Centrality	1.57	15.1	increasing

### D. Qualitative expert-evaluation

We also evaluated our KG construction and contextualization approach with domain experts in two focus groups. The groups included experts in VET and researcher-training programs. A total of nine experts participated and answered three questions on 1) The role of connecting learning goals in learning contexts, 2) KGs as contextual data models in their domain, and 3) Using LO's textual description to mine contextual relations. All participants emphasized the need for connecting individual learning goals for solving real-life, job-related problems. Six participants pointed out that highly connected nodes in the graph can represent transferrable skills among different domains. Three experts found a direct use of connecting multi-lingual LOs in their daily work. They also pointed out that a recommendation of multi-lingual content should avoid repetition of the same content written in different languages. Experts also agreed that textual content and description of LOs are important sources for contextual information. They indicated that content creators should roughly understand the way algorithms extract semantic relations, so that they can enrich the descriptions of their LO content with useful contextual information, allowing better connectivity in the KG. The pedagogical experts also addressed the dynamic change of the learner's context, and the development of the learning domain, which requires a continuous update of the KG.

## V. CONCLUSION

In this paper, we introduced a semantic approach for KG completion to enhance the contextual representation of LOs for personalized learning systems. A concept and a text-mining pipeline for relation extraction are proposed, to transfer hierarchical data models into graph ones, thus enhancing the structural and contextual quality of the data model. Our findings from the TMP and KG evaluation suggest that the KG was able to enhance LOs connectivity on a semantic level. Increased connectivity allowed the KG to represent a context around an LO, through its relations to other similar LOs, which appear in the same or similar learning contexts. Proposed TMP can be used with different hierarchical structures of LOs, such as those

generated from other digital learning platforms, since it utilizes the commonly used title- and description properties of LOs. A limitation here is the dependency of our TMP on the volume and quality of LO's textual metadata. The multilingualism of the proposed solution corresponded to a real-world challenge in VET, but it also raised the concern about repetitive content in a KG-based RS. Although our solution is not responsible for the decision-making process in that scenario, it can still be further developed to expand the similarity description, reflecting potential identical content in multiple languages. Further steps of this research aim to increase the robustness of TMP against textual data sparsity and enrich LOs contextualization with additional domain-specific features.

## REFERENCES

- [1] U. Buchmann, *Subjektbildung und Qualifikation. Ein Beitrag zur Entwicklung berufsbildungswissenschaftlicher Qualifikationsforschung*, 2nd ed. Frankfurt a.M., 2011.
- [2] U. Buchmann, 'Vocational-Scientific Education in Qualification Design – Conditional Factors for Quality Assurance and Development of Qualifications Across Educational Sectors', in *Professional-scientific education: discourses, perspectives, implications and options for science and practice*, V. Rein and J. Wildt, Eds., Opladen Berlin Toronto: Verlag Barbara Budrich, 2022, pp. 341–371.
- [3] X. Wilcke, P. Bloem, and V. de Boer, 'The knowledge graph as the default data model for learning on heterogeneous knowledge', *DS*, vol. 1, no. 1–2, pp. 39–57, Dec. 2017, doi: 10.3233/DS-170007.
- [4] D. L. Trumpower, M. Filiz, and G. S. Sarwar, 'Assessment for Learning Using Digital Knowledge Maps', in *Digital Knowledge Maps in Education*, D. Ifenthaler and R. Hanewald, Eds., New York, NY: Springer New York, 2014, pp. 221–237. doi: 10.1007/978-1-4614-3178-7\_12.
- [5] H. Abu-Rasheed, C. Weber, J. Zenkert, P. Czerner, R. Krumm, and M. Fathi, 'A Text Extraction-Based Smart Knowledge Graph Composition for Integrating Lessons Learned During the Microchip Design', in *Intelligent Systems and Applications*, K. Arai, S. Kapoor, and R. Bhatia, Eds., in *Advances in Intelligent Systems and Computing*, vol. 1251. Cham: Springer International Publishing, 2021, pp. 594–610. doi: 10.1007/978-3-030-55187-2\_43.
- [6] P. J. Giabbanelli, A. A. Tawfik, and V. K. Gupta, 'Learning Analytics to Support Teachers' Assessment of Problem Solving: A Novel Application for Machine Learning and Graph Algorithms', in *Utilizing Learning Analytics to Support Study Success*, D. Ifenthaler, D.-K. Mah, and J. Y.-K. Yau, Eds., Cham: Springer International Publishing, 2019, pp. 175–199. doi: 10.1007/978-3-319-64792-0\_11.
- [7] E. Ilkou et al., 'EduCOR: An Educational and Career-Oriented Recommendation Ontology', in *The Semantic Web – ISWC 2021*, A. Hotho, E. Blomqvist, S. Dietze, A. Fokoue, Y. Ding, P. Barnaghi, A. Haller, M. Dragoni, and H. Alani, Eds., Cham: Springer International Publishing, 2021, pp. 546–562. doi: 10.1007/978-3-030-88361-4\_32.
- [8] K. Verbert et al., 'Context-Aware Recommender Systems for Learning: A Survey and Future Challenges', *IEEE Trans. Learning Technol.*, vol. 5, no. 4, pp. 318–335, Oct. 2012, doi: 10.1109/TLT.2012.11.
- [9] A. Visvizi and L. Daniela, 'Technology-Enhanced Learning and the Pursuit of Sustainability', *Sustainability*, vol. 11, no. 15, p. 4022, Jul. 2019, doi: 10.3390/su11154022.
- [10] Y. M. Hemmler and D. Ifenthaler, 'Indicators of the Learning Context for Supporting Personalized and Adaptive Learning Environments', in *2022 International Conference on Advanced Learning Technologies (ICALT)*, Bucharest, Romania: IEEE, Jul. 2022, pp. 61–65. doi: 10.1109/ICALT55010.2022.00026.
- [11] J. Berri, R. Benlamri, and Y. Atif, 'Ontology-based framework for context-aware mobile learning', in *Proceeding of the 2006 international conference on Communications and mobile computing - IWCMC '06*, Vancouver, British Columbia, Canada: ACM Press, 2006, p. 1307. doi: 10.1145/1143549.1143811.
- [12] P. J. Giabbanelli, A. A. Tawfik, and B. Wang, 'Designing the next generation of map assessment systems: Open questions and opportunities to automatically assess a student's knowledge as a map', *Journal of Research on Technology in Education*, 2022, doi: 10.1080/15391523.2022.2119449.
- [13] M. K. Kim and K. S. McCarthy, 'Using graph centrality as a global index to assess students' mental model structure development during summary writing', *Education Tech Research Dev*, vol. 69, no. 2, pp. 971–1002, Apr. 2021, doi: 10.1007/s11423-021-09942-1.
- [14] Y. Fettach, M. Ghogho, and B. Benatallah, 'Knowledge Graphs in Education and Employability: A Survey on Applications and Techniques', *IEEE Access*, vol. 10, pp. 80174–80183, 2022, doi: 10.1109/ACCESS.2022.3194063.
- [15] J.-J. Decorte, J. Van Haute, T. Demeester, and C. Develder, 'JobBERT: Understanding Job Titles through Skills'. arXiv, Sep. 20, 2021. doi: 10.48550/arXiv.2109.09605.
- [16] H. Luo, S. Ma, A. J. B. Selvaraj, and Y. Sun, 'Learning job representation using directed graph embedding', in *Proceedings of the 1st International Workshop on Deep Learning Practice for High-Dimensional Sparse Data*, Anchorage Alaska: ACM, Aug. 2019, pp. 1–5. doi: 10.1145/3326937.3341263.
- [17] V. S. Dave, B. Zhang, M. Al Hasan, K. AlJadda, and M. Korayem, 'A Combined Representation Learning Approach for Better Job and Skill Recommendation', in *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, Torino Italy: ACM, Oct. 2018, pp. 1997–2005. doi: 10.1145/3269206.3272023.
- [18] M. de Groot, J. Schutte, and D. Graus, 'Job Posting-Enriched Knowledge Graph for Skills-based Matching'. arXiv, Sep. 06, 2021. doi: 10.48550/arXiv.2109.02554.
- [19] M. Tavakoli, A. Faraji, M. Molavi, S. T. Mol, and G. Kismihók, 'Hybrid Human-AI Curriculum Development for Personalised Informal Learning Environments', in *LAK22: 12th International Learning Analytics and Knowledge Conference*, in *LAK22*. New York, NY, USA: Association for Computing Machinery, Mar. 2022, pp. 563–569. doi: 10.1145/3506860.3506917.
- [20] S. Nakatani, 'Language Detection Library for Java'. 2010. Accessed: Feb. 02, 2023. Available: <http://code.google.com/p/language-detection/>
- [21] N. Reimers and I. Gurevych, 'Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks'. arXiv, Aug. 27, 2019. doi: 10.48550/arXiv.1908.10084.
- [22] 'Library Architecture · spaCy API Documentation'. <https://spacy.io/api> (accessed Feb. 03, 2023).
- [23] 'DeepL API'. <https://www.deepl.com/en/docs-api> (accessed Feb. 03, 2023).
- [24] M. Grootendorst, 'MaartenGr/KeyBERT: BibTeX'. Zenodo, Jan. 25, 2021. doi: 10.5281/zenodo.4461265.
- [25] C. Guéret, P. Groth, C. Stadler, and J. Lehmann, 'Assessing Linked Data Mappings Using Network Measures', in *The Semantic Web: Research and Applications*, E. Simperl, P. Cimiano, A. Polleres, O. Corcho, and V. Presutti, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 87–102. doi: 10.1007/978-3-642-30284-8\_13.
- [26] A. Zaveri, A. Rula, A. Maurino, R. Pietrobon, J. Lehmann, and S. Auer, 'Quality assessment for Linked Data: A Survey: A systematic literature review and conceptual framework', *SW*, vol. 7, no. 1, pp. 63–93, 2016, doi: 10.3233/SW-150175.
- [27] H. S. Shin and A. Jeong, 'Modeling the relationship between students' prior knowledge, causal reasoning processes, and quality of causal maps', *Computers & Education*, vol. 163, p. 104113, Apr. 2021, doi: 10.1016/j.compedu.2020.104113.
- [28] L. C. Freeman, 'Centrality in social networks conceptual clarification', *Social Networks*, vol. 1, no. 3, pp. 215–239, Jan. 1978, doi: 10.1016/0378-8733(78)90021-7.
- [29] A. E. Monge and C. Elkan, 'An Efficient Domain-Independent Algorithm for Detecting Approximately Duplicate Database Records', in *Workshop on Research Issues on Data Mining and Knowledge Discovery, DMKD 1997 in cooperation with ACM SIGMOD'97*, Tucson, Arizona, USA, May 11, 1997, 1997.
- [30] U. Brandes and C. Pich, 'Centrality estimation in large networks.', *Int. J. Bifurcation Chaos*, vol. 17, no. 07, pp. 2303–2318, Jul. 2007, doi: 10.1142/S0218127407018403.