

## Creation and Utilisation of Domain Specific Knowledge Graphs for E-Learning

Roy Meissner <sup>1</sup> and Andreas Thor <sup>2</sup>

**Abstract:** Educational domain models are building blocks for educational software. Unfortunately, such domain models require high manual effort to be created, are quickly outdated and are thus rarely used at all. A novel way to create such models is proposed with this paper, utilising knowledge mapping algorithms, natural language processing tools, a semantic web architecture and integration with online knowledge bases. The developed processes and tools make it fast and easy to create domain specific knowledge graphs (DSKG) automatically. Apart from defining DSKGs, two use-cases are presented that show how third-party tools may utilise DSKGs: (1) improved exam generation with EAs.LiT and (2) an assessment item and educational material recommender.

**Keywords:** Knowledge Maps; Knowledge Graphs; Domain Models; Educational Software Engineering; Automation; Semantic Web

### 1 Introduction

Educational software systems work with or are built on some form of educational knowledge bases. Such knowledge bases are either manually provided by domain experts (e.g., a lecturer providing the course content) or are implemented into the system (e.g., taxonomies). Research on education and educational software systems shows that data separation into domain, individual and teaching models are desirable [U108]. Unfortunately, creating such models manually imposes a high effort for domain experts, and those created models are quickly outdated due to changing environments [Ho15]. Because of this experience, appropriate models are slow to be adopted and software that uses such is not yet widely used in education. One way to overcome this situation is to separate models from specific tool implementations and to partially automate the (domain) model creation and extension processes, which is presented in section 4. Section 5 showcases how third-party applications can be integrated with such models and section 3 defines the term DSKGs, which is our interpretation of the term domain model. All these topics lead to the creation of a software called Fachlandkarten-Tool (engl. DSKG-Tool), which is usable by domain experts without programming or data scientist skills.

---

<sup>1</sup> Leipzig University, Institute of Educational Sciences, Dittrichring 5, 04109 Leipzig, Germany, roy.meissner@uni-leipzig.de, <https://orcid.org/0000-0003-4193-8209>

<sup>2</sup> Leipzig University for Applied Sciences, Faculty for Digital Transformation, Zschochersche Str. 69, 04229 Leipzig, Germany, andreas.thor@htwk-leipzig.de, <https://orcid.org/0000-0003-2575-2893>

## 2 Related Work

A literature review on Knowledge Graphs (KG) in education defines the term KG, emphasizes on their demand and the usage of Semantic Web technologies [Ri19]. Presented literature and Rizun's own work focus on structural aspects of KGs, their vocabularies, separation of concerns, introduce into KG creation and state about "bottlenecks" of this process. Information about automated creation and maintenance processes were not presented, and KG usage is limited to one example.

The Open Research Knowledge Graph, presented in [Ja19], focuses on information about semantic scholarly knowledge. It uses automated processes to extract information from text in order to semi automatically extend the knowledge graph. Humans may input, revise, maintain and utilise graph information. In contrast, DSKGs focus on educational domain models and software systems, use different and specialized processes and tools for graph creation and revision purposes, and are composable of (sub-)DSKGs by design.

KnowEdu is a system capable of automated KG generation for educational purposes [Ch18]. It uses a data mining approach called probabilistic association rule mining for data extraction from assessment data and tutoring scenarios. The paper at hand differs to it, as its focus is on pedagogical backed modelling, separation of KGs by their domain, graph enrichment from third party repositories, and on different tools to create KGs.

## 3 Domain Specific Knowledge Graphs (DSKG)

Data models for educational purposes are distinguished into three basic kinds: domain model, teaching (or adaption; didactic) model, and user (or learner; individual) model [UI08, pp. 37]. The term *domain model* is widely used and has various meanings, depending on the context. Ulrich defines *domain model* as "The domain model contains the educational resources, and, depending on the specific system, the domain concepts. Additional information associated to the domain model includes metadata and the domain structure (often represented by a graph) that models the relationships between resources. [...]" [UI08, pp. 37]. This definition can still be widely interpreted and the term *domain model* is often used for large and complex domains. This is why this work focuses on a more precise interpretation of the definition, called *domain specific knowledge graph* (DSKG) (germ. *Fachlandkarte*). A DSKG is an open-world graph which imposes no restriction on the size of its domain or topic and is designed for specific didactic purposes. So DSKGs are usable for e.g., a single seminar, but also for e.g., the whole mathematical domain. They are collaborative pieces of data, curated by domain experts and may even contain conflicting information. DSKGs contain key concepts, which are annotated with further information, like concept types, hierarchies, literature resources, and much more, as depicted in the DSKG part of fig. 2.

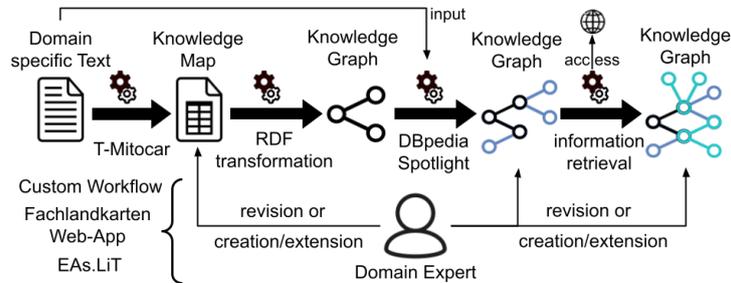


Fig. 1: DSKG creation and extension process, capturing the proposed tools

#### 4 Creation of DSKGs

Manually creating domain models, and thus DSKGs, implies a high workload for domain experts, which is impracticable. This work proposes a semi-automated process to create DSKGs, which is split up into three phases: 1) creation of KGs from source information, using T-Mitocar<sup>3</sup>, 2) extension of these with information extraction processes applied to the source information, using DBpedia Spotlight<sup>4</sup>, and 3) extensions of the resulting KGs with existing third-party knowledge repositories by information retrieval processes. An open-source web app<sup>5</sup> has been developed for domain experts, which guides through manual and automated processes. Manual graph revision, extension and visual exploration is also covered as of this web app. The described process (fig. 1) captures all phases and proposed tools, whereas table 1 summarizes used methods and tools.

Step\Method	Manual	Automated
Creation	Human Domain Modeler	T-Mitocar
Extension	Apps: EAs.LiT, Fachlandkarten-Tool	DBpedia Spotlight

Tab. 1: Used methods and tools for DSKG creation and extension

From a technical point of view the system is modularized and coupled with semantic web technologies according to the *flexible educational software architecture* [MT20]. Even though alternatives might be worth testing, T-Mitocar was chosen because of tech4comp project provisions and implies restrictions on German and English source texts, as well as a minimum text length of 350 words. DBpedia Spotlight was chosen because of it being available as a web-service, providing a concise REST-API and providing data from the “most interconnected freely available knowledge graph on the Web”<sup>6</sup>.

<sup>3</sup> Proprietary Software, see [PDI11], see [https://www.pirnay-dummer.de/?page\\_id=81](https://www.pirnay-dummer.de/?page_id=81) (accessed 21. June 2021)

<sup>4</sup> <https://www.dbpedia-spotlight.org/> (accessed at 21. June 2021)

<sup>5</sup> Source code repository: <https://gitlab.com/Tech4Comp/fachlandkarten-ui/> (accessed 21. June 2021)

<sup>6</sup> <https://www.dbpedia.org/> (accessed at 21. June 2021)

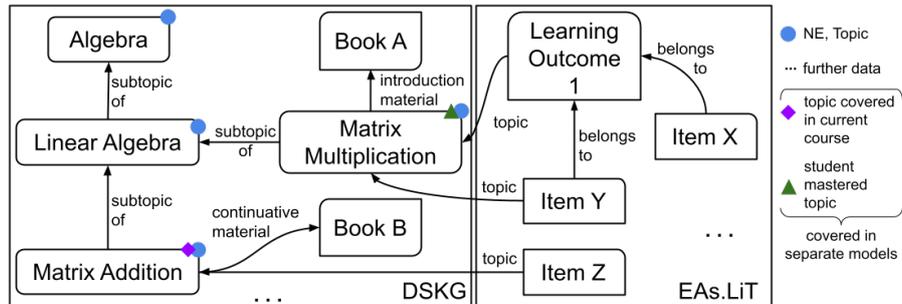


Fig. 2: DSKG integration with EAs.LiT

#### 4.1 Automated Creation

T-Mitocar is a software capable of transferring a written text into an association net (or *knowledge map* (KM)), containing named entities (NE) and weighted associations [PDI11]. Created KMs depend on the text itself, on its length, as well as on an argument dynamically controlling the size of the resulting KM. The KM creation process is fully automated, except for the correct choice of the so called *wordspec* argument, which needs to be determined by domain experts. Created KMs are transferred via custom process to a technical graph format (RDF) for editing, extension, linking and context reasons. This transformation results in a DSKG and the whole process is depicted in the upper left half of fig. 1. As of the simplified example in fig. 2 the so far created DSKG contains the named entities *Algebra*, *Linear Algebra*, *Matrix Multiplication* and *Matrix Addition* only.

#### 4.2 Automated Extension

Several educational use-cases require more specialized information than provided by KMs. One way to gain such information is to analyse the source text with different methods. A tool specifically created for this purpose is DBpedia Spotlight, which offers to extract NEs of a text, providing their types and URLs, leading to further information about the NEs [Me11]. This information is annotated to the created DSKG from step 1), which fig. 2 reflects as all NEs are now referred to as *Topics* (blue dots).

Lastly the DSKG is matched against third party knowledge bases in order to reveal a) hierarchic structures within the DSKG, and b) to gather use-case specific information (not covered as of this work). As of a) contained links of the DSKG are queried to access hierarchical information about the respective NEs on DBpedia, which provides linked and open community knowledge, including hierarchical structures as of topic and subtopic relations [Le15]. In case it is possible to find such structures, the transitive nature of the subtopic relation is used to simplify relations, which are then annotated to the DSKG. Fig. 1 shows both phases in its upper right half and fig. 2 reflects these as *subtopic of* relations.

## 5 DSKG Application & Utilisation

A chatbot, accessing a KG to support students in their literature literacy was presented in [MK20]. The following paragraphs showcase how DSKGs are utilised as of 1) exam generation processes of EAs.LiT, and 2) an e-assessment item and material recommender.

EAs.LiT is a feature unique e-assessment management tool and allows to generate exams and alternative exams of equivalent difficulty [TPW17], and, since a major revision<sup>7</sup>, also as of equivalent item-topic distribution. The latter feature utilises linkage information between EAs.LiT's data (assessment items and learning outcomes) and DSKGs, as well as hierarchic information contained in linked DSKGs (fig. 2). Links are input using a merged visualisation of a DSKG and EAs.LiT's data. Apart of a simple topic categorization, EAs.LiT is able to use the hierarchic information of DSKGs to ensure that an item-topic restriction, e.g., max. 5 items from topic *Algebra*, is not violated by subtopics, e.g., items from *Matrix Multiplication*.

A specified use-case of the tech4comp project is to improve the self-study environment of students by individually recommending assessment items and educational material, based on individual skills, the current course progress and assessment results. The precise concept is described in [MCI21] and integrates DSKGs with data from EAs.LiT and educational material, as well as maps course progress and individual skills of students to this model. The idea is to link assessment items to learning outcomes and to domain topics (fig. 2). By aligning the described data with an educational course, e.g., as a coverage of a specific DSKG (purple square in fig. 2), it is possible to recommend assessment items and educational material to students, based on the current course progress. Such recommendations can be refined by aligning this data with topics mastered by an individual student (green triangle in fig. 2.). Assessment results, utilising such linked items, can provide insights into individual competencies, their alignment with the educational course and topics a student needs to work through in order to reach the desired learning outcomes. For the latter use-case linked educational materials, like introductory material (fig. 2), might be recommended as a supportive measure.

## 6 Conclusion & Future Work

Within this paper the term *domain specific knowledge graph* (DSKG) was defined, which is an implementation of educational domain models, and it was shown how DSKGs can be created automatically. To this end approaches and tools were developed that assist domain experts in all parts of the DSKG creation and maintenance processes. Lastly two use cases were presented, which show how third-party tools can benefit from a DSKG integration.

---

<sup>7</sup> EAs.LiT source code repositories: <https://gitlab.com/Tech4Comp?filter=eas.lit> (accessed at 21. June 2021)

In future work the following topics will be addressed: 1) investigation of challenges like information loss, redundancy, and source text dependencies; 2) extension of the developed tool with further pipeline tools, knowledge repositories, as well as improved visualization and revision capabilities; as well as 3) a study on accuracy, acceptance and efficiency in comparison to other existing approaches and the manual work of domain experts.

### **Acknowledgement**

This work was supported by the German Federal Ministry of Education and Research for the tech4comp project under grant No 16DHB2102

### **Bibliography**

- [Ch18] Chen, Penghe et al.: KnowEdu: a system to construct knowledge graph for education. *Ieee Access*, 6:31553–31563, 2018.
- [Ho15] Hofmann, Sven: E-Learning und ontologisch strukturierte Planung webbasierter Lehr-Lern-Szenarien. Logos Verlag Berlin GmbH, 2015.
- [Ja19] Jaradeh, Mohamad Yaser et al.: Open Research Knowledge Graph. In: *Proceedings of the 10th International Conference on Knowledge Capture*. Association for Computing Machinery, New York, USA, p. 243–246, 2019.
- [Le15] Lehmann, Jens et al.: Dbpedia – a large-scale, multilingual knowledge base extracted from wikipedia. *Semantic web*, 6(2):167–195, 2015.
- [Me11] Mendes, Pablo N et al.: DBpedia spotlight: shedding light on the web of documents. In: *Proceedings of the 7th international conference on semantic systems*. pp. 1–8, 2011.
- [MK20] Meissner, Roy; Köbis, Laura: Annotated Knowledge Graphs for Teaching in Higher Education. In: *Web Engineering*. pp. 551–555, 2020.
- [MT20] Meissner, Roy; Thor, Andreas: Flexible Educational Software Architecture. Preprint arXiv:2104.12557. Presented at the IMHE Workshop of the ITS 2020
- [MCI21] Meissner, Roy; Ruhland, Claudia; Ihsberner, Katja: Kompetenzerwerbsförderung durch E-Assessment. Preprint
- [PDI11] Pirnay-Dummer, Pablo; Ifenthaler, Dirk: Reading guided by automated graphical representations. *Instructional Science*, 39(6):901–919, 2011.
- [Ri19] Rizun, Mariia et al.: Knowledge graph application in education: a literature review. *Acta Universitatis Lodzianis. Folia Oeconomica*, 3(342):7–19, 2019.
- [TPW17] Thor, A.; Pengel, N.; Wollersheim, H.-W.: Digitalisierte Hochschuldidaktik: Qualitätssicherung von Prüfungen mit dem E-Assessment-Literacy-Tool EAs.LiT. *Bildungsräume*, 2017.
- [UI08] Ullrich, Carsten: Pedagogically founded courseware generation for web-based learning: an HTN-planning-based approach implemented in PAIGOS, volume 5260. Springer, 2008.