A Technology-enhanced Smart Learning Environment Based on the Combination of Knowledge Graphs and Learning Paths

Eleni Ilkou and Beat Signer

Web & Information Systems Engineering Lab, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium
{eilkou,bsigner}@vub.ac.be

Keywords: Knowledge Graphs, Learning Paths, E-learning, Smart Education, Smart Learning, Educational Application, Assessment Classification, Personalised Teaching Assistant Tool, Mathematics Education

Abstract: In our position paper on a technology-enhanced smart learning environment, we propose the innovative combination of a knowledge graph representing what one has to learn and a learning path defining in which order things are going to be learned. In this way, we aim to identify students' weak spots or knowledge gaps in order to individually assist them in reaching their goals. Based on the performance of different learning paths, one might further identify the characteristics of a learning system that leads to successful students. In addition, by studying assessments and the different ways a particular problem can be solved, new methods for a multi-dimensional classification of assessments can be developed. The theoretical findings on learning paths in combination with the classification of assessments will inform the design and development of a smart learning environment. By combining a knowledge graph with different learning paths and the corresponding practical assessments we enable the creation of a smart learning tool. While the proposed approach can be applied to different educational domains and should lead to more effective learning environments fostering deep learning in schools as well as in professional settings, in this paper we focus on the domain of mathematics in primary and high schools as the main use case.

1 INTRODUCTION

Have you always been good in maths or physics? Did you have top grades in all your classes or were there one or two classes that you did not particularity like? If that is the case, then the proposed approach aims to provide some insights on why this might happen and how the problem can be addressed. A major issue in today’s educational systems is the restricted amount of individual and customised support that students get from their teachers. This is based on the problem that students can often not identify their own knowledge gaps since they usually do not know what they do not know, making it difficult for educators to assist them properly.

An approach for solving the problem is the creation of a knowledge graph1, a semantic representation of all the knowledge for a given domain and the associations (links) between different topics (Stapel et al., 2016; Rizun, 2019; Lecailliez et al., 2019).

By using a knowledge graph, one can identify any necessary prerequisite knowledge (prerequisite relations) for a given topic and track the areas where a student seems to have a lack of performance. It is obvious that the general knowledge about a domain (e.g. mathematics) is linked in a unique way with a single representation (knowledge graph). Any additional metadata and information that is necessary to support the learning process might refer to such a knowledge graph or be implemented to top of the knowledge graph.

However, a knowledge graph on its own is not enough to point out a student’s potential knowledge gaps. Often students are changing learning environments or moving from one school with a certain educational policy to another school with a different policy. In such a situation, students are facing a new educational system, which makes it difficult for them to perform decently and often leads to school dropouts (South et al., 2007). These problems are based on the fact that each school follows a specific curriculum, which can differ significantly from other curricula. In order to investigate and study the dif-

---

1https://www.blog.google/products/search/introducing-knowledge-graph-things-not/
different systems, the concept of a learning path\(^2\) has been introduced. The learning path of each school represents the order in which the knowledge is being taught—or in other words—how the curriculum traverses the knowledge graph. The learning paths a student has followed in combination with the knowledge graph, as illustrated in Figure 1, could potentially shed light on a student’s secret of success.

Since there are different learning paths resulting from different curricula and educational policies, there should also be a variation in the success of these learning paths. A question we might ask ourselves is whether there are good and bad educational policies. If some learning paths are leading to more successful students, the question is whether we could identify the characteristics of a successful learning path.

Surprisingly, we often see some confusion between the concept of a knowledge graph and a learning path. KnowEdu (Chen et al., 2018b) is an automatic knowledge graph generation tool for educational purposes, which extracts the concepts of subjects or courses from textbooks and exams, and identifies the educational relations between the concepts. Similar to KnowEdu, the K12EduKG system (Chen et al., 2018a) uses the same technique to construct a “knowledge graph”, but with the objective of aiding and improving the flow of teaching and learning, rather than defining course dependencies. The commercial Mathspace\(^3\) solution is an online platform for mathematics making use of a knowledge graph for topics of the U.S. math curriculum and providing teaching assistance for students as well as teachers.

As mentioned above, the problem with some of the presented existing approaches is that even in the scientific and business world, there seems to be some confusion between the concepts of a knowledge graph and a learning path. For example, all the solutions mentioned in the previous paragraph created learning paths based on the specific curriculum of the available teaching material, and not a knowledge graph as they state.

Besides a series of theoretical concepts, a part of the learning path is formed by practical assessments, which differ between learning paths in terms of their content and the specific solving approach. Often students are asked to approach a knowledge topic in multiple different ways in order to be able to apply the theory and multiple methodologies per topic. These methodologies can address different types of assessments and solving approaches, such as assessments solved by calculation or assessments solved by proof. The use of different solving approaches might determine the characteristics of a successful learning path. It is our interest to create a model that can map these correlations between methodologies and groups of assessments. Our approach is trying to integrate the learning outcome with the knowledge topic and to focus on the methodologies that are being taught for the different curricula and their learning paths, in order to construct a knowledge representation that covers all the knowledge components that are applied during the teaching process. Moreover, solving a problem with multiple methods can be beneficial for the knowledge development of students (Cai and Nie, 2007) as well as teachers (Levav-Waynberg and Leikin, 2006).

While there is plenty of educational material available online, the content is often unstructured and not well connected. Practical assessments for the different knowledge domains are usually introduced to students during their study of a specific topic. However, complex assessments with a high difficulty level might require knowledge from multiple domains in order to be solved. This can mislead students and make them believe that they do not have the knowledge about a given topic—such as how to compute the area of a rectangle as illustrated in Figure 5—\(^4\)

---

\(^2\)https://en.wikipedia.org/wiki/Learning_pathway
\(^3\)https://mathspace.co

---

Figure 1: Proposed combination of a learning path and a knowledge graph
although in reality their knowledge gap is in another required knowledge domain (e.g. quadratic equations in the example).

2 SMART LEARNING ENVIRONMENT

We propose the construction of a smart learning environment that can address the problems presented in the previous section. The system will contain and link educational material for a given domain, such as chemistry, mathematics or history. The system is designed around one or multiple knowledge graphs, which can be connected; a physics knowledge graph for instance shares concepts with a mathematics knowledge graph hence there are interconnected topics. On top of the knowledge graphs, we represent the curricula (learning paths) of different schools the students are following. The learning paths can help us to provide personalised aid to students. Besides defining the dependencies between knowledge topics in a knowledge graph, individual assessments should be linked to the corresponding knowledge topics. Therefore there is an interest in developing an assessment classification based on the different characteristics of assessments. Eventually, the results of our research should become available in the form of a smart learning tool supporting learners as well as teachers, as illustrated in Figure 6.

In this position paper, we discuss the potential next steps in research and development for educational material as described in the book “How People Learn” (Council, 2000). We focus on the examination of common practice by reviewing existing curricula and assessments, the extension of the knowledge base by proposing the development of new educational material, and the development of smart learning tools based on key research findings and the principles of learning. We further investigate successful and creative educational practices and explore the foundations of learning to support new research on assessments that focus on improving teaching and achieving a deeper understanding of a specific domain.

2.1 Knowledge Graphs

Educational data has common characteristics as it usually forms hierarchical structures going from the simplest knowledge topic to the more advanced ones, and from the earliest event to the most recent. With that in our mind, we can construct a knowledge representation where knowledge topics are arranged in a directed graph. The graph consists of nodes representing the knowledge topics and their components, as well as links representing the relationships between specific knowledge topics. A framework for the construction of such a graph based on the resource-selector-link (RSL) hypermedia meta-model (Signer and Norrie, 2007) has been proposed in EduKnow (Ilkou, 2019).

The knowledge graph aims to reveal the inner structure of the knowledge topics, by handling the core of the educational material and teaching process, which is represented by the learning outcomes. By focussing on the learners and their deep understanding of a domain, we shall define rules of how the knowledge topics are related to each other and finally create an abstract model of a graph representation for an individual educational domain. For that purpose, we need to have a well-defined structure that can be used in the same way for each learner and at the same time be enriched with the necessary elements to represent all knowledge components. The model needs to be defined theoretically in order to have an abstract foundation for future research. Furthermore, a theoretical foundation of our knowledge components and their relationships is necessary if we thrive to achieve reproducible data and a framework which can be used for knowledge representation in arbitrary educational domains.

Our aim is to develop the foundations for an adaptive learning recommendation system. The core of such a system forms a smart educational application containing a knowledge graph representation for the educational domains. Based on this knowledge representation it becomes possible to implement the learning paths of different curricula and to also track a learners knowledge gaps. Aleks⁴ which is based on the book entitled Knowledge Spaces (Doignon and Falmagne, 1999), uses the concept of learning paths to guide the students based on the curriculum they are following. It is unclear how exactly the knowledge graph space is created, but we know that the nodes represent subjects from the U.S. curriculum. The difference between the proposed adaptive learning recommendation system and existing solutions is that adaptive learning will create a knowledge representation based on how the knowledge itself is structured, and not by some specific curricula. Hence, we will use the knowledge graph “forwards” (in the direction of the directed links) to find the next proposed topics for reaching a knowledge goal, but also “backwards” (opposite to the direction of the directed links) to discover a student’s knowledge gaps. By taking into consideration the mistakes that a learner is making, we can probabilistically determine the prerequisite topics

⁴https://www.aleks.com
that might form the source of a learner’s poor performance. Moreover, we aim to visualise the knowledge structure in order that the actors involved in the learning process are able to view and understand the reasons behind a student’s performance and the possible way to success. Note that this novelty in knowledge representations and knowledge structures represents one of our main contributions.

Let us have a look at an example for the domain of algebra taken from the 8th grade mathematics corpus of the Greek curriculum (B’ Gymnasiou). This example brings light to why it is important to separate the knowledge graph from the learning paths. Figure 2 shows parts of the curriculum (learning path) for the knowledge of ‘linear functions’ and the corresponding knowledge topics of the knowledge graph are illustrated in Figure 3. We can see that the learning path, which is a sequence of knowledge topics does not correspond to how the knowledge is structured and does not reveal all the knowledge connections necessary to reach the end goal. This knowledge graph can help in detecting knowledge gaps by querying the links for a specific knowledge topic as discussed in EduKnow (Ilkou, 2019) and shown in Figure 3 for the topic of ‘linear functions’. In our example, the knowledge topics that correspond to Figure 2 are represented in blue. There are further different types of links representing different kinds of relationships between knowledge topics. We can create a subgraph of parts of the database, such as the one shown in Figure 3 with the ‘linear functions’ query object as a final node. This visualisation might assist the student and their teacher to find all the possible reasons (non-mastered topics) for a student’s poor understanding of a given topic.

**Table of contents**

1. Equations – Inequalities
   1.1. Algebraic expressions
   1.2. 1st degree equations
   1.3. Solving formulas
   1.4. Solving problems with equations
   1.5. 1st degree inequalities
2. Real Numbers
   2.1. Square root of a positive number
   2.2. Irrational and real numbers
   2.3. Problems
3. Functions
   3.1. Definition of a function
   3.2. Graph of a function
   3.3. Function \(y=ax\)
   3.4. Function \(y=ax+b\)

---

**Figure 2**: Part of the table of contents for the 8th grade mathematics corpus of the Greek curriculum showing the learning path leading to the knowledge of ‘linear functions’

**Figure 3**: Parts of the knowledge graph for the 8th grade mathematics corpus of the Greek curriculum highlighting the knowledge components for the topic of ‘linear functions’

We treat a knowledge graph as a multidimensional representation which connects knowledge topics but also contains all the necessary knowledge components for a given knowledge topic. As mentioned previously, a knowledge topic is a node in the knowledge graph and consists of the title of the topic as the highest level of abstraction, and the methodologies, theory and solved examples at a more detailed level. In the suggested model, knowledge is the outcome of the learning process and represents what a student should know by the end of a chapter or academic year. There is often confusion for existing knowledge representation techniques on what should be represented as a node in the knowledge graph, which is another difference between our proposal and existing knowledge mapping techniques. In existing solutions, the nodes are many times formed by knowledge topics with very generic content, such as numbers (Chen et al., 2018b), or with very specific content such as the chapters of a curriculum (e.g. Consumer Arithmetic (11-12))

Therefore, it is of great importance to already have properly identified all the knowledge topics before beginning with the linking process. Hence, this task is carried out by a domain expert who can formulate each knowledge topic based on specific learning outcomes.

Each knowledge topic should have a title and at least one methodology, which explains how an assessment can be addressed in an abstract way together with the illustration of a solved example. The methodology is many times found as part of the introduction of a new chapter for a given subject and forms an important component. By analysing different curricula

---

5https://mathspace.co
When $ax^2 + bx + c = 0$ with the solutions $x_1, x_2$, we know that

$$x_1 + x_2 = \frac{-b}{a}$$

$$x_1 \cdot x_2 = \frac{c}{a}$$

Vieta’s Formula

Completing the Square

$p$-$q$ Methodology

Discriminator

When $a$ is not 0, there are two solutions $x_1, x_2$ to $ax^2 + bx + c = 0$. The solutions are

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

where $b^2 - 4ac$ is the discriminator.

Figure 4: Visualisation of the knowledge components for the knowledge topic of ‘2nd degree equations’

In the domain of mathematics, we found that students are often being taught the same topics, such as multiplication or solving quadratic equations, but they are introduced to a specific technique (methodology) to address assessments which might vary significantly for different curricula. Moreover, based on the focus of the curriculum, students are asked to perform certain tasks for a given topic, such as doing calculations to find a variable, or to learn only the theory of it. In other cases, students are asked to approach a knowledge topic from multiple ways in order to be able to apply the theory and multiple methodologies for a given topic. These methodologies can address different types of assessments, such as assessments solved by calculations and assessments solved by proof. It is in our interest to create a model that can map these correlations between methodologies and groups of assessments. Our approach is trying to include the learning outcomes of a knowledge topic and to focus on the methodologies that are being taught for the different curricula and learning paths to construct a knowledge representation that contains all the knowledge components that are used during the teaching process. An example for the domain of mathematics showing all the knowledge components of the knowledge topics of ‘2nd degree equations’ is shown in Figure 4. In this example, a knowledge topic consists of different methodologies. Each methodology is represented as a branch of the root 2nd degree equations and contains the methodology name, the theory and the solved example (not shown in the figure).

In contrast to existing techniques, our approach represents the knowledge nodes as the learning outcomes, due to the methodologies describing the approaches to address the assessments. Our novel approach handles educational data from the learner’s point of view and lets them know what exactly they should be able to do in order to succeed. We find this format necessary to achieve high scalability and a reproduction of the same knowledge structure. Existing models do not always result in the same knowledge representation, as they are strongly related to the educational material they are processing. Further, these techniques demand a great number of educational data in order to create a decent graph. In the proposed construction of the knowledge graph, all we need is an expert in the field, such as a teacher, in order to first properly identify the knowledge topics, and then associate them via the corresponding links. By the time this procedure is completed, the resulting knowledge representation can be reused and enriched by others. Other experts can build on the existing knowledge graph by adding new methodologies or missing knowledge topics and links. The reason for the possible reproduction is the fundamental structure of the knowledge graph. Since the graph is structuring pure knowledge without taking into account any learning paths (e.g. from textbooks or curricula), it represents fundamental knowledge that will not change from one institution to another. The only thing that might happen is that a curriculum addresses assessments with a different technique, or offers extended courses that cover extra material that is not yet implemented by the knowledge representation of the knowledge graph.

On the practical side, another difference to the existing body of work is the association of assessments with the corresponding knowledge topics. We link assessments with the knowledge domains they depend on in order to be fully addressed. These connections to the necessary knowledge domains are not always obvious as illustrated in the example shown in Figure 5. In order to find the right answer in this example, a student needs to have knowledge about 2nd degree equations, the area of a rectangle and some basic principles of geometry. This assessment illustrates...
the importance of linking an assessment with all its required knowledge topics and also reveals that the connections of the knowledge topics should be based on theoretical as well as practical (assessments) aspects.

**Question:** Find $x$ if the area is $3\text{cm}^2$

**Solution:**

Area $= lw$ [area of rectangle]

$x - 2$

$3 = x(x - 2)$ [variables]

$3 = x^2 - 2x$

$3 + x^2 - 2x + 1$ [2nd degree equations]

$4 = (x - 1)^2$

$2 = x - 1$ or $-2 = x - 1$

$x = 3$ or $x = -1$ [principles of geometry]

$80, x = 3$

Figure 5: Complex assessment requiring knowledge from multiple domains

Assessments have been studied for performance-based applications (Linn et al., 1991), been classified in existing intelligent learning environments (Le and Pinkwart, 2014), and there exist automatic systems for the recommendation of assessments (Harjula, 2008). However, complex assessments that require knowledge from multiple domains to be solved are not well linked to those domains. Usually, an assessment belongs to a single knowledge topic and therefore only inherits any prerequisite connections that the corresponding knowledge topic has. This is also the case that many students are facing, as they solve simple assessments but cannot perform more advanced ones since the advanced assessments require knowledge from multiple domains. In this case, these connections are not obvious to students and hence they cannot realise that an assessment needs extra knowledge (and which extra knowledge) to be solved. Therefore, in the case of a complex assessment, a student might not perform well without knowing the reason for their bad performance. With the proposed knowledge graph approach, a learner is aware of the complexity of an assessment and all the required knowledge. An important aspect of the proposed approach is therefore that it reveals the hidden connections of knowledge that exist on a practical level.

### 2.2 Encoding of Learning Paths

Another important part of the proposed approach is the study of learning paths which will be encoded on top of the knowledge graphs. There is an interest in researching the different learning paths since often students with different learning paths are asked to take the same final exams, which might have a major impact on their future career (baccalaureate diploma and multidisciplinary courses). Our study should provide more insights into the different used curricula, identify their strong and loose connections between topics, and help students in higher education who wish to follow courses from different departments. The goal is to identify these learning paths and to encode their content (components), compare them, and help students understand how their problem matches course prerequisites, and keep track of their performance and knowledge. Students would further understand the goal of each course. Seeing the big picture will help them to increase their performance and prevent them from dropping out, as well as help them to faster adapt to the courses they are following.

The proposed research is challenging, given that it requires a formulation of the main components of each learning path, and inventive when it comes to the study of the strong and successful as well as the weak points of curriculum following different learning paths that ultimately lead to the same diploma. We foresee to contribute to existing knowledge with a number of innovative aspects for educational systems and their performance.

The results of our research might assist future education counsellors to make decisions based on the performance of different learning paths when designing a study path, or to extend the curriculum for any sector that is related to teaching or learning. Our study will also help companies to formulate an expertise learning path based on the knowledge their employees should possess in order to increase their productivity (Željko Šundov and Gregorić, 2014).

### 2.3 Assessments Classification

After having created the learning paths, we will need to enrich them with assessments. We propose two parts for the assessment analysis, based on solving approaches and the knowledge domains an exercise is referring to. Regarding the first case, we are planning to create metadata about different solving ways which will be an innovative feature to be used in an assessment classifier. Moreover, an assessment classifier tool could classify assessments based on the prerequisite knowledge that is needed to solve them. Based on our idea of combining the learning path with a knowledge graph, a new classifier can follow the link to the knowledge graph, get the prerequisites for a given topic and thereby identify all prerequisite knowledge for a given assessment. An example is the exercise in which the square power of 7 is $7^2 = 49$ has a prerequisite set of exercises ($7 \times 7 = 49$). There will be some challenges when classifying complex exercises as we need a multi-dimensional classifier. Hence, the research on the solving ways and data retrieved from the
associated knowledge graph will enable us to come up with a new assessment classification model. By classifying the provided material in different clusters, we will be able to study the relations between topics and identify optimal routes for learning paths, as there will be associations between assessment domains that do not exist as a clear sequence in the learning path, such as in Figure 5 between quadratic equations and the calculation of an area. Also, on the student side, we will be able to provide a larger variety and quantity of assessments for a given topic to be studied. The results of such a study will benefit students who are preparing for their final exams, teachers, as well as companies working in the domain of educational technologies since they will be able to automatically classify and share their advanced learning material.

2.4 Smart Learning Tool

A smart learning tool could enable the centralisation of online and digital exercises to exist in a common pool with more people being able to benefit from. It might also help students to identify in a fast and accurate way their knowledge gaps and the precise goal each student needs to reach in order to succeed in each grade, based on a school’s curriculum, and track their performance based on their knowledge and not only their grades.

All the previous theoretical results will form the basis for the development of a smart learning tool that can potentially help students and teachers in real time and consists of the components shown in Figure 6. A teacher or tutor is often lacking time for a full diagnosis of a student’s problem in order to be able to provide some personal guidance that will clarify all questions of the student. Our tool will come to add and not replace the guidance of a teacher, by offering a smart educational environment, where the student can be aware of their strength, weaknesses and the goal of their level of education (grade). At the same time, the teacher or tutor can provide specialised assistance with rich content based on a student’s previous learning path and knowledge gaps. Given that the tool can classify the exercises and select them in order to close a student’s knowledge gaps, it should help to increase the success rates on the final exams of each grade and diploma examinations. The goal is to develop a recommendation system which, based on the assessments and the existing rich content, will gradually follow up the student’s skills until they have reached the proper level to face difficult exam questions with confidence instead of fear and embarrassment. The benefits of such a system are of financial and pedagogical nature, such as the standard of quality, the re-usability, scalability, convenience and time-saving.

Our tool could be used as an integrated application for already existing educational programs. It might further serve as an indicator in the application processes of teachers as well as other job candidates. In teaching, the tool could potentially be used as a system for examinations or real-time interaction with enriched class content, which could provide mixed media interfaces and allow more interactive learning based on each student’s needs.

3 POTENTIAL APPLICATIONS

The envisioned technology-enhanced smart learning environment is scalable and can be used for any educational material and domains such as mathematics, physics, biology and history. Based on our proposed approach, we could potentially detect the best practices in learning paths, curricula and exam assessments to instruct a more successful next generation of students, who are better prepared for exams and have been individually assisted based on their knowledge gaps. The results of the proposed research might also encourage the mobility of families and students, as educators will be able to identify the learning path of each student. Note that this is in line with policies towards a modernised and adaptive educational system, which try to address inequality and unqualified dropouts and a modernised educational policy that can determine the qualifications of a successful education system. Another application might be in the domain of the screening of applications for enrolment at universities or job vacancies. In this case, the smart learning tool could compare a candidate’s profile to the criteria of a given position, identify the candidate’s chances for success and assist them to improve their profile to match the expected level of expertise.

Furthermore, companies working in the domain of e-learning solutions might benefit from our research as they will be able to enhance their products with a rich model that will help them to share their material. Note that the proposed research might also
benefit companies who are offering courses to assist high school education and companies that are partially working on e-learning projects.

Moreover, there might be an economic impact on companies that offer their employees internal training and educational growth or learning paths, as they will be able to provide more personalised trainings and seminars that will reduce the lost time for overqualified employees, save money for trainings within organisations, and easily and quickly verify a candidates knowledge level for a specific position.

Future educators will be able to provide specialised individual support to learners, which will increase their productivity and satisfaction. On the same page, future counsellors will be able to modify curricula and educational systems in an easy, fast and, most importantly, accurate way to adapt the needs of a challenging and constantly improving the academic and professional environment. Last but not least, students will enjoy personalised assistance that identifies their weak points and potential knowledge gaps, and get to know where they need help in order to reach their educational goals. Of course, they will also learn about their strong points, which will further push them towards pursuing their talents.

4 CONCLUSIONS

We have presented our vision of a smart learning tool that might support students as well as teachers in their daily life. The smart learning tool is going to be based on a knowledge graph which connects the different knowledge topics and also enables the tracking of each student’s learning trajectory, making it easier to identify a student’s knowledge gaps and plan the next steps for a successful learning trajectory. Further, the proposed approach helps in detecting good practices in learning trajectories (learning paths) and methodologies, such as which topic should be taught first and based on which methodology. On the practical level, the smart learning application foresees a large pool of assessments, which are categorised by the knowledge they require in order to be addressed. This aids learners to have a more accurate recommendation of assessments based on their needs and profile. Therefore, the proposed technology-enhanced smart learning environment combining knowledge graphs and learning paths is also of great help for teachers, as it makes the detection of a student’s knowledge gaps a more efficient and accurate process and provides them technology-based assistance for the personalised recommendation of assessments.

REFERENCES


