VRIJE
UNIVERSITEIT
BRUSSEL

Graduation thesis submitted in partial fulfilment of the requirements for the degree of Master of Science in Applied Sciences and Engineering: Applied Computer Science

EDUKNOW: A FRAMEWORK FOR STRUCTURING EDUCATIONAL MATERIAL

Mathematics as a Use Case

ELENI ILKOU

Academic year 2018–2019

Promoter: Prof. Dr. Beat Signer
Advisor: Prof. Dr. Beat Signer
Faculty of Engineering

Abstract

Knowledge representations have been introduced in the last years as a tool to structure and organise knowledge topics and connect related instances. The knowledge mapping techniques can be used as an individual assistant or in a classroom to visualise the content and its connections. Other graph techniques are promoting automatic concept and knowledge graph creation. However, these techniques are based on probabilistic models and they do not use a uniform format for creating graphs. The presented EduKnow framework introduces a strict mathematical formulation based on set inclusion to describe the linking on the knowledge graph, and the knowledge topics represented as nodes in the graph.

Our aim is to create a framework which can structure educational data. The core of our knowledge representation is the knowledge topic which represents the *learning outcomes*, being able to perform on the knowledge topic content. The knowledge topic consists of methodologies. Each methodology has a theory and a solved example part which help learners to have a deep understanding and provide them with ways to address assessments. The difference to existing techniques is our multidimensional format, which creates a knowledge representation with nodes as the knowledge topics, and also analyses the knowledge components of each knowledge topic. A single knowledge topic can be composed of multiple methodologies and many solved examples. Also, the knowledge topics represented as nodes are linked with other nodes via the links the EduKnow framework introduces.

Our linking model is formed from three links including the *prerequisite*, *shared content* and *assessment* link. They are defined based on mathematical formulations and represent the different type of connections and relationships we find in the EduKnow framework. The prerequisite link is the most popular among the techniques that use uniform structure in the related work. However, in our design we do not use any probabilistic metrics. The shared content link defines the relationship between knowledge topics that are not related in a prerequisite relationship, but have some shared content. Further, the assessment link is a very innovative approach compared to the existing models. The idea behind it is that complex assessments require knowledge from multiple domains to be solved. Therefore, there is a hidden connection between these topics, which the assessment link discloses.

Our implementation is based on the resource-link-selector (RSL) hypermedia metamodel. RSL allows us to have a rich model that supports our EduKnow framework. The EduKnow framework implementation allows a large number of visualisations of knowledge compared to the related work techniques. We are offering the representation of the whole database, but

the EduKnow framework is offering additional six representations that are created from the combination of the three links. These provide generous information visualisation to the learners to set a better understanding of their current state and get a clear path to their final goal.

The implementation of the EduKnow framework has been validated based on a concrete example in algebra. By analysing the knowledge representation we can view the benefits of our model compared to similar techniques and discuss its potential use in classrooms. Also, we have a clear perspective on how the EduKnow framework visualisations offer more opportunities for learners to have a straightforward vision of their current knowledge and the ways to excel. Moreover, another benefit of the visualisation is that teachers can detect the learning gaps of a student faster and in a more precise procedure.

Acknowledgements

First and foremost, I would like to express my gratitude for my promoter Professor Dr. Beat Signer for his guidance, support, and continuous feedback. I deeply appreciate the time and faith he added to my research. I cannot recall all the times I was leaving his office highly motivated to work on my project and make it great. It is a unique feeling and I would like to thank him from my heart.

Moreover, I would like to thank Professor Dr. Jean-Paul Doignon, for the valuable discussion, and my mathematician friends Achilles and Eleftheria, for their scientific comments that assisted towards the formulation of my research model.

Further, I would like to express my gratitude to the members of the WISE lab for their friendliness and help throughout the year. I wish everyone in the lab great success in their careers.

Finally, I would like to thank my family, friends, and partner for everything they have done for me, and especially my grandmother. Her love, kindness, and intelligence have shaped who I am today. I feel blessed to be her granddaughter, she is my source of patience, determination, and endurance. This journey would not have been the same without her.

Contents

1 Introduction

Have you always been good at Maths or Physics? Did you have top grades in all classes or were there one or two classes that you did not particularity like? Often students try to succeed but fail to achieve their goals. Many students attend classes and study daily, yet they find their results do not meet their effort [30]. This is due to the fact that they are focusing on the wrong direction, by spending time trying to excellence a topic, while they lack the understanding of some of its prerequisites. Also, often the detection of their knowledge gaps is not a trivial task for the students, as well as for their teachers, leading to more wasted time and a slow growth of student's performance.

An approach to solve this problem is the creation of a knowledge graph, a semantic representation of all the knowledge for a given domain and the associations (links) between different topics. By using a knowledge graph, one can identify all necessary prerequisite knowledge (prerequisite relations) for a given topic, and track the areas where a student seems to have a lack of knowledge.

In this thesis, we are presenting the concepts and the implementation of EduKnow, an educational knowledge graph framework which offers an innovative way of linking data and many knowledge visualisations. A knowledge graph can be referring to a specific category, for example, human diseases, or potentially could contain any information which is not obviously related, such as movies and dog breeds, which could be indirectly connected through movies, their actors, the names of their dog, and the dog breeds. Besides their semantic representations, knowledge graphs are necessary for studying how the knowledge domains are linked. Moreover, there is an interest in the ways different learning systems (curricula) are traversing a knowledge graph.

For this reason, there are three types of links presented in this thesis in order to define the different relations between knowledge domains. The choice of the right type of link (relationship) between the knowledge topics is an important component that needs to be well defined in order to achieve the maximum usability and scalability of a knowledge graph.

In order to help the reader understand the importance of knowledge graph with proper linking structure, we are presenting a small example. This basic case will disclose the confusion on linking the knowledge in a hierarchical model, which is in many cases unclear even between researchers. Hence, let us assume that we want to classify the concepts of *square powers* and the topic of *square roots*. What is the relationship between the two concepts? If you answered that there is a prerequisite from the square powers to the square roots you are mistaken. Both topics, exist in the same knowledge topic $b^n, b \in R$, where the square powers are a methodology for $b^n, b \in R \; and \; n = 2$, and the square roots are a second methodology for $b^n, b \in R \; and \; n = \frac{1}{2}$.

The solution to how the knowledge is linked could bring light to the insights of the knowledge structure and its relations in a specific and uniform solution. Today's world contains plenty of information which cannot be used because it is not well structured, linked and organised. With the proposed EduKnow framework one can have access to the visualised information of the study material, which shows the components of a knowledge topic as well as the connections with other knowledge topics. This provides a view of a learner's current state in combination with the prerequisites and/or shared content topics and the assessments which refer to them. It is becoming simpler to realise all the previous steps that need to be mastered, and all the next steps need to be taken to reach a goal.

The presented EduKnow framework solution could be used in a classroom as the basis material of a personalised assistant, which can detect the knowledge gaps of a student. This teaching assistant could provide in-depth guidance helping students to achieve more with less time and effort. In addition, EduKnow would provide a semantic representation and guidance between multiple courses and knowledge domains, for example connecting the knowledge graphs of mathematics and physics through the connection of quadratic formulas and motion equations respectively. Furthermore, the recommendation pool of assessments can be a helpful tool for teachers, where they could find assessments separated by their difficulty level and complexity. Additionally, the EduKnow framework can be the basis of the learning path implementations[1].

1.1 Definitions

In this section, we provide a brief explanation of the terminology that is going to be used in this thesis.

Educational Domain

An educational domain is a course subject such as mathematics, biology or literature. An educational domain can consist of many knowledge domains.

Knowledge Domain

A knowledge domain is a subpart of an educational domain. For example, algebra and geometry are knowledge domains of the mathematics educational domain.

EduKnow

The EduKnow framework is an educational graph consisting of knowledge topics that are linked together via different types of links. It can be deployed for multiple knowledge and educational domains.

Knowledge Topic

The nodes in the EduKnow graph are knowledge topics, consisting of the title of the topic at the top level, and the methodologies, theory, and solved examples at a more detailed level. We will also refer to it as *knowledge node*.

Learning Path

A learning path is the directed sequence of steps to traverse the knowledge graph based on the curriculum that is followed. This sequence consists of the knowledge topics that are introduced with the specific methodologies and types of assessments.

Example: The Greek and the Swedish mathematics curriculum follow different learning paths, even when they introduce students to the same topic such as quadratic equations. The Greek curriculum solves it with the

[1] `https://en.wikipedia.org/wiki/Learning_pathway`

methodology of the discriminant, while the Swedish curriculum uses the completing square methodology.

Methodologies

Each knowledge topic has one or more methodologies to be studied and some assessments. A methodology can refer only to a certain type of assessments, for example, the assessment type of proof. A methodology consists of its name, theory, and solved examples. A methodology can also be found as *method*.

Example: The 2nd degree equations have multiple methodologies, such as the completing square, Vieta's formula, the discriminant and others.

Theory

Each methodology has a theory part which explains in an abstract way the approach and steps to address assessments.

Solved Example

Each methodology has at least one solved example which explains on a concrete example the way to solve it, step by step. We also refer to it as *example*.

Assessment

Any assignment or exercise, theoretical or practical, that can be used as examination material to check a student's knowledge about a specific topic. An assessment can have a difficulty level from 1 (very easy) to 10 (very difficult), and a type. The type of the assessment is strongly depended on the educational domain the knowledge topic it belongs to. For example, in the mathematics educational domain the types of addressing assessments can be proof or calculations.

Assessment Group

A group of assessments consists of assessments that require knowledge from the same topics and have the same type.

Personal Path

The learning path a user has followed from the beginning of their education tracking point up to the current time. It can be part of a single learning path, or it can consist of multiple sub-paths, in case the student has followed different curricula at different grades.

1.2 Problem Statement

Students nowadays have more access to information than any previous generation, yet they find it difficult to detect the right piece of information they need. 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 often students cannot 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.

Furthermore, many times, students are changing learning environments or moving from one school with certain educational policy and curriculum to another school with very different characteristics. 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 [55]. Moreover, the workload on specific modules [23] and the difficulty of written assignments [6] has been reported to affect dropout rates.

An approach to solve this problem is the creation of knowledge mapping techniques [19] and the most recent knowledge graph. Knowledge graphs were first used by Google in a multi-dimensional way for linking related data. However, knowledge graphs and their predecessors are not using any uniform and theoretical foundation on how a representation is created, and they do not specify which are the referenced knowledge domains and more specifically how the different knowledge topics are linked.

Surprisingly, we often can find quite some confusion between the concept of a knowledge graph and a learning path. KnowEdu [15] is an automatic knowledge graph generation tool for educational purposes. This system extracts the concepts of subjects or courses and then identifies the educational relations between the concepts. Similar to that, the K12EduKG system [14] uses the same technique to construct the knowledge graph that focuses on subject concepts rather than courses, with the objective to aid the flow of teaching and learning rather the definition of course dependencies. In the company sector, Mathspace[2] is an online platform for mathematics that uses its own knowledge graph for topics of the maths curriculum in the USA, and around the world, provides teaching assistance for students as well as teachers. The problem with these previously mentioned techniques is that even in the scientific and business world there is a confusion between the knowledge

[2] https://mathspace.co

graph and learning path. All the previously mentioned techniques created learning paths based on the specific curriculum of the books and material they had available.

On the other hand, we see that there are many semantic representations for the knowledge domain of mathematics, such as knowledge maps and knowledge graphs. However, there is no published and public available graph that contains the knowledge topics content; the theory, solved example and assessments; with the links to other topics. Therefore, there is no knowledge graph in mathematics that specifies the knowledge topics that mathematics education consists of and how they are interconnected.

We also observe that there is often overlapping educational material, in the theoretical parts and as well as on the level of the assessments. This material is often poorly organised and not linked properly so that it classifies the prerequisite relations and the knowledge topics with which it shares content. Furthermore, more complex assessments require knowledge from multiple domains which is not always identified. For example, a student cannot solve motion equations in physics because there is a knowledge gap in quadratics formulas in mathematics. By connecting each assessment to its knowledge domains a student can identify their knowledge gaps and have a deeper understanding of the educational material.

The lack of deeper understanding might also be the reason behind the drop out of students [16]. Moreover, students often lack the ability to transform the abstract knowledge into practice which leads to low motivation for studying and performance. Therefore, a model that would contain the connections between different domains and knowledge topics could solve this problem. Also, the study of Kyle A. O' Connell et al. [48] found that the best predictor of the student's final grade in a college algebra course was the prior performance in past courses.

1.3 Contributions

The work undertaken in this thesis started initially as an attempt to produce a practical model for knowledge graph creation. During the research on the related work, we spotted a new opportunity as we wanted to create a flexible model that would work in many different cases and knowledge domains. The existing models do not provide a uniform framework of creating semantic representations and do not explore the different types of relationships between the knowledge topics. By understanding the characteristics of knowledge mapping techniques, the EduKnow framework could be built to make our own representation framework that can structure educational material. Hence, we realised that this thesis could contribute to more domains, at first by theoretically defining the knowledge representation requirements, which define the different components of a knowledge topic and the relationships between knowledge topics. Secondly, we use the theoretical foundations to create a framework with which we construct the knowledge graph for our use case. With this in mind, the main contributions in this thesis are:

1. The exploration of the most relevant techniques for the development of a knowledge graph representation. The discussion on the main themes related to these areas can be found in the background section.

2. The design and theoretical development of a novel framework. The mathematical notation and definitions which introduce the three different types of links and relationships, compared to a unified existing one. The notions of prerequisite, association and assessment link are introduced giving us the ability to create many sub-graphs in the main knowledge graph.

3. The specification and structure of the knowledge topics components. Each knowledge topic in the graph contains its knowledge components, the methodology, theory and solved example. This extra layer gives the Edy-Know framework the ability to represent knowledge in a multidimensional format, which identifies the steps to perfection and the ways to master each step at the same time.

4. The novel framework modelled via the RSL hypermedia metamodel for creating EduKnow graphs. The RSL framework offers an enriched model for knowledge graphs, which escapes the narrow limits of classical RDF representations.

5. A use case for mathematics based on the EduKnow framework to illustrate the potential interactions with the learners. This use case also serves as technical evaluation of the EduKnow framework.

1.4 Methodology

For the development of the theoretical foundations and the implementation of the EduKnow framework, we are going to adopt the Design Science Research Methodology (DSRM) for information systems research [49] consisting of the following six steps: the problem identification and motivation, the definition of the objectives for a solution, the design and development of artefacts, the demonstration, an evaluation and the communication of the results. The problem identification and motivation includes the definition of the specific research problem that has already outlined in the problem statement and will be analysed in the next two chapters. The definition of the requirements and the development of the artefacts are well defined in the theoretical foundations of the framework, where a mathematical definition of the links is given as well as the construction components of the EduKnow framework. In the design and development phase, we will investigate an innovative hypermedia-based format for the encoding of knowledge representation with the usage of the EduKnow framework. The demonstration is taking place in the form of a use case for the mathematical area of algebra. There will be presented the evaluation and benefits over other models of the new innovative EduKnow framework in a concrete everyday use field in mathematics education. Finally, the communication of our results is happening via this thesis.

1.5 Thesis Outline

In Chapter 2, there is the presentation of the predecessors of knowledge graphs, and linking techniques used in the past. Starting the background analysis from knowledge mapping techniques as they were introduced multiple techniques. The mind maps and concept maps are analysed in detail and there is a brief presentation of the properties of their semantic representations, their structural characteristics.

In Chapter 3, related work is described in two parts. At first, there is the presentation of concept graphs and knowledge graph techniques and an analysis of their characteristics. There is also a brief discussion about the linking techniques they are using to create their knowledge representations. By the end of the first part, there is a comparison exploration of the techniques that have been introduced so far. In the second part, there is a brief introduction to the knowledge representation techniques for educational material that are being used in the private sector. The Mathspace[3] and Aleks[4] systems are presented in more detail.

Right after a comprehensive description is carried out for the creation of the EduKnow framework and its requirements. The main components are introduced and an innovative linking model is being formulated by mathematical foundations based on the requirements definition, in Chapter 4.

In continuous to the abstract model, the implementation of the EduKnow framework is addressed. The implementation has been realised based on the RSL hypermedia metamodel [54] that is used to formulate the components of the EduKnow framework. Also, the visualizations for the different linking techniques as well as a comparative analysis of the related techniques is made, in Chapter 5.

[3] `https://mathspace.co`
[4] `https://www.aleks.com`

Based on the EduKnow framework implementation, Chapter 6 presents a use case created for the topic of algebra in primary and secondary school educational material. The model creation is motivated and the resulting knowledge graph is analysed for potential applications in teaching.

Finally, future work for knowledge representation via the EduKnow framework and linking techniques is explored in Chapter 7, and a summary of the undertaken work is made in Chapter 8.

2
Background

Knowledge is an abstract notion that has been under debate since ancient times. In the state of the art, we find many studies which have been researching knowledge from a structural and organisational point, and creating semantic representations of knowledge. We can find many academic works discussing knowledge structures and representations, accompanied with a lot of terminologies. Although some of these are referring to knowledge extraction and structure techniques unrelated to education or teaching material, we present some of the techniques which are interested in structuring knowledge.

At first, *knowledge representation* [17] is referred to the format of the representation of the knowledge in artificial intelligent systems as described by Brachman and Levesque in 1985 [8] for extracting information, as the principles of knowledge representation are fundamental to work in natural language processing, computer vision, knowledge-based expert systems, and other areas. As Brachman points out, artificial intelligence needs *"descriptions of the world in such a way that an intelligent machine can come to new conclusions about its environment by formally manipulating these descriptions"*, therefore artificial intelligence requires more detailed mapping techniques than what had been already used in other fields. Examples for knowledge representation and reasoning include semantic nets, and ontologies, while many frameworks and functional approaches have been developed for representation systems [45, 37].

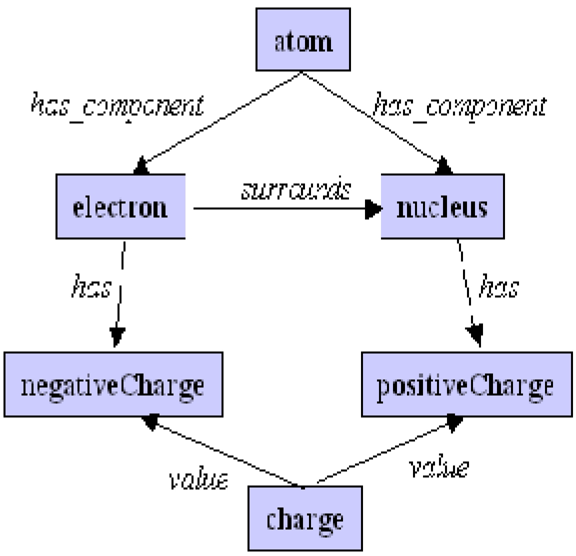

Figure 2.1: Ontology Graph for a fragment of Teaching Unit as presented by Gupta et al. [27]

Also, ontology-based systems have been used for the representation of educational material. An ontology is a *"formal explicit specification of a shared conceptualization"* [24], where the conceptualization is *"an abstract model of how people think of things in the world, usually restricted to a particular subject area"* [25]. Computer science is using ontologies as a model which describes the world that consists of properties, relationship types and objects. Gupta et al. [27] use ontology services in curriculum development. Their model consists of the science curriculum ontology which divides science education into the subparts of settings, clusters, sections and teaching units. Their objects are mapped with ontologies through a set of mapping predicates such as *example_of*, *discovery_of* and *description_of*. They create an ontology graph as shown in Figure 2.1. Their attempt tries to break down the complex notion of knowledge to simple objects which they are connected with a set of connections to ontologies. Although, as it is also clear from Cassel et al. [12] the ontology computing has applications in the education sector, however, the ontologies are extracting their hierarchical structure from content tables of textbooks and curricula, which they define the sequence of how the knowledge is being taught. Often the knowledge is structured differently from the sequence it is being introduced in a curriculum. Figure 2.2 shows the visualisation result of the relations between ontology nodes with various

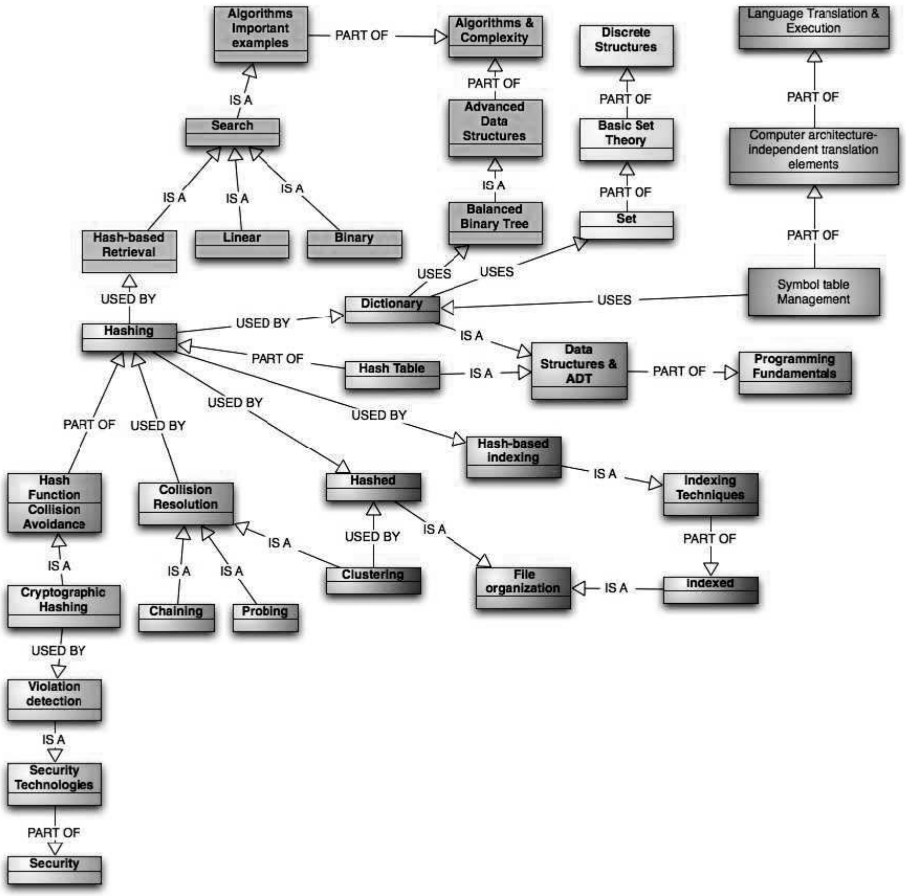

Figure 2.2: Visualisation of the place of hashing and its relations as presented by Cassel et al. [12]

semantic meanings and their relations as presented by Cassel et al. [12].

Knowledge management techniques have been using knowledge as an organisational asset to achieve competitive advantages in the private sector within companies [58], as knowledge-based systems, data mining and in other application domains [38]. They are used to categorise and identify knowledge assets such as people and technology. These techniques have been applied to education with no interest in linking the knowledge that is delivered to the students through the teaching process. They are rather treating knowledge as a key component for the business success of the educational institution. Their design is focused on knowledge for linking people, processes, and technologies, managing and sharing expertise [50].

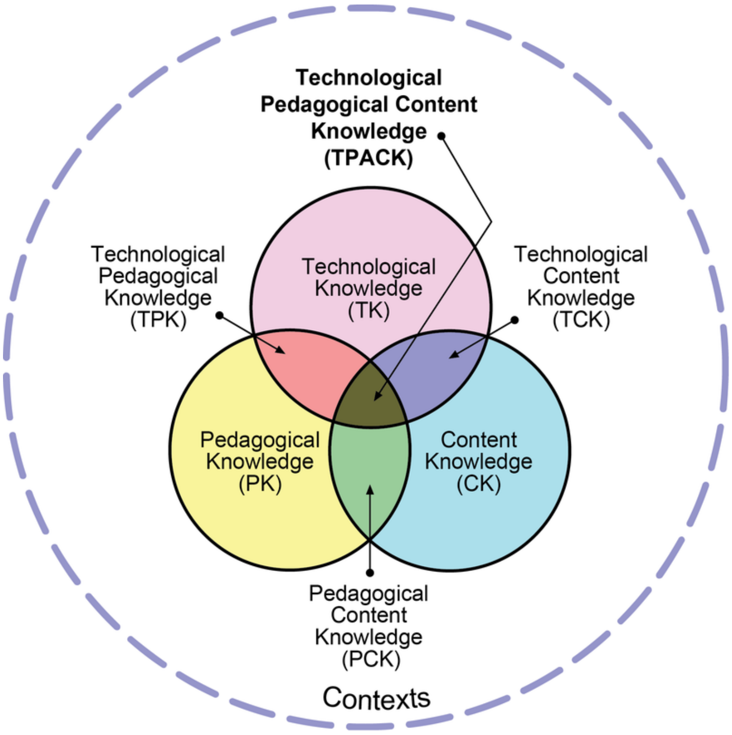

Figure 2.3: Venn Diagram of Technological Pedagogical Content Knowledge [34]

Towards knowledge representation for educational use, there is the idea of pedagogical content knowledge. Gudmundsdottir and Shulman [26] were concerned that trainee teachers could know about their subject (content knowledge) and about teaching (pedagogic knowledge), but still have problems in teaching their subject in an appropriate way. He called the necessary knowledge for doing this *Pedagogical Content Knowledge (PCK)*. He considered PCK as the knowledge of how to teach, not in the abstract or in general, but in the specific context of a subject or discipline. Semantically, the representation is done with the usage of Venn diagrams, where the main circles are representing pedagogical knowledge and content knowledge, and the overlap representing pedagogical content knowledge. In extend to PCK, the idea of *Technological Pedagogical Content Knowledge (TPCK or TPACK)* builds on the work of Gudmundsdottir and Shulman. Their development is the addition of a new circle in the Venn diagram, which represents the technological knowledge as it can be seen in Figure 2.3 [34].

On the other hand, knowledge mapping techniques have been used for specifically aiding learners to acquire a deeper understanding of a knowledge domain.

2.1 Knowledge Mapping

Knowledge mapping is another knowledge representation technique which uses semantic representations of knowledge. These representations typically include the important concepts that are represented in squares, ovals, or circles and connects them with different relationships via a single line. The connections can be unlabeled and thus they represent mere associations without specific relation, or they can have given labels of any type that specify the relationship between the two connected concepts. There are several knowledge mapping techniques and each one uses its own structural model to represent knowledge, although all of them give freedom to learners to add material (concepts and links) whenever it fits their goal. Hence these techniques do not provide strict guidelines of how a representation should be made, what is the content of a concept and many others as we will discuss later. In the book 'Mapping Biology Knowledge' [22] the authors are presenting several metacognitive tools for knowledge mapping in the educational domain of biology. These tools are the cluster maps and webs [42], the mind maps, the concept circle diagrams [59], the semantic networks (SemNet) [56], visual thinking network (VTN) [41] the conceptual graph [44] and others. In the interest of this thesis are the techniques that have been more popular in development, have more research impact and provide some structure in their format. Hence, we will focus our analysis on the techniques that also have more applications in the educational domain of mathematics which are the mind maps and concept maps.

2.1.1 Mind Maps

Mind maps were formally introduced in 1974 in the book *The Mind Map Book*. The book is introducing a revolutionary system, for that time, of planing that can improve the learner's memory and learning skills [11]. A mind map is a knowledge representation technique that captures the associations between ideas. It started by Tony Buzan as a note taking technique and turned out to be a way to capture and reflect the processes in the brain. The mind map format emphasises in increasing the creativity and performance of the learner. Also, mind maps are a useful tool for capturing and analysing complex sets of ideas [10], and they can work as a platform which summarises ideas of several students together for a specific content [9]. The method of mind mapping is taking into account both halves of the brain and

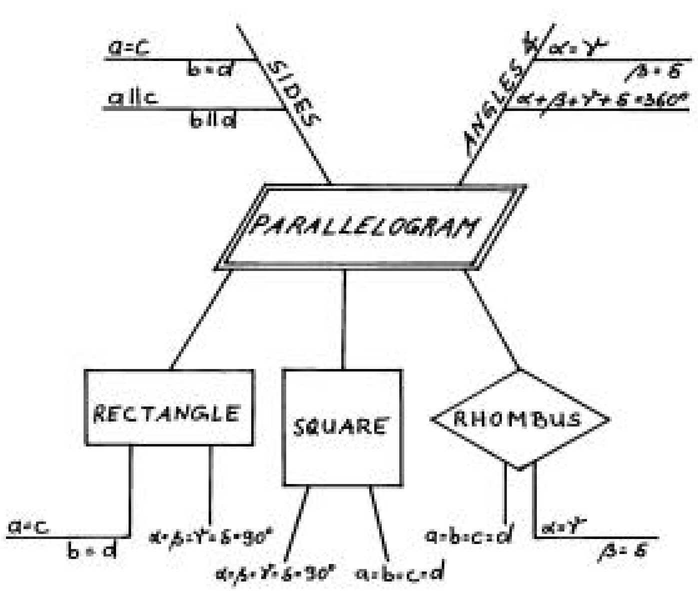

Figure 2.4: Mind Map example with the central topic of the parallelogram as presented by Brinkmann [9]

puts them to work together in order to increase productivity and memory retention. Hence, in order to trigger the right side of the brain, which is mainly responsible for creativity and art, many times the mind mapping will escape the narrow borders of classical semantic representations of knowledge and will include different shapes, colours and artistic pictures. Therefore, mind maps are by far the most artistic knowledge representation, that provides to students a lot of freedom in terms of what they are going to represent and how. However, in order to achieve its goal, a mind map needs to be aesthetically nice. Figures 2.4 and 2.5 show two examples of mind maps creation. As we can observe there are many differences between the graphs, as the representation is highly dependant on the artistic skills and style of the user and not so much dependant on the content.

In order to create a mind map, we need a large sheet of paper. The subject under analysis is placed in the middle, and from it, we draw branches from related ideas and concepts to it. We write down the ideas and concepts as keywords and draw lines. As a general principle, the concepts should go from the abstract to the specialised from the middle to the edges of the paper. The colours, symbols, sketches and shapes are in the preference of the user.

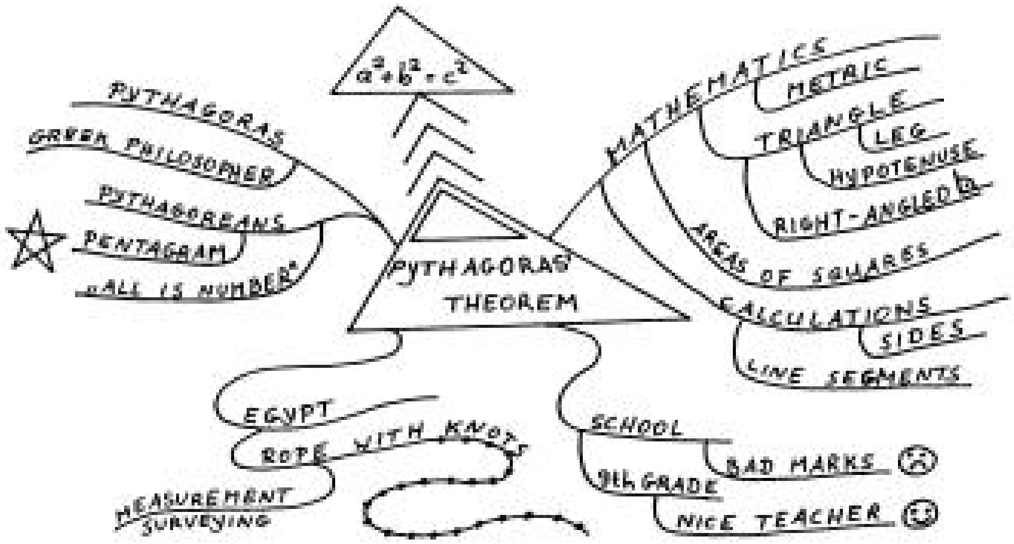

Figure 2.5: Mind Map example with the central topic of the Pythagoras theorem as presented by Brinkmann [9]

2.1.2 Concept Maps

Novak [46] is the creator of concept mapping, a useful tool for science education which can organise and graphically represent knowledge. It is one of the most widely used methods of knowledge representation in science education in the US and around the world. Concept maps facilitate as memory aid of meaningful learning. This model emphasises on the importance of the connections of the individual knowledge construction of each student. It allows the students to start from a concept and unfold their knowledge connections with anything related to the concept they are examining. They can represent the concepts and their components as boxes and use directed arrows or lines to connect them. Each connection can be labelled with no strict restrictions on the nature of the correlations that can be present in the concept map. When students are asked to construct their own personal concept maps they find new meanings in the subject they are studying and new ways to relate what they already know to the new things they are learning [47]. It is often the case that two students starting from the same concept can construct two completely different concept maps, as the resulting concept map is often a product of their understanding of the concept and the connections the students have obtained with other concepts or material. Studies have found that concept mapping has large positive effects on student attitudes, and a

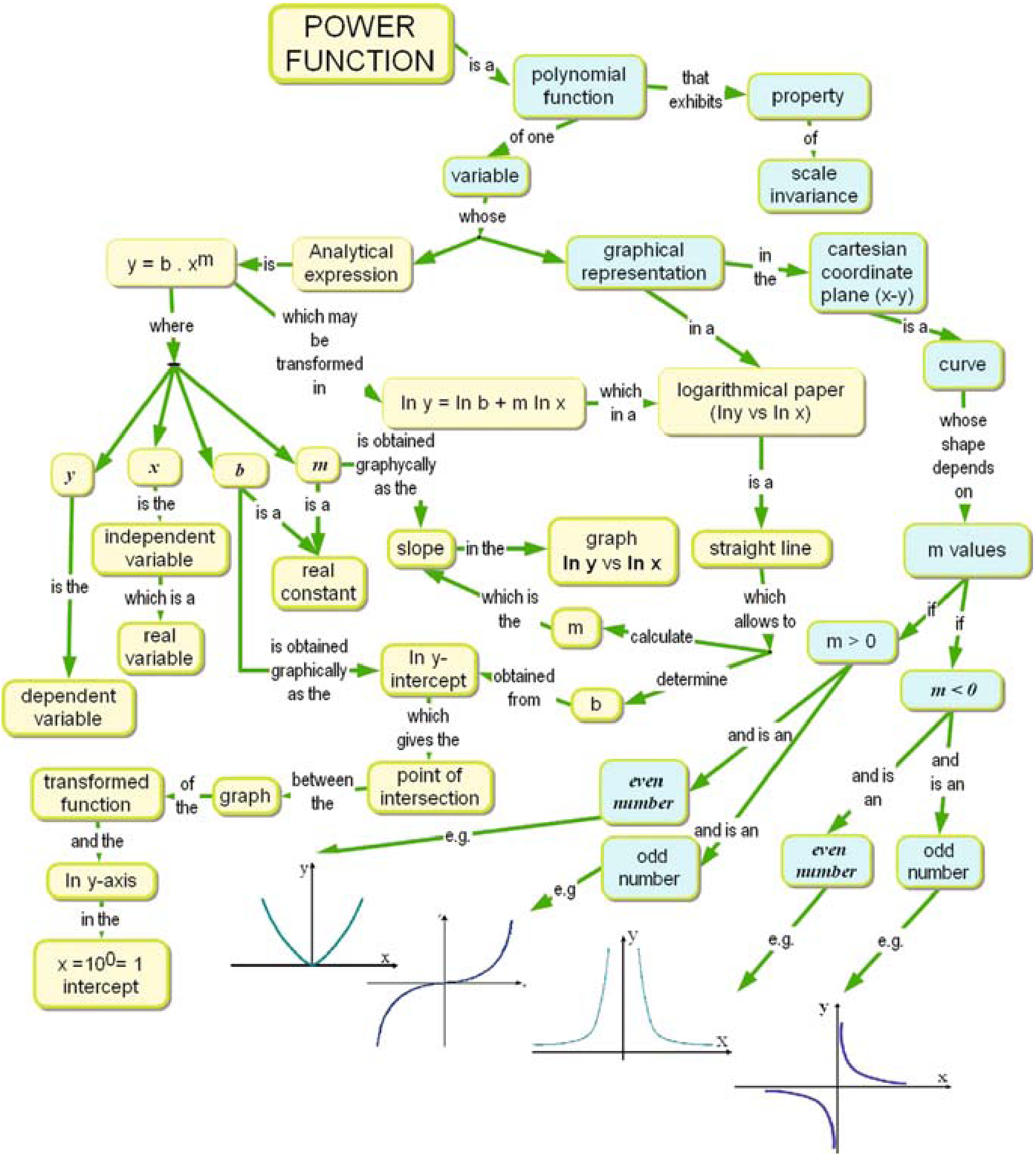

Figure 2.6: Concept Map example for mathematics [2]

medium positive effect on student achievement [31]. Furthermore, it assists students in achieving meaningful learning as it helps them visualise, organise and distinguish the concepts by their importance [43].

In order to create a concept map, we need a large sheet of paper. At first, we place the subject under analysis at the top of the map. We continue by arranging the related concepts that are going to be present on the concept map on several levels beneath the main concept. We draw lines to connect the concepts which are related and we label their relationship at the same time. In the last row of our concept map, we can present some examples to the concepts situated there. Figure 2.6 shows an example of a concept map for mathematics.

Mind maps and concept maps are used by individuals to create mostly manually knowledge representation. However, they do not offer extensibility and scalability features. In the next chapter, we are presenting two more knowledge representation techniques that deploy on a bigger scale and create graphs in a much more uniform manner.

3
Related Work

3.1 Graph Techniques

The techniques discussed in the previous chapter manually construct semantic representations of knowledge. In this section, we are presenting two more techniques which are created automatically via algorithms.

Besides representing the knowledge concepts in a semantic way, there is an interest in the different ways these concepts are linked together. The previous techniques put no constrains on the relations that can exist between concepts, thus there are unlimited types of links that can be represented. However, this creates an issue if we want to develop a model that will enable an uniform representation of any educational domain, and even more if we want a representation on top of which we can study the relationships between the concepts. Some research towards defining relationships between data in large data sets has led to the association rules. Association rules have been introduced in 1993 [3], with the intention to make an algorithm finding transactions in a database [4]. The goal of these rules is to generate all association rules based on the probabilistic metrics of support and confidence [5], of the form $X \to Y$. These models are studying the probability of the interconnection between different entities in a database [57] without specifying the nature of these relations.

The following techniques emphasise in the description of a specific association and introduce the most important type of linking, the prerequisite link.

3.1.1 Concept Graphs

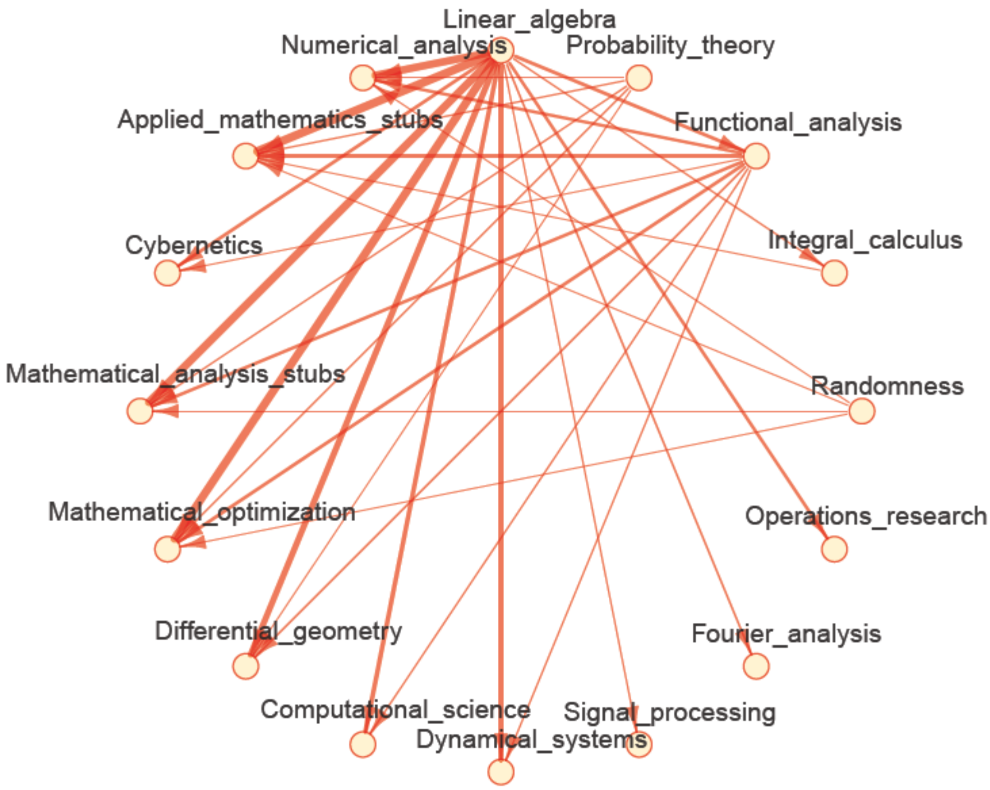

Figure 3.1: A visualisation of concept graph produced by CGL [40]. Each node denotes a concept, and the strength of each link encodes the prerequisite strength between a pair of concepts

A concept graph is the result of the automation extraction of concepts and courses as introduced by the Concept Graph Learning(CGL) system [60]. The concept graph contains a network of concepts, which represent the knowledge topics or domains, and a network of courses from educational institutes. It takes advantage of the huge amount of digital educational material that is available online, in order to train their algorithm and make better predictions on the test data. The implementation of CGL is based on the prediction of prerequisite links among concepts and courses. They use three different optimisation methods to predict the score of the prerequisite link. Their definition of prerequisite relationship is the existence of many directed links from a concept to a course. In Figure 3.1, there is a concept graph as it is produced by CGL, and in Figure 3.2 there is an example of prerequisite courses recommendation on Coursera[1].

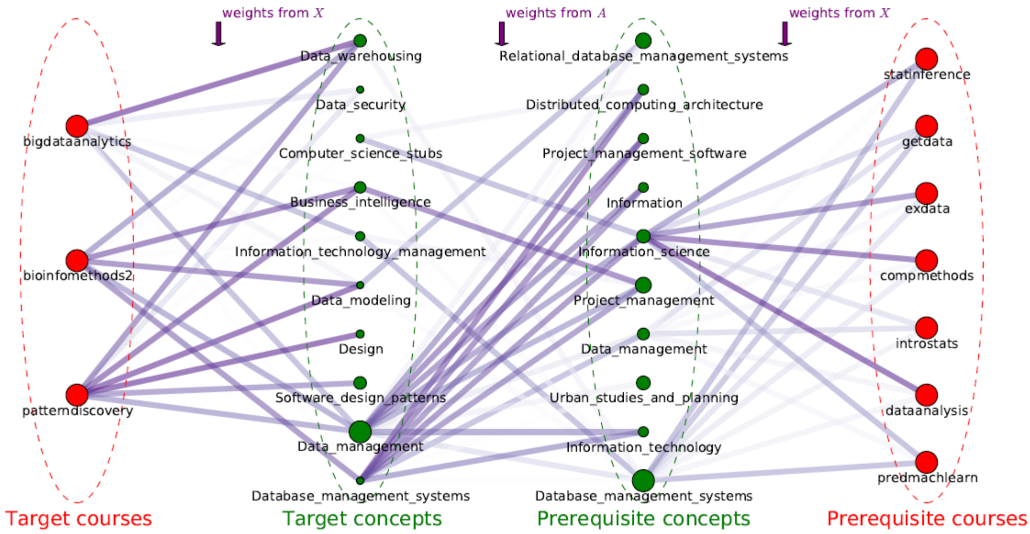

Figure 3.2: An example of prerequisite course recommendation on Coursera using the concept graph by CGL [40]. They map the courses (red, left) that the student wants to learn to the concept space of Wikipedia categories, the prerequisite concepts and then map back to Coursera courses (red, right). The sizes of the concept nodes in the middle (green) are proportional to aggregated weights of the corresponding links, and the strengths of course-concept mapping and concept-concept prerequisite relations are shown by the colour intensity of edges (purple).

Based on their work, the Prerequisite Structure Graphs (PSG) [13] are created. PSG works based on a generic algorithm, which is used in educational material and student activity data to construct an acyclic graph. The PSG is an unsupervised method. In the constructed graph the nodes are educational concepts and the links are directed arrows which define a prerequisite relationship. The PSG follows the prerequisite definition that is found on Doignon and Falmagne [18] and the book *Data Analysis* [32]. In the book, there is a mention about prerequisite links definition. Starting from the research in Didactics of Mathematics, the French professor Regis Gras at the University of Nantes suggested in 1979 a new methodology to approach the following question: *"If a question is more complex than another, then every*

[1]`https://www.coursera.org`

student who answers correct in the more complex one, would their answer be correct to the simplest one too?". Proposals like this, are formulated by the mathematical notation in logic: $a \Rightarrow b$, having the known characteristics of mathematics implication. As it is known, the previous statement stands true for most of the cases, however, exceptions can be found. Hence, the bigger the probability of not finding an exception, the stronger the mathematical implication.

3.1.2 Knowledge Graphs

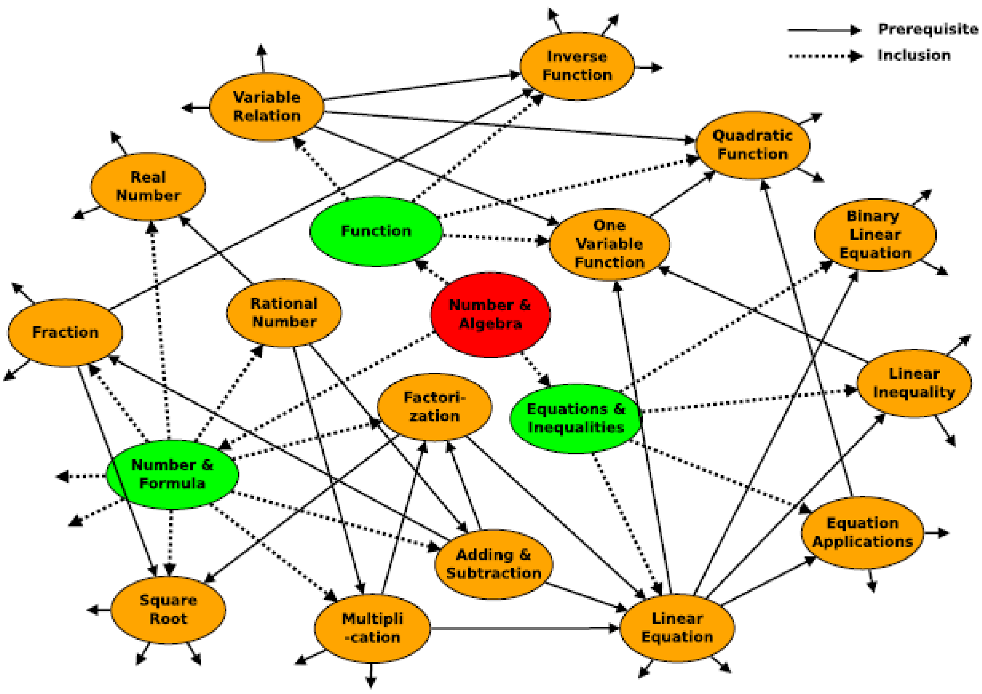

Figure 3.3: A snapshot of built knowledge graph for mathematics by Chen [15]

Knowledge Graph was introduced as a term by Google in 2012, which is a graph database that enhances the value of information returned by Google searches. As a terminology, knowledge graphs have been used also in education, where they handle knowledge from the view of the educational material. The automatic creation of a knowledge graph for educational purposes has already been suggested in KnowEdu by Chen in 2018 [15]. The KnowEdu system first extracts the concepts of subjects or courses, and then identifies the educational relations between the concepts, based on a neural sequence labelling algorithm and probabilistic associations. The knowledge graph is then created with two types of links, the main prerequisite and the inclusion link as can be seen in Figure 3.3. The prerequisite link is defined as a combination of the prerequisite link identification techniques that have been described previously. At first, their model is based on the association rule that was described at the beginning of this chapter [3]. Secondly, they use the PSG method and define that if the concept i is prerequisite of j, then the understanding of concept j implicates the understanding of its prerequisite concept i. Moreover, the poor understanding of concept i leads to poor understanding of the more advanced concept j. Both methods use the probabilistic metrics of support and confidence as thresholds to define whether a prerequisite relationship exists or not. The inclusion link is briefly introduced as the relation between two concepts which indicates that a concept belongs to another one, and they identify it based on the table of contents of a textbook. Thus we can say that the inclusion link is poorly identified because based on the textbook it will contain different inclusion relationships. Moreover, the textbooks are usually part of a curriculum and they indicate a learning path, a sequential way to teach knowledge topics, which can considerably differ from the way the knowledge is structured itself.

Also, the K12EduKG system [14] uses the same technique as KnowEdu to construct the knowledge graph that focuses on subject concepts rather than courses, with the objective to aid the flow of teaching and learning rather the definition of course dependencies. The K12EduKG is focused on the knowledge topics of K12 class and constructs a knowledge graph with nodes subject concepts of K12. For linking, it uses the prerequisite link as defined in KnowEdu.

3.1.3 Comparison of the Techniques

The following Table is a summary of the techniques that have been analysed in the last two chapters. Some of the entries in the table are obtained from the paper of Epper [20]. The goal is to have an overview of the characteristics of all the different techniques together so we can conclude on the benefits and disadvantages of the existing models. In the beginning, the basic characteristics of the techniques are displayed, as they were discussed in the previous sections.

	Mind Maps	**Concept Maps**	**Concept Graphs**	**Knowledge Graphs**
Creation Method	Manual	Manual	Automatic	Automatic
Format of nodes	Central topic bubble and coloured branches with text, colour and sketches	Ovals or boxes with text	Ovals with text	Ovals with text
Format of links	Straight or curved lines, labelled or unlabelled, directed form the centre of the page to the edges	Mostly labelled connector arrows, sometimes directed	Directed links unlabelled	Directed links sometimes labelled
Number of different links	Unlimited links that can define any type of relation between the concepts	Unlimited links that can define any type of relation between the concepts	One	One or two
Focus group	Individual	Individual and learning groups	Educational sector	Private sector
Reading direction	Center-out	Top-down	Any	Any
Application context	As a note taking technique which is based on creativity and increases memorability and learning skills	As an aid to the individual for deep understanding and in the classroom as a tool to increase learning skills of students	Result of text processing that aids to structure educational data	Database formulation as the core of the application development in companies
Scalability	Limited	Limited	Unlimited	Unlimited
Difficulty of creation	Low	Medium to high	High	High
Extensibility	Open	Limited	Unlimited	Unlimited
Memorability	Medium to high (increased by artistic elements)	Low to medium	Very low	Very low
Understandability	Low	High	Medium	Medium

The order in which the techniques have been presented in the background and related work section it is not a coincidence. The format starts from a very general model with no constrains that allows any kind of representation without any requirements and goes on to the most specified and uniform model that exists.

All techniques have advantages and disadvantages. The main advantage of the mind maps is the independent creation that allows users to develop any kind of representation depending on their preference, like a note taking activity. The concept maps allow a user to create openly a map similar to the mind map, however, it gives the representations more structure and increases their readability and understandability, which makes them easier to use in a classroom. On the other side, the graph techniques can be extended unlimitedly and have great scalability features. However, they are not recommended for individual usage as they are very difficult to produce and do not offer characteristics for individual use. Between the two graph techniques, the knowledge graph stands better because it offers an enriched model compared to the concept graph, due to the extra link that it uses.

3.2 Commercial Products

Many frameworks exist for the construction of knowledge representations. The majority of these frameworks is made for concept maps creation because it is the most popular among the presented techniques. However, we do not proceed to any further analysis of the characteristics of these frameworks, as they create their features based on the techniques that were introduced in the background and related work sections. Our focus is on the key elements of the semantic representations of the knowledge and characteristics of each representation, rather than the tool they are created with. Many websites use educational material as assessment-based software to help students improve their knowledge. Most of them, like Math10[2] and Khan Academy[3] for mathematics and courses based on the US curricula, or the Belgian usolv-it[4] for competitions like Kangaroo and Olympiads in mathematics and chemistry, help learners without having a complex recommendation system. They check the correctness of the student's answer and keep recommending assessments on the same topic (or level) until the student chooses otherwise. Their knowledge topics are usually chosen as chapters from a curriculum, from which they obtain their structure. Besides these simplistic implementations,

[2] `https://www.math10.com`

[3] `https://www.khanacademy.org`

[4] `https://www.usolvit.be/servlet/home/index.action`

there have been a few commercial products for educational purposes that use adaptive learning to provide aid to learners with more complex assessment recommendation algorithms. All of them use their own knowledge graph databases which they have created based on some educational material.

3.2.1 Mathspace

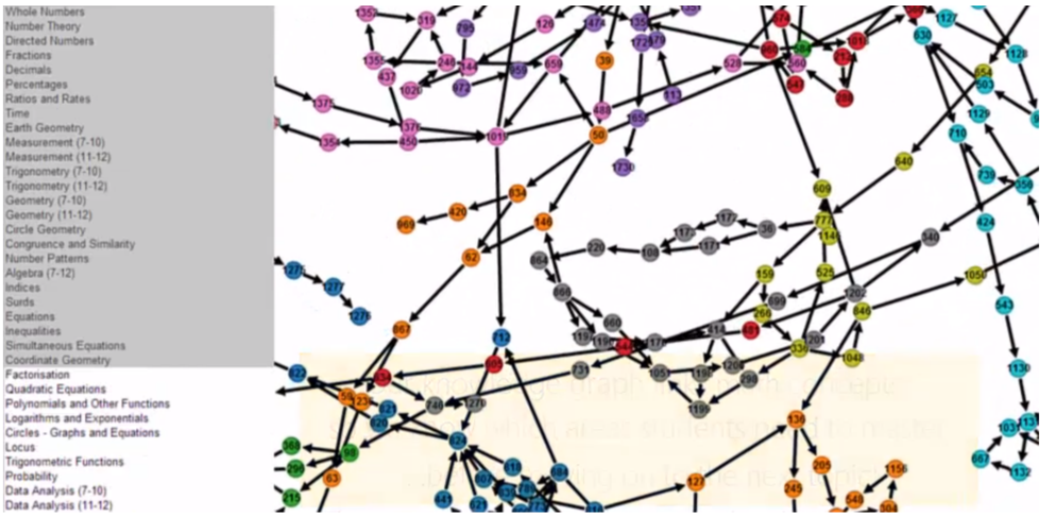

Figure 3.4: Part of the knowledge graph of Mathspace obtained from the website. The topics are numbered and represented as nodes. They are linked together with a single directed link.

In the private sector Mathspace[5], an Australian company working on the US mathematics curriculum, is helping students by suggesting exercises to be solved, based on a user's solving abilities and its own created knowledge graph. Moreover, Mathspace uses a knowledge graph created with nodes as subject concepts based on US courses. It uses directed links, however, it is unclear how the linking between the nodes is defined and what it represents. The knowledge graph works as expert material that shows the next and previous steps of knowledge for a learner. Hence, we can conclude that the representation of the knowledge is the key element in the recommendation of assessments and in the calculation of the students' knowledge gaps. We should also notice that Mathspace uses as knowledge nodes subjects from curricula, such as *Measurements 7-10*, as it can be seen in Figure 3.4. It links them with the sequence they are presented in the curricula so that the

[5] `https://mathspace.co`

knowledge node *Measurements 11-12* comes after the *Measurements 7-10*. Therefore, the knowledge structure is based on the learning paths the curricula are following and not on how the knowledge topics are linked together by their definitions.

3.2.2 Aleks

Doignon and Falmagne in their book *Knowledge Spaces* [19] are presenting a theoretical framework, which proposes mathematical formalisms to operationalise knowledge structures in a particular domain. They are defining their theory based on two key concepts, which are *"the "knowledge state", a subset of problems that some individual is capable of solving correctly, and the "knowledge structure", which is a distinguished collection of knowledge states."* Based on these two concepts the mathematical foundation of Assessment and LEarning in Knowledge Spaces (Aleks)[6] is built. Aleks is a web-based, artificially intelligent assessment and learning system, which uses assessment recommendation to determine a student's level of expertise and knowledge gaps which is developed based on the US curricula. Aleks is used for mathematics, chemistry, and accounting in schools and universities. It uses the concept of knowledge state from Doignon and Falmagne to determine the expertise level of the student and their knowledge gaps. It claims to accurately define the knowledge state of a student in the whole knowledge space of the educational topic. Aleks uses an algorithm based on AI to calculate the next assessment for a student based on all the previous answers of that student. This way it determines the mastered and not yet mastered knowledge and what the student is ready to learn next. Therefore, Aleks only presents topics the student is capable to learn next. An interesting fact is that Aleks also 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, however, we know that the nodes represent subjects from the US curricula. The linking between the nodes is made with the precedence link [21], which is the previous topic that was taught based on the learning path of the curriculum that the students is (possibly) following and not any prerequisite relationship like the other knowledge graph techniques.

The problem with these previously mentioned techniques is that even in the scientific and business world there is a confusion between the knowledge graph and the learning path. All the previously mentioned techniques created learning paths based on the specific curriculum of the books and material

[6] `https://www.aleks.com`

they had available. The knowledge itself is often structured differently from what a learning path is suggesting as the next topic to be taught.

4
EduKnow Definition and Theory

Education has been well studied by didactics, pedagogical and social sciences. The EduKnow framework aims to reveal the inner structure of the knowledge topics, by handling the core of educational material and teaching process, which is the learning outcomes. This approach is developed with the focus on the learners and their deep understanding. The EduKnow framework defines rules of how the knowledge topics are related and finally creates an abstract model of a graph representation for an educational domain.

While there is plenty of educational data online, they still remain unstructured and not well connected. We find that students are being taught similar courses during the standard educational policy in the European Union and around the world, until the age of 15 or 18 years old. Hence, the high-school diploma is a guarantee that a student has reached a specific level of expertise, which is relatively similar to the level of expertise another student has achieved by finishing high-school in countries with the same educational standards. Especially, in the main educational domains, like mathematics and physics, the level of knowledge for specific knowledge domains is the same between different countries with not crucial differences between curricula. Hence, there is an interest in trying to organise the knowledge topics up to this expertise level. By studying the knowledge topics presented up to the level of high school education, as well as their connections and their content, we can view their detailed characteristics and structure. Moreover, since there is a final knowledge level which students are reaching by graduation, it is easy to assume that we can make a representation that will contain all the knowledge information as it is only a limited amount of knowledge topics per knowledge domain.

The goal of this chapter is to create the foundations of a framework which can structure educational material in any educational domain. Educational data have common characteristics as they usually form hierarchical structures going from the simplest knowledge topic to the more advanced, and from the earliest event to the latest. With that in our mind, we can construct a knowledge representation where knowledge topics exist in a directed graph. The graph consists of nodes, which represent the knowledge topics and contain their components, and links, which are the relationships between the knowledge topics.

We need to have a well defined and strict structure so it will be able to be used in the same way for every learner, and at the same time enriched so it contains all the necessary elements to represent all knowledge components. The model needs to be theoretically defined so it has an abstract foundation where a researcher can realise its use and importance. Furthermore, a theoretical foundation of our knowledge components and their relationships is necessary if we thrive to achieve a reproducible framework, which can deploy a knowledge representation for any educational domain, and have the ability to construct the same representation every time it has the same data as a starting point.

Our aim is to create the basis for an adaptive learning recommendation system. In the core will be the EduKnow framework containing a knowledge graph representations for the educational domains. Based on these knowledge representations it would be possible to program learning paths of different curricula and also to track the knowledge gaps of learners. The difference between the adaptive learning recommendation assessments systems with the simplistic existing ones, is that the adaptive learning will use the knowledge graph forwards (in the direction of the directed links) to find the next topic; as it is done already; but also backwards (opposite to the direction of the directed links) to discover the knowledge gaps of a student. By taking into consideration the mistakes a learner is making, we can probabilistically calculate the previous topics that might have been responsible for the lack of a learner's performance. Moreover, we aim to visualise the knowledge structure so that the actors related to the learning process are able to view and understand in depth the reasons behind a student's performance and the way to success. We consider this one of the main contributions of novelty to the knowledge representations and knowledge structures.

4.1 Requirements

In order to proceed with our design, we must first identify the necessary requirements. For the construction of our model, we need to have an educational domain we construct the framework for, such as mathematics. Right after, we can choose one or multiple knowledge domains for our development, such as algebra, geometry and statistics.

The next step is the identification of our main component, the knowledge topic. A knowledge topic, as defined previously, is a node in the EduKnow graph and it consists of the title of the topic at the top level, and the methodologies, theory and solved examples at a more detailed level. We view knowledge as the outcome of the learning process, what a student should know by the end of a chapter or academic year. Many times there is confusion in existing knowledge representation techniques about what to represent as a node in the knowledge graph. This is one more separation from the related knowledge mapping techniques. They often represent the nodes in their representation as knowledge topics with very generic content, such as *numbers* [15], or with a very specified content as chapters of a curriculum, *Consumer Arithmetic (11-12)*[1]. Therefore, it is of great importance to have already properly identified all the knowledge topics before beginning the linking process. Hence, this important task comes to the hands of an expert developer, who is aware of the educational domain and can formulate each knowledge topic, based on the specific learning outcomes.

Each knowledge topic has a title and at least one methodology, which will explain how an assessment can be addressed in an abstract way and illustrate it with a solved example. The methodology parts are often found as part of the introduction of a new chapter for a given subject. The methodology is a very important component. Studying different curricula we find that it is often the case that students are being taught the same topics, such as multiplication or solving quadratic equations, however, they have been introduced to a specific technique to address assessments (methodology) which can be completely different from one curriculum to another. Moreover, based on the focus of the curriculum, students are asked to perform certain tasks given a topic, for example, to perform 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, to be able to apply the theory and multiple

[1] `https://mathspace.co`

methodologies per topic. These methodologies can address different types of assessments, such as assessments solved by calculations and assessments solved by proof. It is 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 outcome on the knowledge topic, and 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 taking place during the teaching process.

4.2 Linking in EduKnow

As far as it is in our concern, linking techniques have no uniform method of representing knowledge. The focus is being on probabilistically guessing the possible transactions among the entries of a dataset based on the metrics of support and confidence. In contrast to the current techniques, our goal is to create a strict, uniform and simple way of representing all relationships between knowledge topics by using only three types of links; the *prerequisite*, the *shared content* and the *assessment* link. We introduce the linking model based on theoretical foundations in order to have a scalable model which can be easily used in multiple domains.

In our model, we find three types of relationships between the knowledge topics from which we create the three links. We identify two links that refer to relationships between knowledge topics as content. We find only two links in our model as under some research we spotted that the prerequisite link is the most significant in terms of related work and obviously necessary to represent a hierarchy of data. Additionally, the second link (shared content) came to surface when the prerequisite link was not capable to construct a full knowledge representation graph alone. Hence, the new type of relationship was revealed with a bidirectional link that links knowledge topics together. The third link is the assessment link which comes as an outcome of complex assessments that require knowledge from multiple domains in order to be solved. During the linking process for the formulation of the knowledge graph representation, the knowledge topics are nodes represented by their title and linked via the three links that have different semantic illustrations. There is great interest in the assessment link, as it shows connections between knowledge topics which are not obviously remarked. Often, on exams students are asked to address assessments with a high level of complexity, which requires knowledge from multiple domains. Research on these connections could help students perform better and reveal hidden connections between knowledge topics and domains. Last but not least, although complex assessments have been studied [39], it is still a relevant new approach towards knowledge representation to use an assessment link between knowledge topics according to the published related work.

4.2.1 Prerequisite Relationship

The prerequisite relationship is the one that appears most often in the literature. However, in contrast to the existing techniques that use the probabilistic metrics of confidence and support to define it, we are using a strict mathematical model. Therefore, we are also counting on well defined knowledge topics in order to be able to identify properly the links and relations between them.

Therefore, the prerequisite relationship which defines the relationship between any knowledge topic A and B, (A, B), such as the A is a subset of B, then we define that A is a prerequisite of B. We also require the prerequisite relationship to have the principle of transitivity of implication which means that if A has a prerequisite relationship with B, (A, B), and B has a prerequisite relationship with C, (B, C), then the A has a prerequisite relationship with C, (A, C).

Example: The *1st degree equations* and the *2nd degree equations* are prerequisites of the *3rd degree equations*, because all the knowledge needed to study *1st degree equations* and *2nd degree equations* is required also in the *3rd degree equations*.

4.2.2 Prerequisite Link

In order to avoid redundancy on links, we define the prerequisite links only on the direct prerequisite topics, with which we can construct all the prerequisite relationships. The prerequisite link is directional and points from the simpler knowledge topic to the more advanced.

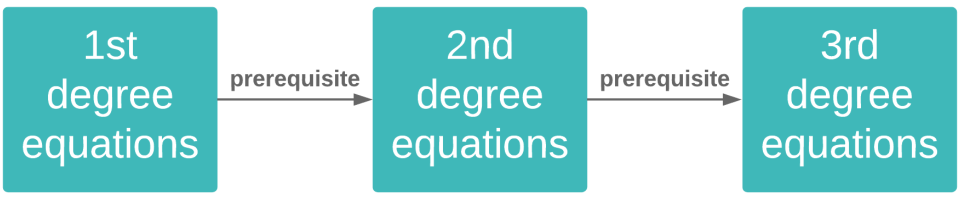

Figure 4.1: The prerequisite link is created only between the direct prerequisite topics.

This link comes from the prerequisite relationship between two knowledge topics and is defined such that there is a prerequisite link from A to B, if and only if, A is the greatest subset of B.

Example: The *1st degree equations* are direct prerequisite of the *2nd degree equations*, and the *2nd degree equations* are direct prerequisite of the *3rd degree equations*. The link is created only between the direct prerequisites. Therefore there is no link between the *1st degree equations* and the *3rd degree equations*.

4.2.3 Shared Content Relationship and Link

The shared content relationship is defined between two knowledge topics when they are not connected via a prerequisite relationship and they share some knowledge content.

We define the shared content relationship, such as A is sharing content with B, if A and B are not connected with a prerequisite relationship, and $A \cap B \neq \emptyset$.

The shared content link is non-directional. This link connects two knowledge topics if they are under a shared content relationship. There is no reduction policy on this link. This is happening due to the fact that different topics share different content, therefore it is wise to point out the relationship right away with a link.

Example: *Identities* are sharing content with *2nd degree equations* and *3rd degree equations*, because the *identities* are consisting of square, cubic, quadratic formulas etc, of which the square and the quadratic identities are necessary for solving *2nd degree equations* and *3rd degree equations* accordingly.

4.2.4 Assessment Link

An assessment can exist only in a single domain, or it can require knowledge from multiple domains. In the first case the assessment does not create any assessment link, while on the second case an assessment link is created from the knowledge topic the assessment exists towards each other knowledge topic the assessment requires knowledge from. This link identifies the connections between different knowledge topics that are linked together as knowledge domains of a complex assessment. The assessment link is non-directional. It exists only if the knowledge topics are not connected via a prerequisite or shared content relationship. An assessment group is formed when there are many assessments of the same type that require knowledge from the same domains. When a group of assessments requires knowledge from multiple

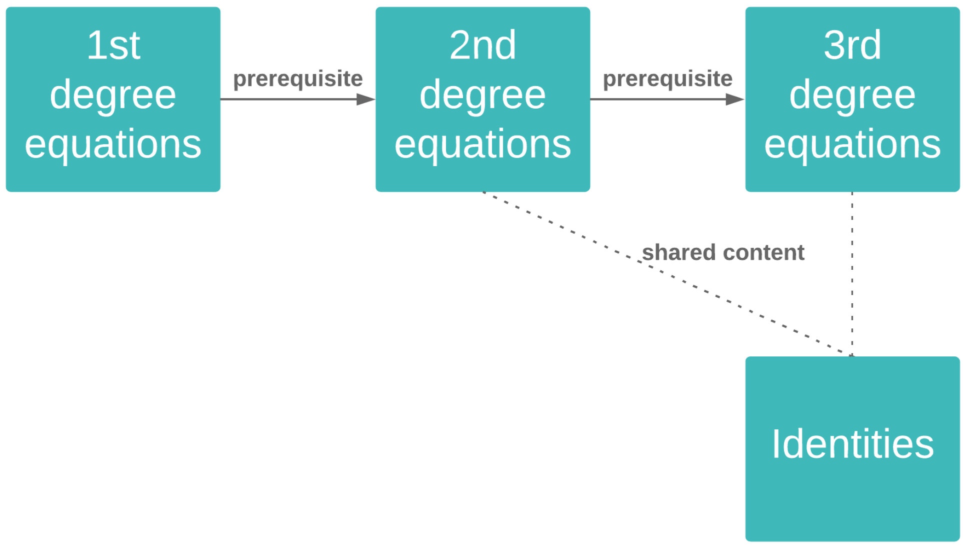

Figure 4.2: The shared content link is created if there is a shared content between topics that share content.

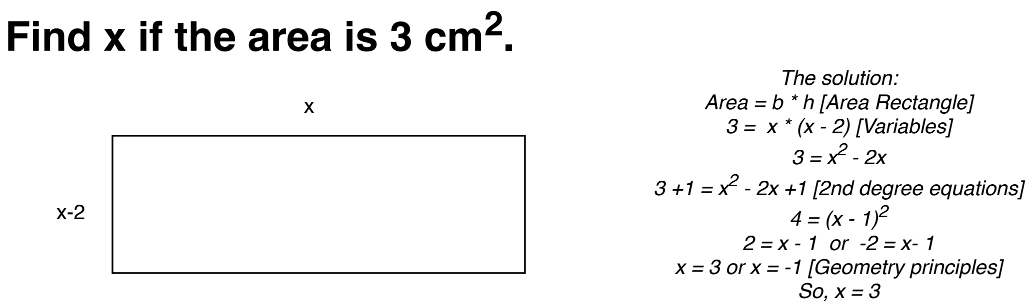

Figure 4.3: To find the right answer the student needs to possess knowledge from many domains.

topics, only a single link is created with source the knowledge topic of the assessment group and targets the required knowledge topics. Assessments and assessment groups exist as part of a knowledge topic.

Therefore, if an assessment (or a group of assessments) exists in the knowledge topic A but requires knowledge from knowledge topic B, then an assessment link between the topics A and B is created, if the two are not connected via a prerequisite or shared content relationship.

Example: Figure 4.3 presents an assessment that links to multiple knowledge topics. This assessment connects via assessment link the knowledge topics *2nd degree equations* and *Area Rectangle* as they are not in a prerequisite or shared content relationship, although their knowledge topics are linked via this assessment as shown in Figure 4.4.

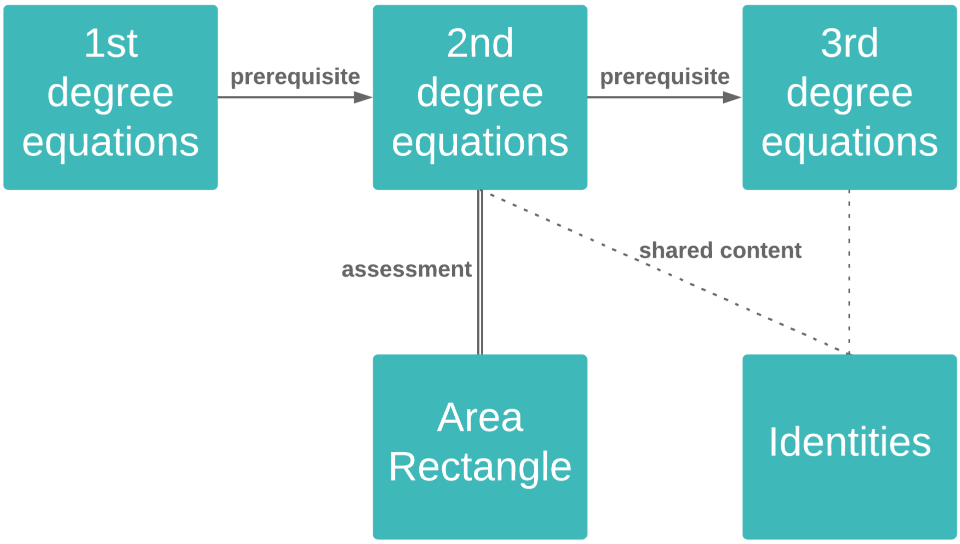

Figure 4.4: An extra link is added to connect the 2nd degree equations from algebra with the area of rectangle from geometry. This link would not obviously be included as it does not naturally come from the prerequisite or the share content relationship. However, due to the exercise in Figure 4.3 there is a connection.

The importance of the EduKnow framework lays on its definition. In contrast to the existing techniques, the EduKnow framework represents the knowledge nodes as the practical learning outcomes. This happens due to the methodologies that describe the approaches to address assessments. This novel approach handles educational data from the learner's point of view, and lets them know what exactly the should be able to do in order to succeed. Moreover, the EduKnow framework introduces a uniform way of linking between knowledge nodes with strictly formulated relations and links. We find this necessary in order to achieve high scalability and 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 EduKnow framework all we need is an expert to the field, such as a teacher, in order to first identify properly the knowledge topics, and then connect them with the proper links. By the time this procedure is completed, the resulting knowledge representation can be reused and enriched by others. The next experts can build on top of the existing one, by adding new methodologies or missing knowledge topics and links. The reason behind this reproduction is the fundamentals of the structure in the EduKnow framework. Since it is counting on structuring the knowledge itself without taking into account any learning path (e.g. textbooks or curricula) it creates the representation of how the knowledge exists. Therefore, this cannot be altered or modified from one institution to another. The only thing that can happen is that a curriculum may address assessments with a different technique, or offer extended courses that cover extra material that is not yet implemented in the knowledge representation of the EduKnow framework.

One more benefit of the EduKnow framework is the uniform and well structured linking method, which covers the possible connections that can occur between knowledge topics. This solution is well defined compared to many techniques in the literature that are not using a uniform way of creating relationships and links in their representation. Even the latest techniques that introduce up to two different types of links, often miss creating all the connections, as they are highly dependant on the educational material they are using. Moreover, their links are defined by the probabilistic metrics of support and confidence, which means that they can only identify a relationship if it exists many times in the data they are processing. The difference of the EduKnow framework is that it defines the links based on the content of the knowledge topics and not in the sequence they are presented. This allows us to structure the knowledge by content and not by the learning paths, which at the end creates a more accurate model that can be reproduced in any educational system whatever the leaning path is.

Another addition is the introduction of the assessment link. Assessments have been studied for performance based applications [39], have been classified in existing intelligent learning environments [36], and there are existing automatic systems for recommendations of assessments [29]. However, the 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, from which it has only the connections the knowledge topic has. These links can be any type, like in mind maps or single prerequisite type like in concept graphs, and therefore the assessment will miss the connections to the knowledge domains it needs in order to be addressed. This is also the case many students are facing, as they solve simple assessments but they cannot perform more advanced ones, because the advanced assessments require knowledge from multiple domains. In this case, these connections are not obvious to students. Hence they cannot realise that the assessment needs extra knowledge—and which—to be solved. This is due to the fact that students may have knowledge gaps that are difficult to spot and identify as they usually do not know what they do not know. Therefore, in the case of a complex assessment, a student cannot perform and does not know the reason. With the EduKnow framework, a learner is aware of the complexity of an assessment and all the required knowledge it needs. This approach is very important as it reveals the hidden connections of knowledge that exist on a practical level.

5 Implementation

In the previous chapter, we defined the theoretical foundations and requirements of the EduKnow framework. As described earlier, the purpose of the EduKnow framework is to model the knowledge based on its inner structure. The EduKnow framework defines a multidimensional graph that links knowledge topics in dimensions, at first with its components, the methodologies, solved examples and assessments, and secondly with other knowledge topics. Hence, we need a model which can support complex structures with rich content, which is scalable and can potentially link data on the Web. The reason is that there is plenty of educational online material that is free to access. Furthermore, most of the online materials are components of a knowledge topic that are not properly connected, therefore they remain unstructured. The EduKnow framework could be the basis of the interconnection of the knowledge components under a single model. Hence, for the implementation, we are using a model that can support the requirements of our framework, the resource-selector-link (RSL) hypermedia metamodel.

5.1 RSL Model

The Resource-Selector-Link (RSL) hypermedia metamodel has been introduced by Signer and Norrie [54]. RSL allows us to represent complex structures in a uniform and simple representation and enables us to link to external

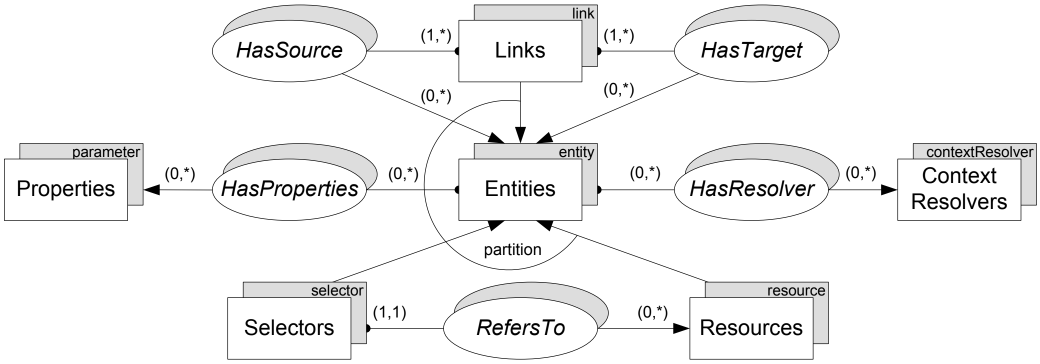

Figure 5.1: Core link metamodel by Signer and Norrie [54]

components (hyperlinks) with our local database. RSL has further proven its extensibility and richness for various document and hypermedia formats in the past [53] and we think that it is an ideal candidate to metamodel the structure and characteristics of the EduKnow framework. Our EduKnow framework implementation based on the RSL model will form a uniform architecture with a scalable and content-rich representation of knowledge that can be integrated into multiple cases, as an educational and learning product.

The RSL hypermedia metamodel is developed in Java and Gradle, and offers the formulation of the schema via a JSON (JavaScript Object Notation) file. For the EduKnow framework implementation, we created the JSON format that can be found in Appedix A.2. When compiling the code, the JSON file generates the Java classes for each instance. In each class, we can use the RSL predefined methods and further add our functionality by creating more methods that fit our model, as can been seen in the class diagram in Appedix A.1. For our implementation, we are using the database from ObjectDB[1]. We populate the database via POST requests on the server. For our model the HTTP requests are made through Postman[2].

In RSL the main components are the resource, the selector and the link, which are represented as entities as shown in Figure 5.1. The resource can be a topic or subject, such as a video. The selector allows choosing a specific part of the resource, such as the video frames starting from the 15th second until the 34th second. The link is the entity that connects resources, such as a video that is connected with the knowledge topic it explains.

The resource can be a knowledge topic as a whole, an assessment, a methodology as a whole, the theory of the methodology, or the solved example of the methodology. Besides our local database, the RSL hypermedia

[1]https://www.objectdb.com
[2]https://www.getpostman.com

metamodel allows us to use external resources such as educational material that we can obtain online. Hence, we can link a knowledge topic with videos, audio, and text that we can find online and create a rich model that takes into consideration all the different ways of learning [52] and gives the opportunity to special need students to have access to specific aid material, such as an explanation video in sign language. Considering the scalability and future extensions of the EduKnow framework, it is wise to use the RSL model in our implementation with which we can potentially link any external resource with our local database, and create a powerful tool with lots of content.

The selectors can choose a specific part of the resources, making it easy to select a part of the resource and not the whole object. The selectors are important to obtain specific information from the online data, which will provide the learners with the exact information they need and give a better experience to the users. Moreover, we can view the selectors as data labelling actors, that will declare which source is linked with a knowledge topic and create metadata that can be used for data mining processing and other purposes.

The RSL hypermedia metamodel introduces two types of links, the navigational and the structural links. For the EduKnow framework implementation we use the structural links to glue together the parts of the knowledge topic and keep the concrete structure in a single node in the knowledge graph. The navigational links identify the relationships between knowledge topics and create the knowledge graph structure with the connected nodes. Therefore, the prerequisite and shared content links are navigational links and have knowledge topics as target and source, while the assessment link is also a navigational link which has target knowledge topic(s) and as a source an assessment or a group of assessments, in which all the assessments have the same characteristics (type) and require knowledge from the same domains in order to be solved. On the visualisation level, we are representing the assessment link with source the knowledge topic the assessment or the assessment group exists.

At the first view a reader might wonder if the EduKnow framework could be implemented by the Resource description framework (RDF) model [35] which is widely used and capable of addressing a variety of problems [51]. The RDF has become a general way for conceptual description and a method for modelling information. It can use a variety of syntax notations and data serialisation formats, and it can be implemented in the semantic web. We can also find knowledge management applications of RDF triple together with ontologies, that are linking data creating knowledge representations [7] and information representations [28]. Therefore, it is logical to wonder why not use the RDF on our implementation.

The answer lays on the richness of RSL that comes from the JSON file, its structural model and its capabilities of creating and handling metadata. For increasing data volumes the use of JSON offers more successful management than RDF [33]. Also, a migration tool has been developed for the translation of the RDF triples to JSON formats proving the need for effective transitions from the RDF model to novel architectures. Moreover, the implementation of the EduKnow framework with the RSL hypermedia metamodel fulfils our goal for a framework which can include all the knowledge components of the learning process, support all types of educational material, and being able to provide a sufficient number of visualisations to the user. Also, in a later stage, our aim is to extend the current EduKnow framework by encoding the learning paths of different curricula on top of the created knowledge base. The EduKnow framework with the RSL implementation offers us a plethora of ways to program the learning paths in a detailed way with the usage of the metadata.

5.2 Visualisations

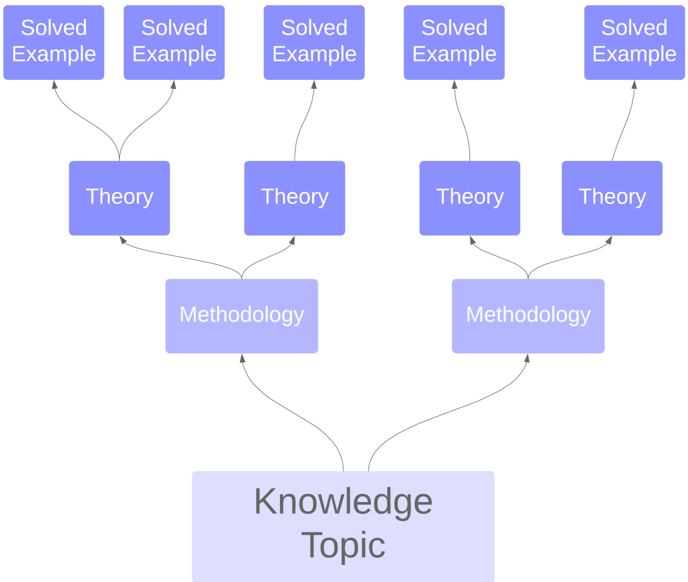

Figure 5.2: The knowledge topic components visualised as an abstract model

The EduKnow framework offers visualisations at two levels. At first, it is the semantic representation of the connections between the knowledge topics with the three different types of links that have been introduced, and also, for a single knowledge topic, we create a representation of the knowledge topic components. The latter consists of the title, one or multiple methodologies with the theory and solved example parts. This illustration points out the different ways different curricula can introduce the same knowledge topic. It is very important to observe in the representation of the techniques that are used to address assessments as well as the types of assessments that the student is asked to deliver.

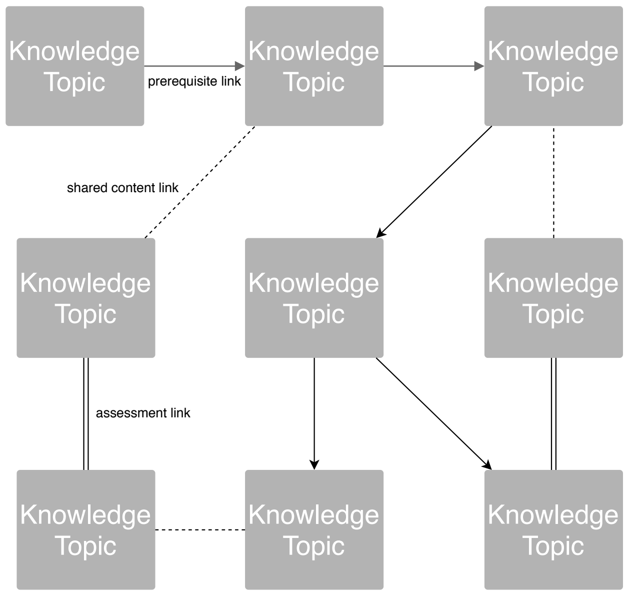

Figure 5.3: The main semantic representation of our database. The prerequisite link is the directed arrow, the shared content link is the dashed line, and the assessment link the double line connector.

In Figure 5.2 we can see an abstract representation of the semantic graph of the components of a single knowledge topic. We have added two different methodologies, however, there can be from one to as many as needed. Also, there can be more solved examples present as part of a methodology. This is an extra layer of visualisation on our representation. To the best of our knowledge, exists no knowledge representation that is suggesting the representation of the knowledge components for each knowledge topic in their research model combined with the visualisations of the knowledge components. The EduKnow framework implementation connects the knowledge topics with via the methodologies structural links as a whole. The methodologies are also connected with the theory and solved example via structural links. The idea of structuring the knowledge components of a topic is a novel contribution to the knowledge structures and representations. The visualisation of them provides the learners with all the different tools they can use to succeed in the knowledge topic they are studying.

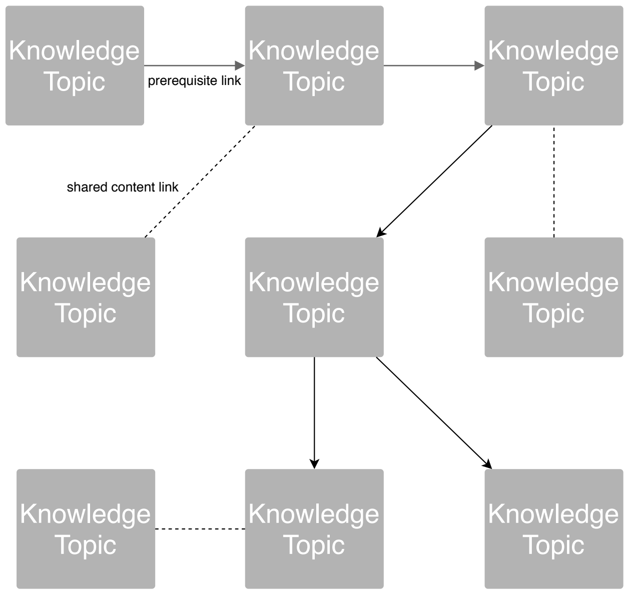

Figure 5.4: The semantic representation in our database of the prerequisite links (directed arrows) and the shared content links (dashed lines).

One of the benefits of the EduKnow framework compared to existing techniques is that it allows us to have multiple visualisations, besides the representation of the entire database. Depending on the links we want to have in our representation, there are different visualisations created. There have been introduced three types of linking in the EduKnow framework. Each one of the links creates a separate knowledge representation. In order to help the reader better understand the possible representations, we have created an example that shows possible links in our database.

The EduKnow framework allows us to have six additional representations, besides the main which is the general representation of the entire database. The additional six different representations are based on the prerequisite, shared content, and assessment link. The prerequisite links are the directed arrows, the shared content links are the dashed lines, and the assessment links are the double line connectors. The next figures show the additional representations that are created based on the main representation of the entire database.

We can choose to view only the prerequisite relations in our database, by querying only the prerequisite links as shown in Figure 5.7. With the same

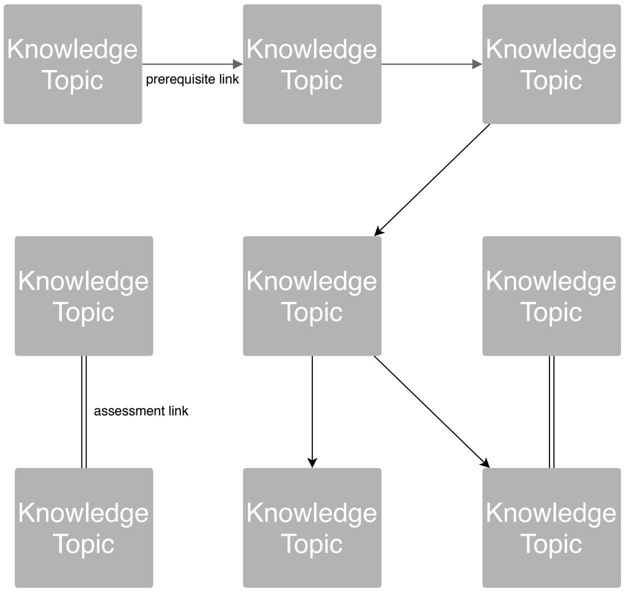

Figure 5.5: The semantic representation in our database of the prerequisite links (directed arrows) and the assessment links (double line connector).

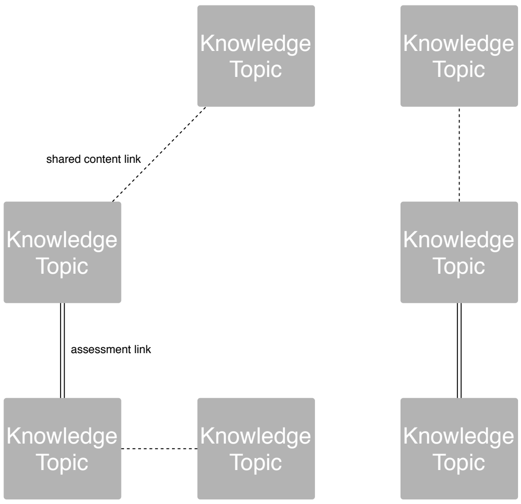

Figure 5.6: The semantic representation in our database of the shared content links (dashed lines) and the assessment links (double line connector).

procedure, we can have a specific view only for the shared content relations between knowledge topics, Figure 5.8. By querying only the assessment links we create a graph visualisation which contains only the knowledge topics that are linked together via complex assessments as shown in Figure 5.9.

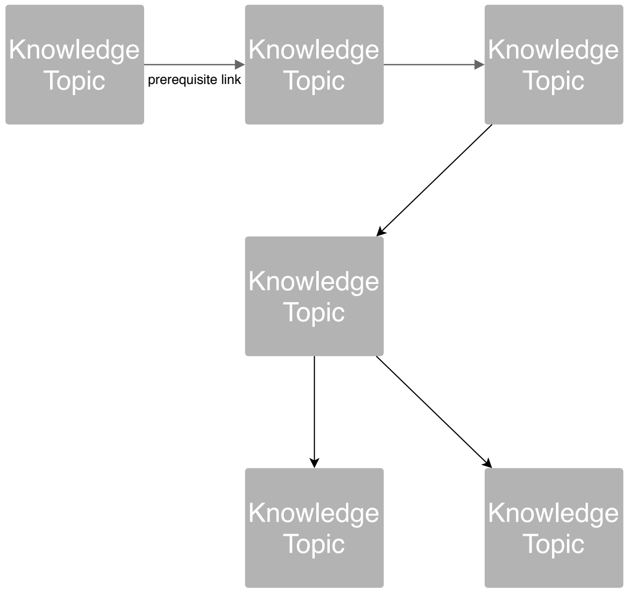

Figure 5.7: The semantic representation in our database of the prerequisite links only.

Moreover, we can select to have two links in our representation. The combinations that can be created are prerequisite links and shared content links co illustrated, in Figure 5.4. The same way the prerequisite links with the assessment links are shown in Figure 5.5, and the shared content links with the assessment links as presented in Figure 5.6.

The added visualisations provide more detailed information visualisation that can assist the learners to have a better understanding of the different steps they have to excel in order to achieve their goal. Also, it is becoming possible by tracking the prerequisite relations and links to spot the knowledge gaps of students, whose performance does not match their effort. Many times, especially in well hierarchically structured educational domains as mathematics, a student cannot perceive the intended meaning due to their lacks of knowledge on simpler topics that were prerequisites to the one the student is facing at the moment. Furthermore, the visualisation with the assessment links can reveal hidden correlations between knowledge topics which seem theoretically unrelated. The assessment link visualisation can also display the knowledge topics that are more connected with other topics in terms of assessments. This can be useful to know the most probable knowledge that will be needed in a set of knowledge topics, that can be used for example in the final exams.

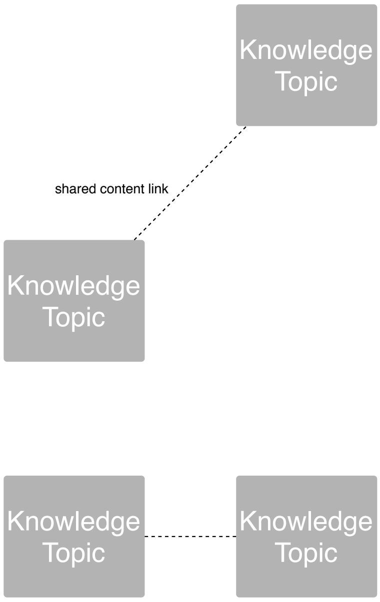

Figure 5.8: The semantic representation in our database of the shared content links only.

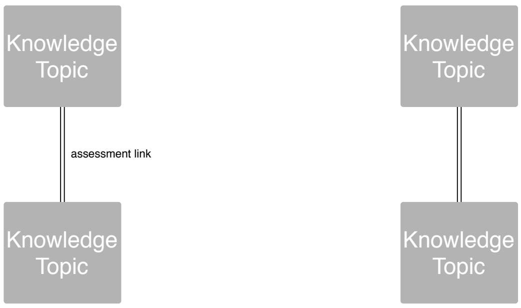

Figure 5.9: The semantic representation in our database of the assessment links only.

The EduKnow framework offers a general knowledge representation with all the links and the knowledge nodes. The similarities to existing techniques rely on the use of the prerequisite link and on the representation of the nodes as knowledge content. However, there are plenty of differences in the EduKnow framework implementation compared to the other techniques. At first, the prerequisite link that is used in the EduKnow framework is more strict defined compared to the probabilistic model that related solutions are using. The nodes of the EduKnow framework are the learning outcomes of each knowledge topic; they are neither too general nor too specific as discussed earlier. Moreover, the EduKnow framework implementation offers representations with all the possible combination of the three links and created six extra subgraphs from the main representation of the entire database. Finally, it allows a detailed representation of the knowledge topic components. This creates a multidimensional graph in which all the ingredients of educational knowledge are combined.

6 Use Case: Algebra

In this chapter, we are examining a use case of the EduKnow framework based on the RSL implementation. The use case is also going to play the role of a technical evaluation of the EduKnow framework, as it will test whether the framework has all the necessary components to create a complete representation of the knowledge.

6.1 Motivation

For the deployment of our use case, we choose the educational domain of mathematics. The subject of mathematics is being taught in all schools and curricula around the world. It is one of the universal courses that exist in curricula from the first grade of primary school until the end of high school. This has created a big amount of educational material and data related to didactic techniques of mathematics, the theory of mathematics, and assessments, which can be found in printed form and online. Our aim is to extract the inner structure of the mathematical topics, organise them properly and link them by using the EduKnow framework. More specifically our focus is on the knowledge domain of algebra topics, that are presented to pupils during the primary and secondary school. Algebra has the majority of mathematical topics that are presented in schools at early stages. The reason behind this is that counting and basic algebraic operations are the foundations of more complex theories like analysis and statistics.

Moreover, mathematics and mostly algebra are one of the most well structured knowledge domains. It starts from the foundation of the arithmetic and algebra and each new topic is built on top of the previous one. Therefore structuring the knowledge domain of algebra is going to disclose the full potential of the EduKnow framework.

6.2 Knowledge Representation

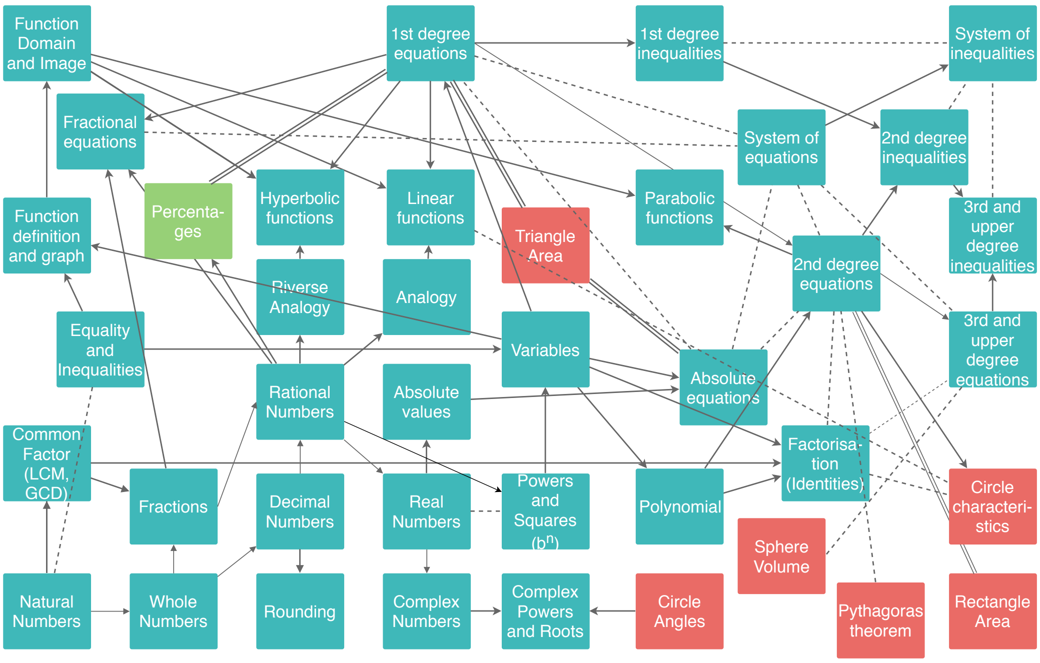

Figure 6.1: Part of the algebra that has been used as the use case for the EduKnow framework.

For the construction of the knowledge representation, we first need to fulfil the requirements of the EduKnow framework. Our first step is to identify the knowledge topics that are going to be present in our graph. The knowledge topics should represent the learning outcomes of the topic. For example, the knowledge topic *1st degree equations* represents the learning outcome of a student being able to perform 1st degree equations. At the first stage, we ignore how a student will be able to learn and expertise a topic. The first goal is to clarify all the knowledge topics and make sure there is no overlap between nodes and that all the learning outcomes are present.

Once we have identified all the knowledge topics, we can proceed to the linking part. At first, we begin by distinguishing the prerequisite relationships. By spotting the direct prerequisites as defined in Chapter 4, we add the first links to our representation, the prerequisite links. We continue with the indication of the shared content relationships and links. The shared content links will possibly not be as many as the prerequisite links. However, they are also very critical in the construction of an accurate knowledge representation and therefore it is important to identify them properly.

Before we establish the last type of links, we should populate the knowledge topics with assessments. The complex assessments and their analysis will set the foundations from where the assessments links are going to be created. By adding the assessment links to our graph the knowledge representation is complete.

However, there are still the knowledge topic components that need to be filled in. We complete the parts of each knowledge topic with one or more methodologies, each containing a title, theory and solved example parts.

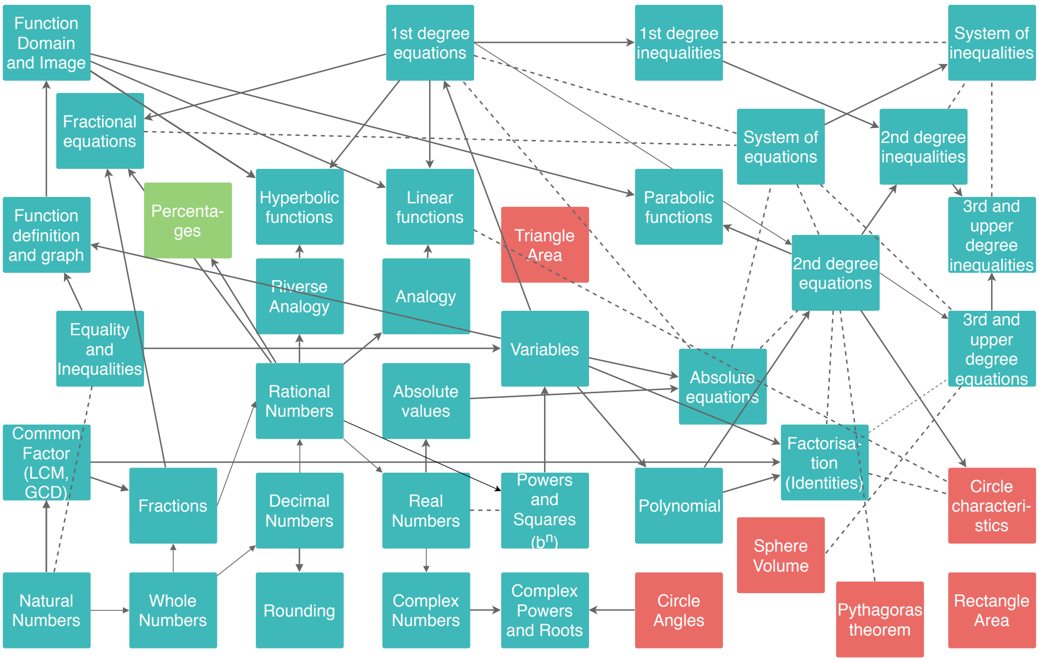

Figure 6.2: Visualisation using only the prerequisite and shared content links.

Figure 6.1 shows the representation of our database for parts of the algebra with extension to other domains. In the figure the knowledge topics of algebra are shown in blue, the red nodes are knowledge topics from the knowledge domain of geometry and the one green node is from statistics. The first conclusions from our representation compared to the related work techniques is that the knowledge topics are represented with more detail on the content. Note that the visualisations have been created with illustration software based on the principles of the EduKnow framework implementation, and the final visualisation will still have to be implemented.

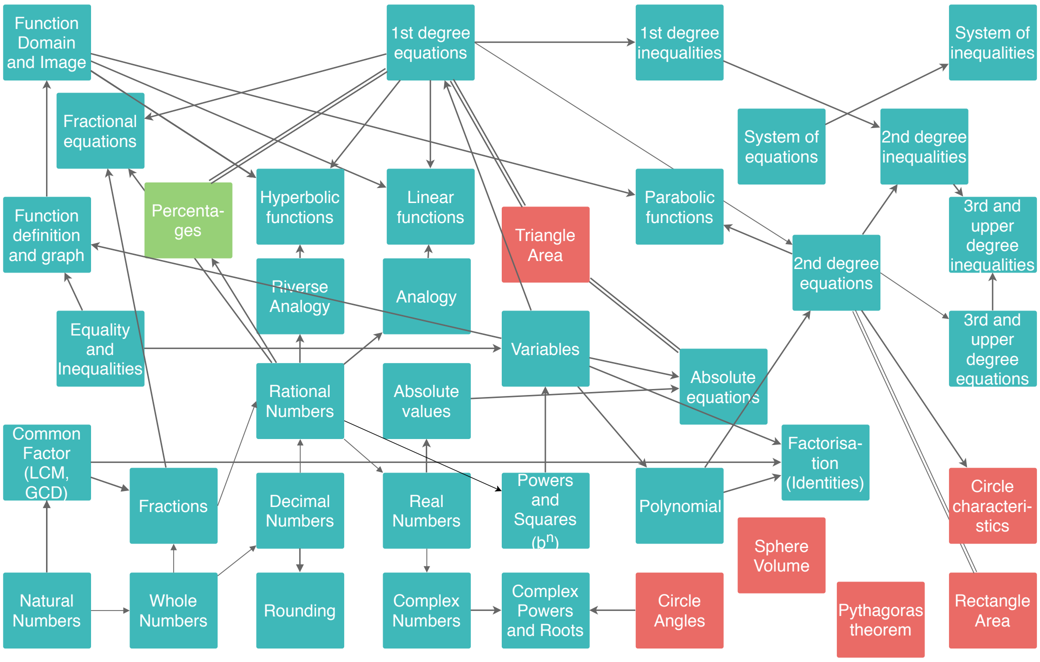

Figure 6.3: Visualisation using only the prerequisite and assessment links.

The visualisation of all the links is the main illustration that the EduKnow framework is offering. It is similar to the existing techniques, as it contains prerequisite links and knowledge nodes, although there are fundamental differences, as there are two more links and the knowledge nodes describing a more specific topic.

As we have shown in Chapter 5, the implementation of the EduKnow framework offers six extra visualisations besides the main visualisation of the entire database. In Figure 6.2 we see the result of only the prerequisite and shared content link part. These two links are the most common in the EduKnow framework and therefore there are only a few differences compared to the main representation.

Figure 6.3 shows the representation of the prerequisite links in combination with the assessment links. In Figure 6.4 we see the representation of the shared content links with the assessment links. This is obviously a

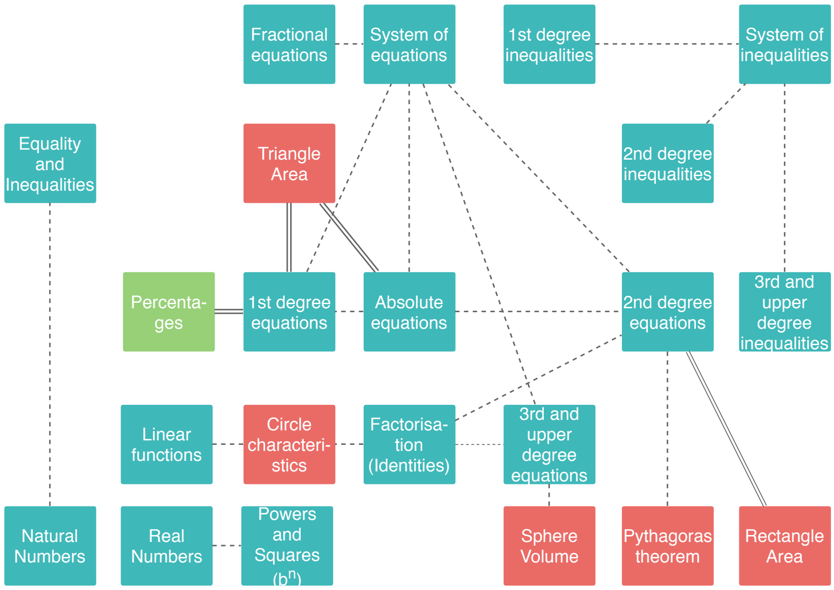

Figure 6.4: Visualisation using only the shared content and assessment links.

small sub-graph of the full visualisation. There we can also realise that the prerequisite links are the most common connections we find in the primary level of knowledge in algebra. This phenomenon might not be visible in other educational domains such as biology.

Figure 6.5, Figure 6.5, and Figure 6.7 show the single links representations. In Figure 6.5 the prerequisite link connects most of the nodes, therefore only a few nodes from the original graph are not present. The shared content comes second in terms of appearance frequency and the assessment link last, although the last might be more interesting to be further investigated. Shared content links alone create a visualisation of a subpart of the main graph as shown in Figure 6.6, and even smaller is the representation of only the assessment links as highlighted in Figure 6.7.

Another useful visualisation that the EduKnow framework is offering is the illustration of the knowledge components of a knowledge topic. The knowledge topics are represented with their title entry in the connections between other topics. However, they also consist of the methodology titles, the theory, and the solved examples. The visualisation of the insights of a knowledge topic is a guide towards understanding the knowledge topic and approaching the possible problems this topic is introducing. The different methodologies can often address different types of problems or offer an alternative method to solve the same assessment. The different perspectives that

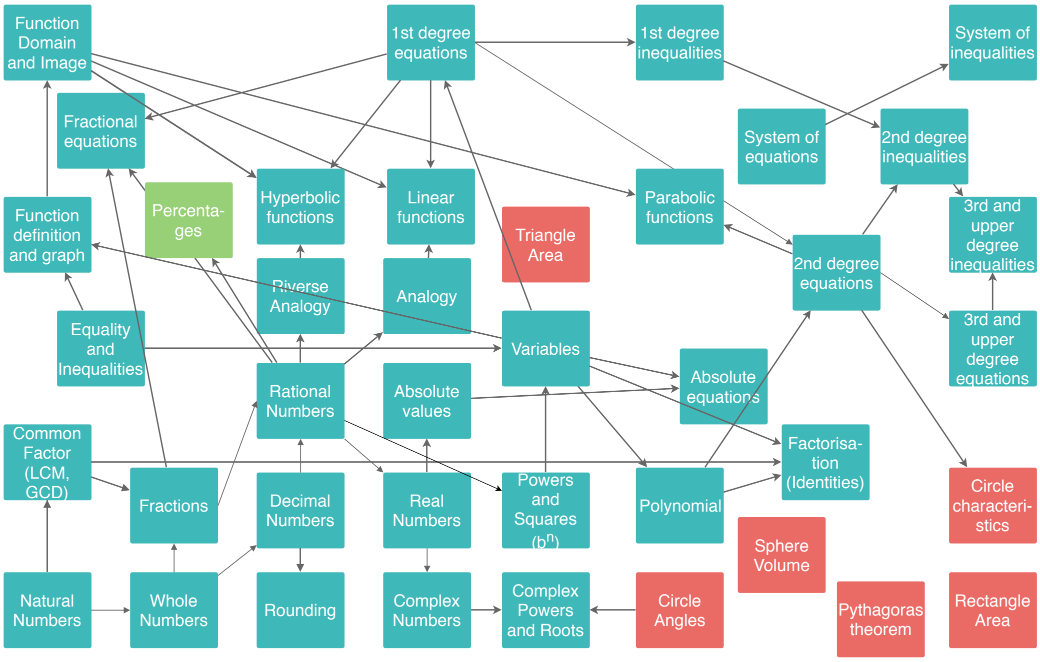

Figure 6.5: Visualisation using only the prerequisite links.

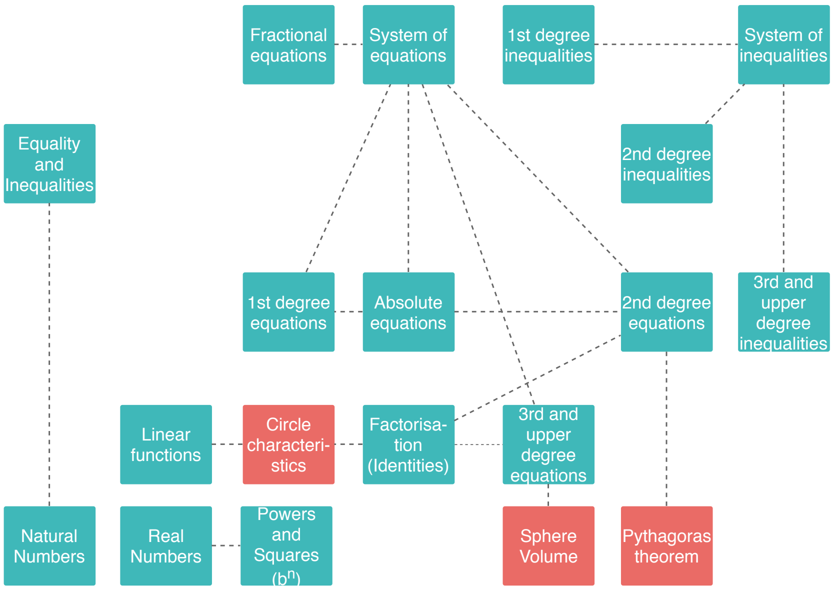

Figure 6.6: Visualisation using only the shared content links.

are presented in the methodologies provide auxiliary methods that broaden the knowledge of learners. An example of the visualisation of the knowledge

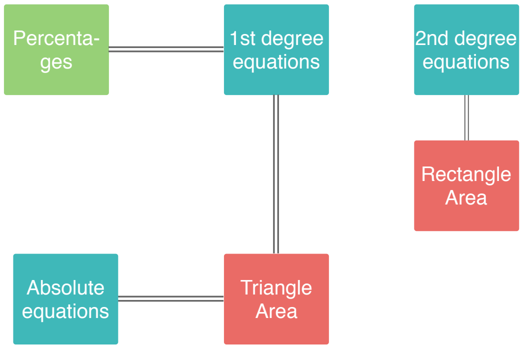

Figure 6.7: Visualisation using only the assessment links.

components of the knowledge topic of *2nd degree equations* is illustrated in Figure 6.8.

6.3 Applications

The EduKnow framework structures knowledge topics in order to create knowledge representations. At the first view, it might not be possible to realise the full potential and necessity of our approach. EduKnow's superiority compared to existing models lies in its structure and visualisations. Let us focus on a specific knowledge topic to continue the discussion with a concrete example. Let us assume that a student who is currently studying the topic of *linear functions* is underperforming. The student is attending classes and studying individually, however, the results do not reflect the effort and time the student is investing. Our aim is to help the student understand the current topic and moreover, realise and detect any knowledge gaps that may occur. The first is accomplished by representing the knowledge components of the knowledge topic, as shown in Figure 6.8.

The EduKnow framework can help in the detection of knowledge gaps by querying the links for a specific knowledge topic. We can create a subgraph of a part of the database, which has as a final node the query object. Figure 6.9 could be a result of this process. This visualisation can assist the student and the teacher to find all the possible reasons behind the poor understanding of the student. It might be that there is a knowledge gap in the direct prerequisite topic, although it could be that the knowledge that is missing comes from the shared content material, or even that it exists at the beginning of the graph.

Therefore, the EduKnow framework can assist in the detection of knowledge gaps, which is otherwise a long procedure. The learners and teachers can

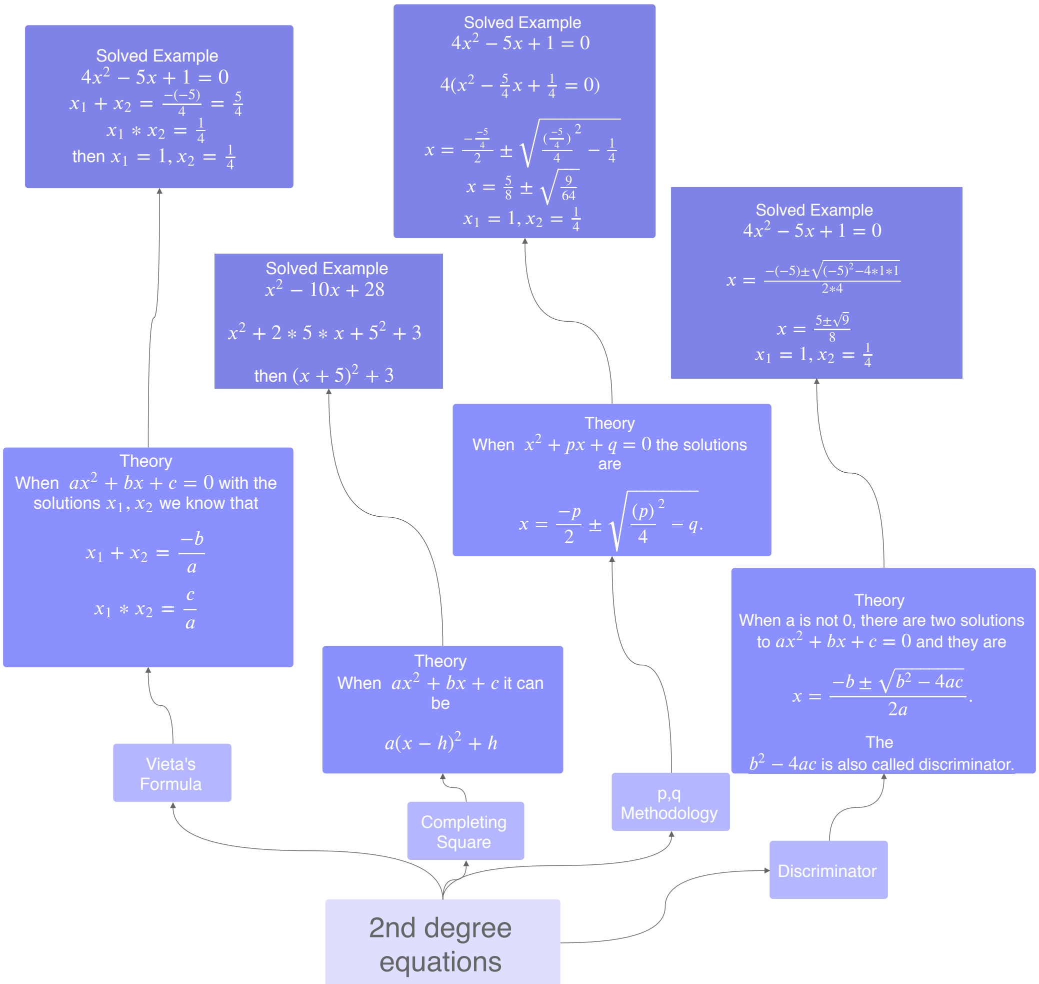

Figure 6.8: Visualisation of the knowledge components of the knowledge topic of 2nd degree equations.

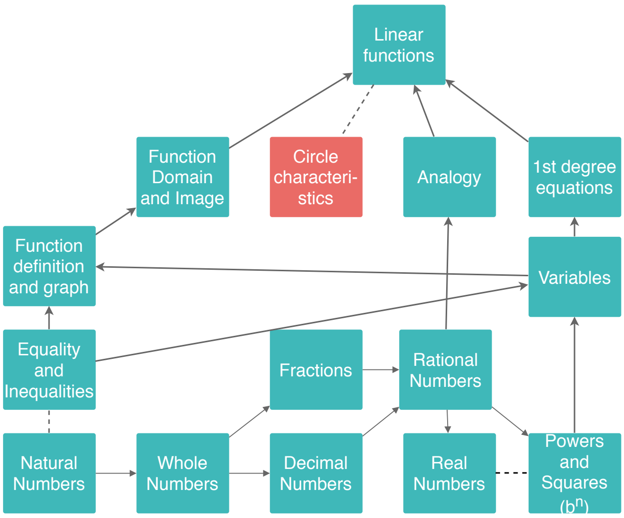

Figure 6.9: Visualisation of parts of the database for the *linear functions* knowledge topic.

save valuable time and resources and achieve more with the representations the EduKnow framework has to offer.

7 Future Work

The EduKnow framework has proven its applicability, scalability, and usability in the use case. Therefore it can form the basis for future applications, such as the knowledge representation in other educational and knowledge domains. The deployment of more knowledge representations with the EduKnow framework would lead to uniform graphs that could be interconnected and create a big complete representation of all the educational material in all domains. This will be an evolutionary attempt to classify, organise and connect the knowledge that is taught in educational institutions, that could help learners, teachers and educational experts to have a broad view and understanding of the taught knowledge. The EduKnow framework could provide knowledge visualisations that will assist self-taught learners to achieve faster results, and create a personal path of knowledge of each individual. This path can help educational institutions and companies to realise the expertise and knowledge gaps of an employee or a student and make their training more precise and less expensive.

The use case in algebra can be extended with more nodes covering the more advanced material. Also, the current representation of algebra can be enriched by implementing more methodologies and assessments for each node. This could possibly lead to more assessment link connections. Furthermore, in cooperation with educational researchers, the EduKnow framework could host educational and psychological parameters for its representation. Hence, based on the type of the learner it could characterise the material. This could be for simple separations such as a visual learner to students with special needs. This way it could implement specific material that fits better to each learner category, such as blind, deaf or autistic learners. Furthermore, the EduKnow framework could further be evaluated by students and tested as an assistant to the learning process. This evaluation could examine its benefits and provide valuable feedback that could make the visualisations and design more user-centred [1].

Moreover, in future work, the implementation of the content of knowledge nodes could be automated by natural language processing that could process educational material, recognise knowledge topics and add new methodologies, theories and solved examples where they are missing. Some automation could also be used in the recognition of the multiple domains of complex assessments. An AI algorithm could suggest assessment links for new assessments by analysing the existing links in the database. One step further would be the automatic creation of solved examples and assessments, based on the current ones. The EduKnow framework has unlimited scalability and extensibility. On a large scale, it could be used as a pool for AI algorithms to process and analyse the content and connections in order to make conclusions and create new content and connections.

Another addition would be the development of the learning paths based on the current knowledge representation. The learning paths are a sequence of knowledge topics in the order they are presented in a curriculum. More precisely we can observe that different learning paths are examining different knowledge components of the same knowledge topic by, for example, using different methodologies to solve the same problem. The EduKnow framework is designed in such a way that it allows the coexistence of different curricula (knowledge topics) and highlights the fundamental differences at the same time (methodologies). Learning paths can also reveal the characteristics of different educational policies. The similarities and differences of these policies could highlight the reasons for poor and good performance, and start new research on an ideal curriculum or learning path for general knowledge and per study case.

8 Conclusion

The work undertaken this thesis highlighted some current limitations of knowledge representation techniques, such as the representation of the knowledge components of each knowledge topic. Therefore, we contribute to the research towards solving this problem by developing a comprehensive framework and educational data visualisation. Also, our model suggests a method how the educational knowledge can be structured and organised based on the learning outcomes. One might begin to realise improvements in understanding the knowledge space of educational data and its applications. Therefore, we provide some mathematical foundations for our framework for the relations between knowledge topics and assessments. By representing relationships and links through mathematical notations one can get a better insight over the knowledge space through different visualisations.

The presented EduKnow framework represents the start of new ground in knowledge representation for educational material. It may serve as a useful starting point for the representation of knowledge domains that can assist classrooms during the teaching process as well as independent individuals. The presented solution constitutes of several requirements and different types of links, which are combined to realise the novel idea of knowledge representation that can identify knowledge prerequisites and backtrack topics. The latter can be a useful tool in detecting the knowledge gaps of students, and help the learners achieve better results and have a deep understanding of the knowledge.

The novelty of the EduKnow framework is the fact that it represents and connects all knowledge components in a multidimensional graph representation. It links the knowledge topics with three types of links, depending on the relationship between the topics, and also interconnects the knowledge components of each topic. The presented model offers learners a rich representation of the knowledge they are studying with all the connections between different topics and shows possible ways of mastering a topic. Moreover, our solution also covers assessments, making it a fully capable teacher assisting tool. Also, the components of each knowledge topic can be the start of research towards the identification of different learning paths and the methodologies they introduce based on their curricula.

The main contributions of the EduKnow framework are:

- The exploration and in-depth analysis of the knowledge representation techniques and their comparison, which revealed the advantages and weaknesses of current models.

- An examination of the knowledge representation requirements and the mathematical foundations of the EduKnow framework, which introduces a concrete method for representing knowledge topics as the nodes in a graph and three types of links that cover the possible relations between topics.

- The structure of the knowledge components for each knowledge topic. Each knowledge topic consists of the different methodologies based on which it can be mastered. The EduKnow framework offers a rich representation for each knowledge topic, which aids the learners to understand the different ways they can master a topic.

- The novel framework modelled via the RSL hypermedia metamodel for creating the EduKnow framework graph and subgraphs. The RSL framework offered an enriched model for the knowledge representation graphs of the EduKnow framework with many applications.

- A use case for algebra with the usage of the EduKnow framework which illustrates the potential interactions with learners, and also served as technical evaluation of the EduKnow framework.

As shown in this thesis, the EduKnow framework overcomes the shortcomings of existing models in the visualisation, structural definition, and specification of knowledge connections and content. The EduKnow framework enables the representation of the complete knowledge base as well as combinations of the three links with six different representations. Moreover, for each single knowledge topic, there is a visualisation of its knowledge components. Also, the importance of the EduKnow framework is highlighted in the variety of future work applications. Therefore, the EduKnow framework is a promising tool which can contain all the necessary components of the taught knowledge. It has been developed after research of necessary knowledge elements which determined its structure with a solid mathematical foundation. Finally, the EduKnow framework has great potential for future applications supporting students as well as teachers.

A
Appendix

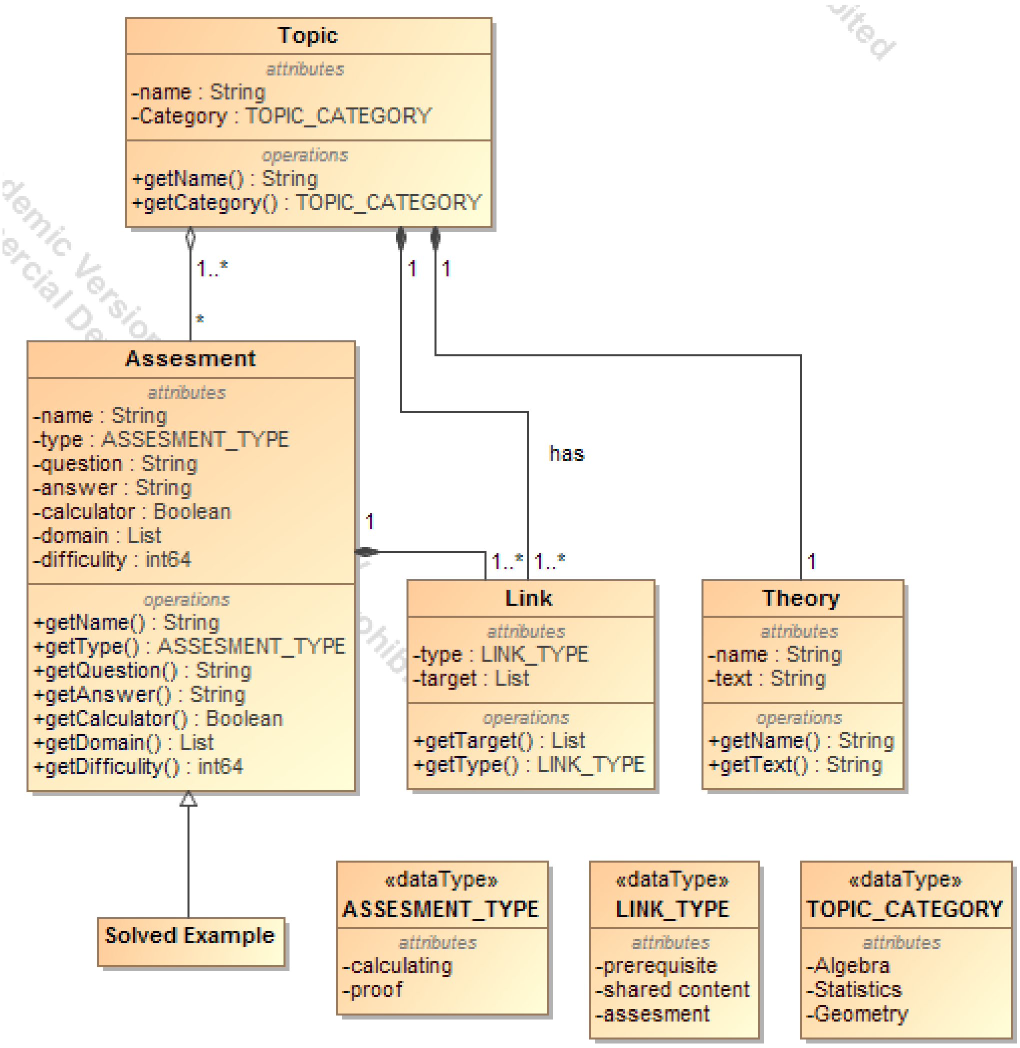

Figure A.1: The class diagram of the EduKnow framework. The solved examples are considered as assessment with its answer presented.

```json
{
    "name": "Knowledge Graph",
    "version": "0.0.1",

    "resources": [
        {
            "name": "KnowledgeTopic",
            "properties": [
                {"name": "title", "type": "string"},
                {"name": "theme", "type": "string"}
            ]
        },
        {
            "name": "Assessment",
            "properties": [
                {"name": "theme", "type": "string"}
            ]
        },
        {
            "name": "Methodology",
            "properties": [
                {
                    "name": "title",
                    "type": "string"
                }
            ]
        },
        {
            "name": "Theory",
            "properties": [
                {"name": "title", "type": "string"}
                ]
            },
        {
            "name": "SolvedExample",
            "properties": [
                {"name": "title", "type": "string"}
                ]
            }
    ],

    "selectors": [
        {
            "name": "TheorySelector",
            "properties": [
                {"name": "title", "type": "string"}
            ],
            "refersTo": "Theory"
        },
        {
            "name": "SolvedExampleSelector",
            "properties": [
                {"name": "title", "type": "string"}
            ],
            "refersTo": "SolvedExample"
        },
        {
            "name": "AssessmentTypeSelector",
            "properties": [
                {"name": "title", "type": "string"}
            ],
            "refersTo": "Assessment"
        }
```

```json
    ],

    "links": [

        {
          "name": "PrerequisteLink",
          "sourceRestrictions": {"type": "whitelist", "list": ["KnowledgeTopic"]},
          "targetRestrictions": {"type": "whitelist", "list": ["KnowledgeTopic"]},
          "properties": [
             {"name": "followCount", "type": "int64"}
             ]
        },
        {
          "name": "SharedContentLink",
          "sourceRestrictions": {"type": "whitelist", "list": ["KnowledgeTopic"]},
          "targetRestrictions": {"type": "whitelist", "list": ["KnowledgeTopic"]},
          "properties": [
             {"name": "followCount", "type": "int64"}
          ]
        },
        {
          "name": "AssessmentLink",
          "sourceRestrictions": {"type": "whitelist", "list": ["Assessment"]},
          "targetRestrictions": {"type": "whitelist", "list": ["KnowledgeTopic"]},
          "properties": [
             {"name": "followCount", "type": "int64"}
          ]
        },

        {
          "name": "HasMethodology",
          "sourceRestrictions": {"type": "whitelist", "list": ["KnowledgeTopic"]},
          "targetRestrictions": {"type": "whitelist", "list": ["Methodology"]},
          "properties": [
             {"name": "followCount", "type": "int64"}
             ]
        },
        {
          "name": "HasAssessments",
          "sourceRestrictions": {"type": "whitelist", "list": ["KnowledgeTopic"]},
          "targetRestrictions": {"type": "whitelist", "list": ["Assessment"]},
          "properties": [
             {"name": "followCount", "type": "int64"}
          ]
        },
        {
          "name": "HasTheory",
          "sourceRestrictions": {"type": "whitelist", "list": ["Methodology"]},
          "targetRestrictions": {"type": "whitelist", "list": ["Theory"]},
          "properties": [
             {"name": "followCount", "type": "int64"}
          ]
        },
        {
          "name": "HasSolvedExample",
          "sourceRestrictions": {"type": "whitelist", "list": ["Methodology"]},
          "targetRestrictions": {"type": "whitelist", "list": ["SolvedExample"]},
          "properties": [
             {"name": "followCount", "type": "int64"}
          ]
        }
    ]
}
```

Figure A.2: The JSON file of the EduKnow framework implementation.

Bibliography

[1] Chadia Abras, Diane Maloney-Krichmar, Jenny Preece, et al. User-centered design. *Bainbridge, W. Encyclopedia of Human-Computer Interaction. Thousand Oaks: Sage Publications*, 37(4):445–456, 2004.

[2] Karoline Afamasaga-Fuatai. *Concept Mapping in Mathematics*. Springer, 2009.

[3] Rakesh Agrawal, Tomasz Imieliński, and Arun Swami. Mining Association Rules between Sets of Items in Large Databases. In *Acm sigmod record*, volume 22, pages 207–216. ACM, 1993.

[4] Rakesh Agrawal, Heikki Mannila, Ramakrishnan Srikant, Hannu Toivonen, A Inkeri Verkamo, et al. Fast Discovery of Association Rules. *Advances in knowledge discovery and data mining*, 12(1):307–328, 1996.

[5] Rakesh Agrawal, Ramakrishnan Srikant, et al. Fast Algorithms for Mining Association Rules. 1215:487–499, 1994.

[6] J. Baath. Assignements in Distance Education—an Overview. *Epistolodidaktika*, 1:13–20, 1994.

[7] Marcia Barros and Francisco M Couto. Knowledge representation and management: A linked data perspective. *Yearbook of medical informatics*, 25(01):178–183, 2016.

[8] Ronald J. Brachman and Hector J. Levesque. *Readings in Knowledge Representation*. Morgan Kaufmann Publishers Inc.

[9] Astrid Brinkmann. Knowledge Maps—Tools for Building Structure in Mathematics. *International Journal for Mathematics Teaching and Learning*, pages 1473–0111, 2005.

[10] Tony Buzan. *Mind Mapping*. Pearson Education, 2006.

[11] Tony Buzan and Barry Buzan. *The Mind Map Book*. Pearson Education, 2006.

[12] Lillian N Cassel, Gordon Davies, William Fone, Anneke Hacquebard, John Impagliazzo, Richard LeBlanc, Joyce Currie Little, Andrew McGettrick, and Michela Pedrona. The Computing Ontology: Application in Education. *ACM SIGCSE Bulletin*, 39(4):171–183, 2007.

[13] Devendra Singh Chaplot, Yiming Yang, Jaime Carbonell, and Kenneth R Koedinger. Data-driven automated induction of prerequisite structure graphs. *International Educational Data Mining Society*, 2016.

[14] Penghe Chen, Yu Lu, Vincent W Zheng, Xiyang Chen, and Xiaoqing Li. An Automatic Knowledge Graph Construction System for K-12 Education. In *Proceedings of the Fifth Annual ACM Conference on Learning at Scale*, page 40, London, UK, 2018. ACM.

[15] Penghe Chen, Yu Lu, Vincent W Zheng, Xiyang Chen, and Boda Yang. Knowedu: A system to construct knowledge graph for education. *IEEE Access*, 6:31553–31563, 2018.

[16] Sunita Chugh. Dropout in Secondary Education, A Study of Children Living in Slums of Delhi, 2011.

[17] Randall Davis, Howard Shrobe, and Peter Szolovits. What Is a Knowledge Representation? *AI magazine*, 14(1):17–17, 1993.

[18] Jean-Paul Doignon and Jean-Claude Falmagne. Spaces for the Assessment of Knowledge. *International Journal of Man-Machine Studies*, 23(2):175–196, August 1985.

[19] Jean-Paul Doignon and Jean-Claude Falmagne. *Knowledge Spaces*. Springer-Verlag Berlin Heidelberg GmbH, 1999.

[20] Martin J. Eppler. A Comparison Between Concept Maps, Mind Maps, Conceptual Diagrams, and Visual Metaphors as Complementary Tools for Knowledge Construction and Sharing. *Information Visualization*, 5(3):202–210, 2006.

[21] Jean-Claude Falmagne, Eric Cosyn, Jean-Paul Doignon, and Nicolas Thiéry. The Assessment of Knowledge, in Theory and in Practice. In *Formal Concept Analysis*, pages 61–79. Springer, 2006.

[22] Kathleen M. Fisher, James H. Wandersee, and David E. Moody. *Mapping Biology Knowledge*, volume 11. Springer Science & Business Media, 2001.

[23] Suresh Garg, V Panda, and S Panda. A Preliminary Study of Student Workload for IGNOU Physics Elective Courses. *Mathematical Methods in Physics*, 2(2), 1992.

[24] Thomas R Gruber. A Translation Approach to Portable Ontology Specifications. *Knowledge acquisition*, 5(2):199–220, 1993.

[25] M. Gruninger and J. Lee. Ontology Applications and Design. *Commun. ACM*, 45(2):39–41, 2002.

[26] Sigrun Gudmundsdottir and Lee Shulman. Pedagogical Content Knowledge in Social Studies. *Scandinavian Journal of Educationl Research*, 31(2):59–70, 1987.

[27] Amarnath Gupta, Bertram Ludäscher, and Reagan W Moore. Ontology Services for Curriculum Development in NSDL. In *Proceedings of the 2nd ACM/IEEE-CS joint conference on Digital libraries*, pages 219–220. ACM, 2002.

[28] Asaad Sabah Hadi and Enass Mahdi Abdulshaheed. Twitter information representation using resource description framework. *International Journal of Engineering & Technology*, 8(1.5):497–502, 2019.

[29] Matti Harjula et al. Mathematics Exercise System with Automatic Assessment. 2008.

[30] Syed Tahir Hijazi and S.M.M. Raza Naqvi. Factors Affecting Student's Performance, a Case of Private Colleges. *Journal of sociology*, 3(1):44–45, January 2006.

[31] Phillip B. Horton, Andrew A. McConney, Michael Gallo, Amanda L. Woods, Gary J. Senn, and Denis Hamelin. An Investigation of the Effectiveness of Concept Mapping as an Instructional Tool. *Science Education*, 77(1):95–111, January 1993.

[32] Dimitrios Karaistolis. *Data Analysis, Statistics without Models.* Anikoula, 2008.

[33] Adrienne Kitts and Stephen Sherry. The single nucleotide polymorphism database (dbsnp) of nucleotide sequence variation. *The NCBI*

Handbook. McEntyre J, Ostell J, eds. Bethesda, MD: US National Center for Biotechnology Information, 2002.

[34] Matthew Koehler and Punya Mishra. What Is Technological Pedagogical Content Knowledge? *Contemporary issues in technology and teacher education*, 9(1):60–70, 2009.

[35] Ora Lassila, Ralph R Swick, et al. Resource description framework (rdf) model and syntax specification. 1998.

[36] Nguyen-Thinh Le and Niels Pinkwart. Towards a Classification for Programming Exercises. In *Proceedings of the 2nd Workshop on AI-supported Education for Computer Science*, Honolulu, USA, 2014.

[37] Nicola Leone, Gerald Pfeifer, Wolfgang Faber, Thomas Eiter, Georg Gottlob, Simona Perri, and Francesco Scarcello. The DLV System for Knowledge Representation and Reasoning. *ACM Transactions on Computational Logic (TOCL)*, 7(3):499–562, 2006.

[38] Shu-hsien Liao. Knowledge Management Technologies and Applications—Literature Review from 1995 to 2002. *Expert systems with applications*, 25(2):155–164, August 2003.

[39] Robert L. Linn, Eva L. Baker, and Stephen B. Dunbar. Complex, Performance-Based Assessment: Expectations and Validation Criteria. *Educational researcher*, 20(8):15–21, November 1991.

[40] Hanxiao Liu, Wanli Ma, Yiming Yang, and Jaime Carbonell. Learning Concept Graphs from Online Educational Data. *Journal of Artificial Intelligence Research*, 55:1059–1090, 2016.

[41] P.J. Longo. *Visual Thinking Networking promotes Long-Term Meaningful Learning and Achievement for Ninth Grade Earth Science Students.* PhD thesis, Teachers College, Columbia University, New York, N.Y, 2001.

[42] G. Lusser Rico and Jeremy P Tarcher. Writing the Natural Way. *Using Right-Brain Techniques to Release Your Expressive Powers*, 1983.

[43] Zorica Marinković. Concept Maps in Math Teaching. 2014.

[44] M. L. Mugnier and M. Chein. Conceptual Graphs: Fundamental Notions. *Revue d'intelligence artificielle*, 6(4):365–406, 1992.

[45] John Mylopolous and Michael L. Brodie. *Readings in Artificial Intelligence and Databases*. Morgan Kaufmann, 2014.

[46] Joseph D. Novak. Concept Mapping: A Useful Tool for Science Education. *Journal of research in science teaching*, 27(10):937–949, 1990.

[47] Joseph D. Novak. Clarify with Concept Maps. *The Science Teacher*, 58(7):44, October 1991.

[48] Kyle A. O'Connell, Elijah Wostl, Matt T. Crosslin, Lisa Berry, and James P. Grover. Student Ability Best Predicts Final Grade in a College Algebra Course. *Journal of Learning Analytics*, 5(3):167–181, 2018.

[49] Ken Peffers, Tuure Tuunanen, Marcus A. Rothenberger, and Samir Chatterjee. A Design Science Research Methodology for Information Systems Research. *Journal of management information systems*, 24(3):45–77, 2007.

[50] Lisa Petrides and Thad R. Nodine. Knowledge Management in Education: Defining the Landscape. 2003.

[51] Shelley Powers. *Practical RDF: Solving Problems with the Resource Description Framework*. O'Reilly Media, Inc., 2003.

[52] Alan Pritchard. *Ways of Learning: Learning Theories for the Classroom*. Routledge, Taylor & Francis Group, 2017.

[53] Reinout Roels and Beat Signer. MindXpres: An Extensible Content-Driven Cross-Media Presentation Platform. In *International Conference on Web Information Systems Engineering*. Springer.

[54] Beat Signer and Moira C. Norrie. *As We May Link: A General Metamodel for Hypermedia Systems*. Springer, 2007.

[55] Scott J South, Dana L Haynie, and Sunita Bose. Student Mobility and School Dropout. *Social Science Research*, 36(1):68–94, 2007.

[56] John F. Sowa. Semantic Networks. *Encyclopedia of Artificial Intelligence*, 1987.

[57] Ramakrishnan Srikant and Rakesh Agrawal. Mining Generalised Association Rules, 1995.

[58] Amrit Tiwana. *The Knowledge Management Toolkit: Practical Techniques for Building a Knowledge Management System*. Prentice Hall PTR, 2000.

[59] E. R. Tufte, N. H. Goeler, and R. Benson. *Envisioning Information*, volume 126. Graphics press Cheshire, CT, 1990.

[60] Yiming Yang, Hanxiao Liu, Jaime Carbonell, and Wanli Ma. Concept Graph Learning from Educational Data. pages 159–168, February 2015.