IANUA
lingue, culture, educazione

La collana, di carattere interdisciplinare, accoglie contributi inerenti agli ambiti della linguistica educativa, della comunicazione interculturale, della formazione e delle tecnologie per l'apprendimento.

IANUA
lingue, culture, educazione

Collana diretta da
Daria Coppola

Segreteria di redazione
Cristiana Cervini
Dipartimento di Interpretazione e Traduzione, Università di Bologna
Francesca Gallina
Dipartimento di Filologia, Letteratura e Linguistica, Università di Pisa

Comitato scientifico
Camilla Bardel
Department of Language Education, University of Stockholm
Sandro Caruana
Faculty of Education, University of Malta
Daria Coppola
Dipartimento di Scienze Umane e Sociali, Università per Stranieri di Perugia
Matteo Santipolo
Dipartimento di Studi Linguistici e Letterari, Università di Padova
Massimo Vedovelli
Dipartimento di Studi Umanistici, Università per Stranieri di Siena

Ogni contributo proposto per la pubblicazione viene prima sottoposto al vaglio del direttore e del comitato scientifico e successivamente alla procedura "doppio *peer reviewing* anonimo", affidata a studiosi ed esperti esterni alla direzione e al comitato scientifico.

Il colore che caratterizza la copertina di ogni volume viene scelto dall'autore

Chiara Alzetta

Unlocking Knowledge in the Digital Age

A Guide to Modelling Propaedeutic Relations in Educational Texts

visualizza la scheda del libro sul sito www.edizioniets.com

Edizioni ETS

www.edizioniets.com

Volume pubblicato con il contributo
del Dipartimento di Lingue e Culture Moderne
dell'Università degli Studi di Genova.

© Copyright 2024
Edizioni ETS
Palazzo Roncioni - Lungarno Mediceo, 16, I-56127 Pisa
info@edizioniets.com
www.edizioniets.com

Distribuzione
Messaggerie Libri SPA
Sede legale: via G. Verdi 8 - 20090 Assago (MI)

Promozione
PDE PROMOZIONE SRL
via Zago 2/2 - 40128 Bologna

ISBN 978-884676923-7
ISSN 2785-4078

Contents

Acknowledgments 9

Introduction 11

Chapter 1
The Building Blocks of Knowledge Modelling in Education 19
1.1. Concepts: The Cornerstone of Understanding and Knowledge Sharing 19
1.1.1. How to Acquire Domain Concepts from Educational Resources 23
1.2. Prior Knowledge Matters: Prerequisite Relations 24
1.2.1. Formal Properties of Prerequisite Relations in Graph-based Knowledge Representations 27
1.2.2. Approaches to Concepts Organisation 30
1.2.3. How to Acquire Prerequisite Relations from Educational Resources 33
1.3. Knowledge Modelling for Educational Enhancement 37

Chapter 2
Enriching Educational Texts: The Annotation of Corpora 39
2.1. How to Build Manually Annotated Corpora 40
2.1.1. Defining the Annotation Task 41
2.1.2. Testing, Evaluating, and Revising 43
2.2. Choosing the Right Tool: Annotation Interfaces 43
2.3. Quality Control in Annotation 45
2.4. Manual Annotation Initiatives in Prerequisite Research 48

Chapter 3
A Protocol for Annotating Prerequisite Relations in Educational Texts 51
3.1. PREAP User Guide 53
3.1.1. Setting up the Annotation Project 55
3.1.2. Enriching the Educational Text with Annotations 58
3.1.3. Quality Control and Gold Dataset Creation 67
3.2. Educators' Insights: A Comprehensive Assessment of PREAP 71
3.2.1. Methodology for the User Evaluation of the Protocol 72
3.2.2. Study Procedure 75
3.2.3. Results and Discussion 76

Chapter 4
PRAT Tool: An Annotation Platform Aligned with PREAP Principles 81
4.1. Platform Overview 82
4.2. Pre-Processing Module 83
4.3. Annotation Module 86
4.3.1. Manual Annotation of Concepts and Relations 87
4.3.2. Annotation Revision 90
4.3.3. Post-Annotation Processes 91
4.4. PR Extraction Module 92
4.5. Analysis Module 95
4.5.1. Quantitative Analysis 95
4.5.2. Data Visualisation 97
4.6. User Evaluation of the Annotation Tool 98
4.7. Comparison with Existing Annotation Tools 99

Chapter 5
Case Study: An Annotation Project for Building a Gold PR-Annotated Dataset 103
5.1. Project Set Up 103
5.2. Performing the Annotation on the Text 106
5.2.1. Concept Annotation 106
5.2.2. Prerequisite Relation Annotation 106

5.3. Annotation Revision 107
5.4. Quality Control and Gold Dataset Creation 109
5.4.1. Inter-Annotator Agreement 109
5.4.2. Annotations Combination 110
5.5. Gold-PR Dataset Statistics 110

Chapter 6
Gold-PR Dataset Applications 113
6.1. Insights from PREAP Annotations 113
6.1.1. Exploring Linguistic Patterns in Educational Texts 114
6.1.2. Time as a Factor in Prerequisite Relations 116
6.1.3. Exploring Text Complexity Based on Concept Learning Levels 120
6.1.4 Lesson Learned from PR Explorations 125
6.2. Automatic Acquisition of PRs from Texts 126
6.2.1. Workflow of the Prerequisite Classification Model 127
6.2.2. Prerequisite Prediction Phase 132
6.2.3. Assessment of PREL Performance 134
6.2.4. Lesson Learned from PR Classification 137

Conclusions 139

Bibliography 145

Acknowledgments

I am deeply grateful to all those who contributed to the research presented in this volume, without whom this work would not have been possible.

First, I express my gratitude to the Department of Modern Languages and Cultures at the University of Genoa for supporting this volume. I also extend my sincere appreciation to the members of the Technology-Enhanced Learning & Digital Humanities Research Programme at the University of Genoa, and the researchers from the ItaliaNLP Lab at the Institute for Computational Linguistics "Antonio Zampolli" of the National Research Council, Italy. Their insights and expertise have been invaluable throughout the course of this work.

Special thanks are due to Professors Ilaria Torre and Simonetta Montemagni for their guidance, support, and encouragement. Their mentorship has been fundamental in shaping both the direction of the research and the contents of this volume.

Lastly, I would like to express my gratitude to the anonymous reviewers for their constructive feedback and valuable suggestions, which significantly enhanced the quality of the work, and to Professor Daria Coppola, the editor of the Ianua collection, for her enthusiasm in appreciating this contribution and the invaluable suggestions for improving its quality.

Introduction

In contemporary times, the educational landscape is undergoing a profound transformation driven by digital innovations. The traditional classroom setup, with its rows of desks and chalkboards, has gradually been joined by alternative, dynamic, and interconnected learning environments. These significant changes in education were driven by the integration of technology into every aspect of learning, hastened by the COVID-19 pandemic. During the sanitary emergency, even the most traditional educational institutions have been compelled to adapt to online education. As a result, there has been an explosion on the Web of educational materials and tools that enhance the educational experience, freely accessible to learners worldwide. The dynamic nature of the internet and the continuous creation of new educational resources make it challenging to estimate the volume of data presently available, but it is a certainty that new materials continue to emerge on a daily basis.

As the Web started housing massive quantities of educational materials and user-generated content, educational technologies became, on the one hand, more advanced in order to keep up with novel needs and, on the other hand, vital for organising and delivering such content to a varied audience of learners. Indeed, nowadays, a learner might be anyone who wants to acquire new knowledge, either in a traditional classroom environment or from home using the novel opportunities of autonomous learning. Clearly, such different populations of learners have their unique educational needs and backgrounds. For instance, lifelong learners of all ages might want to acquire new skills and expand their knowledge using online courses, webinars, and open educational resources for continuous

growth. One of the greatest opportunities offered by digital innovation in education is the enhanced support for individuals with diverse learning needs. Adaptive technologies and personalised learning platforms now cater to students with varying learning styles, or even special needs, ensuring that education is more inclusive and adaptable than ever before. Online learning allows students to tailor their education to their individual needs and schedules, whether it is asynchronous courses that let students learn at their own pace or live virtual classrooms that connect learners with peers and experts worldwide.

However, it is crucial to acknowledge the challenges that come with this transformation. First of all, many digital educational resources, born out of necessity during the sanitary emergency, are disseminated on the Web without the structure and design typically associated with well-thought-out learning plans. Consequently, the internet has transformed into a rich but chaotic educational environment, presenting a significant challenge for learners to navigate effectively. Students often find themselves overwhelmed by the volume of information and struggle to identify reliable and high-quality resources. Meanwhile, teachers may find it challenging to integrate the most valid resources into their courses. This is especially true for large classes, where the different backgrounds of students require individual planning, but addressing individual needs in those contexts is a significant challenge that requires specific training. Indeed, this digital transformation also places new demands on educators, requiring them to adapt to these novel circumstances. Teachers are not just disseminators of information; they are facilitators, mentors, and guides in the digital age. They curate content, design engaging online courses, and foster collaborative learning communities that transcend geographical boundaries. Preparing teachers to confront and manage these new educational realities is an equally formidable challenge. This issue stands at the forefront of Digital Transformation, even recognised by major organisations such as the European Union, which has indeed outlined an action plan related to Digital Education[1].

[1] https://education.ec.europa.eu/focus-topics/digital-education/action-plan

Striking a balance between harnessing the potential of technology and addressing its pitfalls will be essential as we continue to navigate this transformative journey in education. The future of learning is undoubtedly digital, but it is our collective responsibility to make sure it is equitable, accessible, and meaningful for all. In a world where information is constantly evolving and expanding, the need for effective strategies to organise, categorise, and personalise educational content has never been more pressing. This is where educational technologies can step in to offer invaluable support in developing systematic frameworks to enhance the efficiency and effectiveness of handling complex information, such as the domain-specific knowledge contained in educational resources. To meet this multifaceted challenge, knowledge modelling emerges as a highly promising solution. Knowledge modelling involves representing information and relationships in a structured format, such as graph representations, to facilitate understanding and analysis of knowledge. Employing formal frameworks, such as ontologies and semantic networks, knowledge modelling captures key concepts and their interconnections, ultimately creating a systematic and organised representation of knowledge. This representation seamlessly integrates into – and can be leveraged by – educational technologies, enhancing their capacity to deliver tailored and impactful learning experiences.

The significance of managing the knowledge contained in learning resources stems from the wealth of knowledge contained within their content. Educational materials, such as textbooks, lecture notes and teacher summaries, continue to hold immense value as dependable and effective resources for acquiring knowledge, even in the digital era. The significance of managing learning materials stems from the wealth of knowledge contained within their content, carefully curated by experts in the domain as well as in learning practices. Such curated content serves as a guiding path for readers and learners, facilitating knowledge acquisition. However, this comprehensive structure can sometimes overwhelm learners in search of specific information. For example, browsing an entire resource might not be the most efficient choice for a learner seeking information on a particular concept. Imagine the possibilities if we

could uncover the inherent knowledge structure within educational materials, such as a textbook. We could then provide readers with precisely the section of the textbook that aligns with their needs, tailoring their learning experience. Simultaneously, we could identify gaps in explanations and offer supplementary materials to enhance their understanding or provide learning aids to enrich their reading experience. However, the process of automatically extracting the knowledge embedded in textual materials, whether they are concise knowledge capsules or extensive textbooks, and integrating it into educational technologies still presents a complex challenge. This multifaceted process involves identifying the concepts mentioned in the resource and uncovering the relationships that connect them, making it a non-trivial endeavour.

The work presented in this volume aims to contribute to the faceted and wide area of educational technology by addressing, in particular, the issues related to modelling the content of textual educational resources in order to obtain a text-bound representation of the instructional content. The ultimate goal is to propose an approach for modelling the content of educational resources that can serve both researchers and teachers. In particular, the methodology seeks to extract the propaedeutic relationships between the concepts mentioned in textual educational materials. These relationships hold great importance in the educational domain since, as already discussed in the pioneering work of Gagne (1968), they enable the modelling of the most effective sequence in which concepts should be acquired by a learner to ensure content understanding while preventing confusion.

The significance of knowledge structures representing these propaedeutic (also called prerequisite) relationships has been demonstrated across various educational scenarios to the point that automatic identification of prerequisite relationships between concepts has been identified as one of the key requirements for modern, large-scale online education by Gasparetti *et al.* (2018); Talukdar and Cohen (2012); Manrique *et al.* (2019), just to name a few. This relevance is confirmed by the fact that such representations are key elements that enable technology-enhanced learning systems to incorporate knowledge about domains and enhance their functionalities.

Along this line, we mention the work of authors that explored these possibilities for, e.g., automatic synthesis of study plans (Gasparetti, Limongelli, and Sciarrone 2015; Agrawal, Golshan, and Papalexakis 2016; Zhao *et al.* 2020), reading list generation (Koutrika, Liu, and Simske 2015; Gordon *et al.* 2017), and automatic educational content creation (W. Lu, Ma *et al.* 2019). Additionally, knowing the knowledge structure of a domain allows locating students' competence level (i.e., what they already know about the subject matter) with respect to the knowledge structure (Osenova and Simov 2010) and supporting them with personalised recommendations. Hence, the task of identifying the prerequisite relationships within an educational resource lays the foundations to assist the personalisation of the learning experience by offering the necessary resources to surpass the "one size fits all" philosophy.

In this volume, we will present a comprehensive study aimed at modelling the content of educational resources based the prerequisite dependency structure underlying the content of educational texts. The methodology proposed is based on a thorough investigation of the main issues related to uncovering the prerequisite structure of an instructional resource. Accordingly, the volume is organised into main three parts.

The first part discusses the related work and background research. Specifically, Chapter 1 presents the basic notions related to concepts and propaedeutic relations: we outline the main properties of concepts and prerequisite relations and provide examples to clarify how they might be represented in different instructional materials. In Chapter 2 we complete the overview of the research context by discussing the fundamental principles of two main research areas: the science of annotation and the literature related to the manual and automatic extraction of prerequisite relations.

The second part of the volume is focused on the methodology for educational content modelling and prerequisite relation identification. The proposed methodology offers solutions for modelling the content of educational texts and extracting the prerequisite structure of the concepts therein. Its most fundamental principle is the choice of modelling only the content of textual instructional resources rather than the abstract domain knowledge. Although

ideally simple, and possibly close to the way in which students acquire knowledge from learning materials, formally incorporating such principle into prerequisite identification is a novelty of the prerequisite annotation methodology. In Chapter 3, we introduce the *PRErequisite Annotation Protocol* PREAP, a protocol intended as a knowledge engineering procedure for the annotation of prerequisite relations. PREAP represents a novel knowledge modelling approach for uncovering the prerequisite structure of educational materials, thus it is distributed along with an annotation manual that includes guidelines providing recommendations and a set of good practices for its application to educational texts. This is done to support the protocol dissemination and reuse in future annotation projects. Chapter 4 presents PRAT (*PRerequisite Annotation Tool*), the annotation interface designed to support the application of PREAP principles on educational texts. Both chapters 3 and 4 should be intended as user guides to facilitate the use of the protocol and tool.

The last part of the volume shows use case scenarios and possible applications of the PREAP protocol and PRAT tool annotation interface on the part of researchers and teachers. Specifically, Chapter 5 reports an annotation project that led to the annotation of a chapter of a computer science textbook. Chapter 6 demonstrates how the annotated dataset produced in the project can be used by teachers and researchers depending on their purposes and goals, as well as their satisfaction with respect to the objectives achieved. The last chapter of this book concludes the work: it summarises the contributions of the research presented in the book, and discusses limits, future improvements and applications of this research.

Given the nature of the content at hand, this volume is intended for a diverse audience. Researchers will find in these pages a methodology to model educational content, aligning it with a fundamental pedagogical principle: that of didactic propaedeutics. School and university educators can readily integrate this methodology into their daily work. In the classroom, it serves as a collaborative and hands-on practice to enhance knowledge assimilation through direct engagement with information. In the context of lesson planning, this method provides a powerful tool for swiftly navigating

diverse learning materials, a resource not only for educators but also for students seeking effective study aids. Furthermore, we believe that educational editors can extract substantial value from this approach as it facilitates the comparative analysis of multiple educational materials.

The content of this volume, the challenges addressed, and the findings reported stem from an interdisciplinary work: the proposed knowledge modelling approach draws upon insights from the fields of Education, Information Extraction and Natural Language Processing to address the problem of uncovering the knowledge structure of educational texts with the goal of identifying the propaedeutic relationships between concepts in educational materials. The novel methodology for knowledge modelling that emerges from this work promises to empower users by facilitating their navigation and access to educational resources. While its most natural application lies in the creation of tools capable of harnessing this structured knowledge, its impact extends far beyond. It reaches into the broader learning ecosystem and provides invaluable support to educators at all levels.

IANUA

lingue, culture, educazione

L'elenco completo delle pubblicazioni
è consultabile sul sito

www.edizioniets.com

alla pagina

https://www.edizioniets.com/view-Collana.asp?col=Ianua.%20Lingue,%20culture,%20educazione

Pubblicazioni recenti

1. Daria Coppola, *Educazione linguistica e insegnamento*, 2019, 214 pp.
2. Edith Cognigni, *Il plurilinguismo come risorsa. Prospettive teoriche, politiche educative e pratiche didattiche*, 2020, 212 pp.
3. Victoriya Trubnikova, Benedetta Garofolin, *Lingua e interazione. Insegnare la pragmatica a scuola*, prefazione di Matteo Santipolo, 2020, 168 pp.
4. Francesca Gallina (a cura di), *Itinerari di formazione: spunti di riflessione per i docenti di lingue seconde*, 2021, 300 pp.
5. Roberta Ferroni, *Strategie e dinamiche di comunicazione. Verso un'educazione linguistica strategica plurilingue e pluriculturale*, 2022, 180 pp.
6. Paolo Della Putta, Silvia Sordella, *Insegnare l'italiano a studenti neo arrivati. Un modello laboratoriale*, 2022, 208 pp.
7. Raffaella Moretti, Davide Delle Chiaie (a cura di), *Educazione linguistica, documenti programmatici e proposte didattiche. Alcuni esempi dall'Italia e dal mondo*, 2024, 220 pp.
8. Chiara Alzetta, *Unlocking Knowledge in the Digital Age. A Guide to Modelling Propaedeutic Relations in Educational Texts*, 2024, 164 pp.

Edizioni ETS
Palazzo Roncioni - Lungarno Mediceo, 16, I-56127 Pisa
info@edizioniets.com - www.edizioniets.com
Finito di stampare nel mese di giugno 2024