



Intelligent Textbooks: Themes and Topics

Sergey Sosnovsky¹(✉), Peter Brusilovsky², and Andrew Lan³

¹ Utrecht University, Princetonplein 5, 3584 CC Utrecht, The Netherlands

s.a.sosnovsky@uu.nl

² University of Pittsburgh, 135 North Bellefield Avenue, Pittsburgh, PA 15260, USA

³ University of Massachusetts Amherst, Amherst, MA 01003, USA

Abstract. The transition of textbooks from printed copies to digital and online formats has facilitated numerous attempts to enrich them with various kinds of interactive functionalities, link them with external resources or extract valuable information from them. As a result, new research challenges and opportunities emerge that call for the application of artificial intelligence methods to enhance digital textbooks and students' interaction with them. In this report, we summarize our workshop series on Intelligent Textbooks from 2019 to 2022. We focus on the evolution of the topics covered in the workshops' programs and identify the main themes that have been proposed by the intelligent textbooks community.

Keyword: Intelligent textbooks

1 Historic Overview

This year, the workshop on Intelligent Textbooks is organized for the fourth time. It builds on the success of the three previous workshops conducted as a part of the satellite program of the International Conference on Artificial Intelligence in Education in 2019, 2020 and 2021. Overall, 36 (20 full and 16 short) contributions were published in the workshop proceedings over these years; 14 intelligent textbooks prototypes and technologies were showcased as demo presentations.

At the first workshop in 2019, a majority of accepted submissions have focused on various aspects of making textbooks adaptive through navigation, recommendation, or problem solving support. Other popular topics were integration of interactive content, orchestration of learning around digital textbooks and automated analysis of the textbook content for various purposes.

In 2020, adaptivity and interactivity remained important aspects of intelligent textbooks. However the trend on leveraging machine learning, natural language processing and semantic technologies to automate processing or construction of textbook content became much more prevalent. Several papers and demos have presented approaches for textbook generation, transformation, linking to external content and extraction of knowledge from textbooks.

The third workshop explored a variety of topics. Two new trends that separated it from its predecessors were: demonstration-based papers presenting prototypes of domain-oriented textbook applications and projects exploring automated approaches to extract from textbooks different kinds of learning objects.

In 2022, the workshop has attracted ten submissions. And while the review process is still ongoing at the moment of writing this report, we can already provide a preliminary overview of the topics covered by the potential program of the Intelligent Textbooks 2022 workshop and relate them to the main themes explored by the workshop participants over the last 4 years. Table 1 presents the summary of these topics together with counts of corresponding papers presented at each of the workshops from 2019 to 2022¹.

2 Workshop Themes

Intelligent interfaces for online textbooks could be considered as an “end product” of several other research directions. This is where the “intelligence” reaches the readers augmenting their textbook experience with a range of functionalities not available in traditional textbooks. While these new functionalities are typically based on modeling and knowledge processing technologies, the main concern of papers focused on intelligent interfaces is not these foundation technologies, but how to present the new functionality to the readers. Among interface-focused papers presented at the past workshops, several papers focused on open student models - a visual presentation of learned knowledge and progress (computed by AI student modeling algorithms) to the learners. Other papers focused on such technologies as personalized guidance within the textbook, run-time recommendation of external learning content, and augmenting user interactions with interactive tools such as concept maps and chatbots.

One of the appealing ways to extend online textbooks with additional functionalities not available in traditional textbooks is adding so-called “smart content” items - interactive activities, which engage students into learning by doing (rather than just by reading). Since most of these “smart content” items are in fact learning exercises supporting automatic assessment, working with smart content also enables the students to check their content understanding and receive feedback on their performance. While only a fraction of “smart content” activities could be intelligent by themselves (i.e., ITS problems with intelligent scaffolding) the work on smart content is important for intelligent textbooks as a whole. Most importantly, “smart content” activities produce a much richer volume of learning data that is crucial for both student modeling and knowledge extraction. The work on smart content has been well-represented at the workshop, especially in papers focused on computer science textbooks, the domain where smart content is becoming increasingly popular. The key research issues examined in the workshop papers focused on smart content are the infrastructure (how to connect an interactive item to a textbook while maintaining authentication and data collection) and content matching (how to assign a smart content

¹ The numbers for 2022 are projections.

Table 1. Number of papers under each topic over the years.

Topic/Year	2019	2020	2021	2022
Intelligent interfaces	5	2	1	0
Smart content	1	1	2	4
Knowledge extraction	1	2	0	1
Learning content construction	0	1	3	2
Intelligent textbook generation	1	2	0	1
Interaction mining and crowdsourcing	1	1	3	1
Domain-focused textbooks and prototypes	1	0	2	0
Miscellaneous	4	0	2	1

item to its proper place within a textbook). Smart content grouped together with the intelligent interfaces forms the first large research theme that is focused on textbook enrichment with extra features and functionality.

A textbooks can be seen not only as an object of enrichment, but also as a source of domain knowledge whereas intelligent textbooks are often written in formats that enable automated extraction of this knowledge. Works published in our workshop have covered many different subtopics related to knowledge extraction. Several papers discussed the construction of knowledge representations from textbooks, using a wide range of approaches, from human-labeling to automated construction via both traditional feature-based natural language processing (NLP) techniques and modern embedding-based NLP techniques. Several works focused on extracting prerequisite relations and studied the noisy nature of this process. The extracted knowledge representations can be useful in many downstream tasks, including entity relationship visualization, matching across textbooks and between textbooks and external learning content, monitoring ideas in student discourse, and personalized learning support.

In recent years, with rapid development in neural language models in NLP research, especially generative ones such as GPT, our workshop has seen an uptick in the number of works on automated content construction for intelligent textbooks using these models. These models are highly capable of effectively transferring what they learn from pre-training on web-scale text to different contexts, resulting in high levels of fluency and consistency of the generated text. The vast majority of these works use generative language models to automatically produce assessment questions for intelligent textbooks. Works have focused on generating both different formats of questions, from multiple-choice to short-answer, and different types of questions, from factual to reasoning. Additionally, several works have also considered generating other types of learning content, such as textbook indices and concept definitions. Learning content construction and knowledge extraction constitute the second theme of research on intelligent textbooks aimed at utilizing the textbooks themselves as a source of an added value, be it (elements of) domain knowledge or (elements of) learning material.

Another stream of research that has been explored by the workshop participants is textbook generation and assembly. The works on this topic are characterized neither by a uniform methodology, nor by the common attributes of the final product. Yet, they have had a common objective - propose a technology that facilitates creation of digital textbooks from external resources. These resources range from Wikipedia content to specially-formatted material such as Jupyter notebooks to existing digital textbooks in PDF format. The proposed technologies ranged from a community-oriented authoring platform for digital textbook assembly to a framework for automated generation of intelligent textbooks enriched with semantic and adaptive services. Generally speaking, textbooks generation represents the third workshop theme that focuses on a textbook itself as the final product.

Most technologies presented at the workshop are aimed at developing and/or supporting complex applications built around textbooks. These applications provide their users with various methods of interaction with the actual content of textbooks, integrated smart content items and or value-adding services enriching textbook functionality. Evaluation of these interactions to data-mine typical patterns, model parameters or characteristics of students has been a common topic for several workshop contributions. Another related line of research has become an organization of interaction between users and textbooks in such a way that the textbook application could crowdsource execution of challenging tasks to its users. The outcomes of these interactions would provide the textbook application with the elements of crowdsourced “intelligence” (e.g., concept map-based exercises helping extract types of relations between domain terms, or text highlighting behavior helping to train a student modelling approach). Interaction mining and crowdsourcing constitute the fourth large theme of research on intelligent textbooks that is concerned with the link between the textbook and the user and aims at extracting “intelligence” from it.

Intelligent textbooks are broadly applicable in many subject domains. Therefore, tools that are customized to the specifics of each subject domain are necessary. For example, in the domain of reading education, intelligent textbooks can benefit from embedded tools for authoring support, student modeling, personalized activity, and mini-games. In the domain of quantum cryptography, intelligent textbooks can benefit from built-in coding environments, interactive visualizations, and self-graded quizzes, driven by learning styles and student objectives. In medical domains such as dentistry and cardiovascular anatomy, intelligent textbooks can benefit from chatbots built into the textbook to ask learners questions and interact with them or take the form of a mobile application helping students learn the logical connections behind key technical terms.

Finally, over the years, the workshop has attracted several contributions that are hard to categorize under a single label. Some of these papers proposed technologies that are too unique (e.g., a prototype using paper-based workbooks and a mobile application scanning and grading hand-written solutions). Others focused on very particular tasks (e.g., a new format facilitating representation and retrieval of math formulae). In addition, we have had several position papers envisioning new way to organise and orchestrate lessons around intelligent textbooks of the future.