Building a better eTextbook

M. L. Hollingsworth, N. H. Narayanan

Abstract— **ETextbooks are increasingly more common in the classroom, yet no consensus on the best eTextbook design has emerged. There are several features that characterize eTextbook products currently on the market. This paper will describe six adaptive and intelligent features that we believe, based on an extensive survey of commercial and research-based eTextbooks, are important to learning. This is supported by the existence of research-based results in the literature showing that these features improve student learning when implemented in eTextbooks or other learning systems. The paper concludes with a discussion of what research challenges need to be met to build a next generation eTextbook.**

Index Terms— **E-textbook design; Electronic publishing; Artificial intelligence; Natural language processing; Knowledge representation; Computer aided instruction**

I. Introduction

As part of a research project on designing and evaluating the next generation of eTextbooks, we undertook a comprehensive review of eTextbooks, some of which are products of academic research while others are commercial products. This effort is the first step toward envisioning a next generation eTextbook that relies on existing technology but addresses technological and interactional gaps of current eTextbooks.

One of the earliest visions of an interactive, electronic learning resource for children was presented by Alan Kay as the "DynaBook." We live in the world Kay envisioned where we would use these devices and, "… a global information utility such as the ARPA network …" [14] would bring the information of the world into our homes and offices. Modern tablets have moved well past Kay's original goals of storage to hold, "…about 500 ordinary book pages…" or "…several hours of audio … files" [14].

This article focuses on five features that emerged as important from the aforementioned literature review. tour selection was influenced by the fact that research conducted on the teaching efficacy of these features is available in the literature. The features are concept maps found in the eTextbook CoMPASS [1], peer collaboration and assessment used by the WileyPLUS Learning Space package [8], dynamic pages provided by Wolfram Research's Computable Document Format [2], adaptive learning featured by CogBooks [10], and Inquire: Biology's capability of answering student questions using natural language [11]. In the next section, we discuss each of these features followed by a discussion of the limitations of the tools we have and what innovation could propel future eTextbooks to a new level. We conclude the paper by presenting a vision of the next generation eTextbook. It should be noted that the features we describe are not the only features offered by each of the systems; reviewing all features of each system or reviewing all systems that contain a specific feature goes beyond the scope of this paper.

Michael Hollingsworth is a Computer Science and Software Engineering Graduate from Auburn University, USA (hollimi@auburn.edu).

N. Hari Narayanan is a Professor at the Department of Computer Science and Software Engineering of Auburn University, USA (naraynh@auburn.edu).

II. Concept Map

A concept map is a graphical way of organizing information to show how each concept relates to another concept, which allows students to approach the subject matter in a nonlinear way. This is different from a traditional textbook, which presents information in a linear fashion one chapter after another.

CoMPASS is a non-commercial research project developed at the University of Wisconsin, and its topics target 7th-8th grade science subjects (the project is currently undergoing a rework and has been renamed VidyaMap.). CoMPASS presents the learning material in the form of a concept map, which shows relationships among concepts using a directed graph. As of now, CoMPASS only covers small subsets of science subjects: simple machines in physics and composting in biology. CoMPASS also features simple natural language query processing, which allows the students to interact with the eTextbook in a more natural manner. For example, students ask a question in English and the system will show a concept map showing concepts in its knowledge base that matched words in the query.

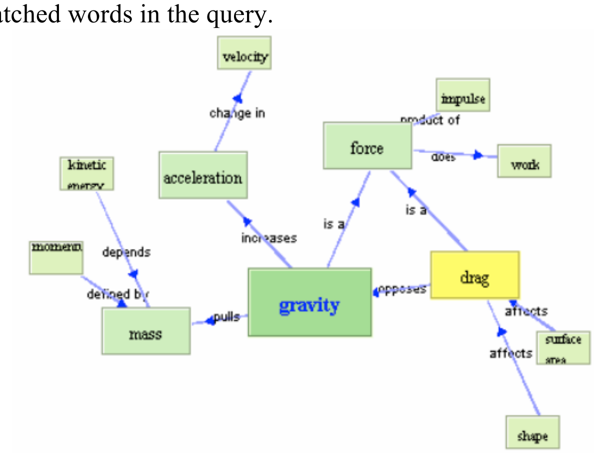

Fig. 1. A concept map from an early version of CoMPASS [6] showing how gravity interacts with drag, force, acceleration, and mass.

The knowledge base that is at the backend of CoMPASS is quite complex. It requires a domain expert to construct it. This knowledge base takes a significant amount of time to construct, yet the instructor cannot modify it in anyway.

Studies suggest that students learn the material better when using CoMPASS over a traditional textbook [1], [6].

III. Assessment

Assessment is typically implemented as a posttest that students take after reading a section of the material. This posttest can be used by either the student, instructor, or both to determine if the material should be reviewed. Using this information, the instructor can detect trends in the class to find areas that need reinforcing. Students can instantly get specific feedback on the extent of their understanding of the material for personal review.

WileyPLUS is a commercial product that uses assessment to reinforce learning. WileyPLUS is a learning program used with several textbooks, specifically for high school and higher education students. Numerous topics are in existence ranging from architecture to science to foreign language.

WileyPLUS' user interface is a familiar social media like interface. It presents the student with a dashboard. Among the features of this dashboard are 'feeds' where students and/or the instructor can highlight topics from class for everyone in the class to read and comment. The dashboard also contains class rankings, so each student can see how their grades on the posttest questions rank among their peers.

WileyPLUS does not allow for custom knowledge base creation. The particular textbook is the basis for all knowledge base information for the particular WileyPLUS module. It does allow the instructor to highlight topics for the class and construct assessment quizzes.

Studies have suggested that WileyPLUS improves the grades of the student who use the product over those students that use a traditional textbook [8], [9].

IV. Dynamic Widgets

Wolfram Research developed the Wolfram Computable Document Format (CDF). Wolfram CDF is a document format which allows embedded dynamic widgets. CDF positions itself as superior to other document formats such as PDF, JavaScript/HTML, Flash, and Microsoft Office.

Figure 2 shows an example widget. The student can input different data values using the sliders and buttons, and the graphs are dynamically redrawn by the eTextbook built using the CDF. These widgets allow authors to incorporate dynamic simulations that students can manipulate and interact with while reading the eTextbook.

There have been no studies directly linking a book based on CDF to improved learning; however, studies suggest that using simulations can improve learning [12], [13].

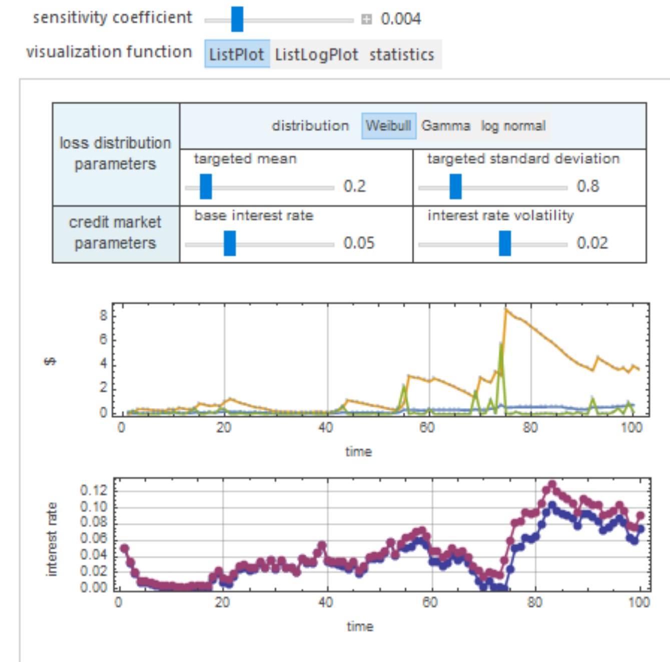

Fig. 2. An example widget from CDF [2].

V. Adaptive Learning

Adaptive learning is defined as the ability of an eTextbook to change (adapt) the way it selects and displays the content material to the student. An adaptive eTextbook should contain the logic to detect when a student has a strong grasp of the basics of a subject and push the individual further than other students who are behind. Alternately, if a student is struggling with a topic, the eTextbook should be able to detect that also, and present earlier material to help the student catch up. Thus, the material presented is personalized for each student.

CogBooks [19] is an adaptive eTextbook. Unlike other systems discussed, CogBooks also allows the instructor to create the knowledge base from scratch. The knowledge base is structured as a graph, and the instructor can create nodes and edges[1]. The nodes will contain snippets of text from a textbook or the instructor's custom text. The edges are posttests to assess how well the student understands the information in the node, which works much like the assessment in WileyPLUS.

Another interesting aspect related to adaptivity is that CogBooks do not present information in a linear fashion. The instructor-created directed graph allows the system to guide students along different paths through the graph depending on how well they do on each posttest assessment. For example, if a student starts in node A and scores a 30% on the posttest, he is returned to A. A 60% score might move the student to node C, whereas an 80% score might move the student to node R. A 100% would take the student to yet another node Q. The way the student navigates is determined by a combination of CogBooks machine learning algorithm as well as the instructor created graph.

There have been no studies of effectiveness of CogBooks per se, but studies of other learning systems have suggested

[1] Nodes are defined as "modules" and edges are defined "outcomes" in the CogBooks knowledge base creator

that adaptive learning leads to better student outcomes than traditional linear teaching [10],[11].

VI. Intelligent Question Answering

Inquire: Biology is a product of Project Halo[2]. One powerful feature of Inquire: Biology is that it allows students to pose natural language questions and to receive a natural language answer even if the answer does not exist anywhere in the text. Inquire: Biology can infer the answer and dynamically construct natural language text due to the combination of a complex knowledge base and powerful Artificial Intelligence (AI).

Inquire: Biology is based on the ninth edition of the *Campbell Biology* textbook. It only contains a small subset of the entire textbook. This is probably due to the immense amount of work needed to convert human readable text into an AI readable directed graph. Much like the CoMPASS knowledge base, this work must be completed by a domain expert. Each sentence from the textbook must be broken down word by word to create a graph with nodes that contain a single concept and edges that show how one concept relates to another.

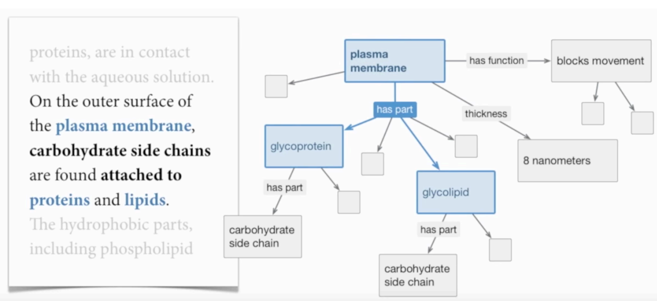

Fig. 3. An example showing how one sentence from a traditional textbook is transformed into a concept map that becomes part of the knowledge base of Inquire: Biology. [5]

An example of a question a student might ask is, "What is the structure of the plasma membrane?" The answer to this question is not directly in the textbook; however, due to the powerful AI and complex knowledge base, Inquire: Biology can infer the answer, give additional facts about the plasma membrane, provide figures from the book about the plasma membrane, and give examples of how one concept interacts with other concepts (e.g. how a plasma membrane facilitates cell-cell recognition) [11]. Thus, the power of this feature is that the eTextbook is able to piece together various facts and figures from the underlying knowledge base to dynamically construct a cohesive answer to a student question.

A study compared a traditional text, an electronic version of the text, and Inquire: Biology in a course. Student grades were consistently the highest with Inquire: Biology. Also, no students in the Inquire: Biology group received a grade of D or F, while approximately 33% of students in the other two groups scored either a D or F [4]. It appears that Inquire: Biology is no longer under further development.

[2] Project Halo started in 2004 sponsored by Vulcan Inc. with the goal to create a "Digital Aristotle" i.e. a system that would have a broad scientific knowledge base. The combination of this knowledge base and AI could infer the answer to novel questions that people could ask. Project Halo is currently called Aristo. It is being developed by the Allen Institute for Artificial Intelligence. Project Aristo seeks, much like Inquire: Biology, to answer novel questions. It covers more topics than just biology.

VII. The missing Pieces

The benefits of AI and powerful natural language querying and answering in eTextbooks are almost entirely absent in current eTextbooks. The ability for a student to ask a book a question using natural language and to receive a dynamically generated answer to this question is the holy grail of eTextbooks. Perhaps, the reason that we do not see this feature in more eTextbooks is due to the complexity of the underlying computational problems of knowledge representation, inference, and natural language processing. The other significant lacuna is that the knowledge base undergirding an eTextbook must be manually constructed for each subject, which requires many hours of highly skilled labor. The answers provided by Inquire: Biology are the type of answers a student would expect from an instructor. With powerful AI, these answers would be available to each student in a way similar to each student having his/her personal instructor on call at all times. This underscores the work that needs to be done in order to transform a textbook into a machine-readable, graph-structured knowledge representation. This may require the creation of a universal standard for eTextbook knowledge representation. Once such a standard is established, a tool to turn any piece of text into this representation needs to be designed. This would open the door to turn any textbook into a machine-readable knowledge representation upon which the aforementioned features of question answering, personalization, assessments, dynamic interactions and learning analytics could be built. Addressing these challenges would signify progress toward the ultimate goal of adaptive and intelligent learning systems - giving every student access to a virtual instructor embedded in an eTextbook any place and any time.

VIII. Concluding Discussion

WileyPLUS and CogBooks both have a self-assessment tool. From the instructor side, WileyPLUS requires the instructor to evaluate the scores and make adjustments based on these scores. From the student side, WileyPLUS requires students to compare their scores to that of the class and make a decision as to what they should review. CogBooks' AI uses the self-assessment scores to tailor the material from the book displayed to the student.

However, both are not using all available data to its full potential. The instructor should not have to manually review the scores of all the students in the class to find trends. This task should be handled by learning analytics. The instructor should not be limited to pulling this data from a single class using the eTextbook. Instead, trends should be computed, displayed, and acted upon (automatically) from the aggregated results of all their classes using the same eTextbook. CogBooks uses the assessment results to construct a custom book for the user. This information could be used to find trends in the class where the text is lacking or unclear to the students. This kind of feedback would free up the instructor to make modifications to the content and to make learning objectives clear while the eTextbook takes on the task of automatically tailoring itself to meet the needs of individual students.

CogBooks, CoMPASS, and Inquire: Biology all allow the student to interact with the material in a nonlinear fashion, but this feature is enabled by an underlying knowledge base, which requires a large amount of manual effort from a domain expert to build. In the case of CogBooks, the instructor is tasked with, essentially, writing a textbook. In the case of CoMPASS and Inquire: Biology, a complex directed graph is constructed by a domain expert from an existing textbook.

Two improvements could be made to this process. One, CoMPASS and Inquire: Biology lack the adaptive learning experience offered by CogBooks. They already have knowledge structured as complex directed graphs but do nothing to exploit this structure to customize the learning experience based on feedback from students. Two, the process of transforming an existing textbook into a machine-readable directed graph could be semi-automated through the use of natural language processing and AI. This would eliminate the need for a domain expert to create the graph (in the case of CoMPASS and Inquire: Biology) or write a textbook (in the case of CogBooks). Instead, the expert's role will be the less labor-intensive, which is one of reviewing and revising the knowledge base constructed by the machine. This is a nontrivial task, but currently, it appears that no one is looking into this problem.

Having interactive games/simulations in eTextbooks seems to an important feature. Decades of research suggests that this improves student learning [15], [16], [17], [18]. While there are dozens of eTextbooks available, very few offer this feature. Facilitating rich interactions with the learning material in the form of games and/or simulations should be a standard feature of future eTextbooks.

In summary, we have identified the following features as essential for the next generation of eTextbooks:

1. *Learning Analytics:* The ability to collect, store, and analyze student assessment and interaction data over time. use it to both adapt/personalize the student experience, and to display trends and student progress to the instructor.
2. *Knowledge Engineering:* The ability to automatically or semi-automatically construct a structured knowledge base that captures the learning topics and content by processing existing text-based resources.
3. *Learning Interactions:* Provide a rich learning interface to the students that include nonlinear navigation, natural language question answering, interactive charts and graphs, simulations and learning games, as well as traditional forms of self-assessment.

References

[1] Puntambekar, Sadhana, Agnes Stylianou, and Roland Hübscher. "Improving navigation and learning in hypertext environments with navigable concept maps." Human-Computer Interaction 18, no. 4 (2003): 395-428.

[2] Uses and Examples of the Computable Document Format (CDF)Available: http://www.wolfram.com/cdf/uses-examples/

[3] EPUB 3 Overview Recommended Specification 26 June 2014 Available: http://www.idpf.org/epub/301/spec/epub-overview.html#sec-intro-overview

[4] Ongoing Research Available: http://inquireproject.com/#research

[5] Technology Available: http://inquireproject.com/#tech

[6] Puntambekar, Sadhana, and Agni Stylianou. "Designing navigation support in hypertext systems based on navigation patterns." Instructional Science 33, no. 5-6 (2005): 451-481.

[7] Tawil, Norngainy Mohd, Nur Arzilah Ismail, Izamarlina Asshaari, Haliza Osman, Zulkifli Mohd Nopiah, and Azami Zaharim. "Learning Process in Mathematics and Statistics Courses towards Engineering Students: E-learning or Traditional Method?." Asian Social Science 8, no. 16 (2012): 128.

[8] Tawil, Norngainy Mohd, Nur Arzilah Ismail, Izamarlina Asshaari, Haliza Osman, Zulkifli Mohd Nopiah, and Azami Zaharim. "Comparing lecture and e-learning as learning process in mathematics and statistics courses for engineering students in Universiti Kebangsaan Malaysia." Procedia-Social and Behavioral Sciences 60 (2012): 420-425.

[9] VanLehn, Kurt. "The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems." Educational Psychologist 46, no. 4 (2011): 197-221.

[10] Steenbergen-Hu, Saiying, and Harris Cooper. "A meta-analysis of the effectiveness of intelligent tutoring systems on college students' academic learning." Journal of Educational Psychology 106, no. 2 (2014): 331.

[11] Watch our introduction video Available: http://inquireproject.com

[12] Moreno, Roxana, and Richard Mayer. "Interactive multimodal learning environments." Educational Psychology Review 19, no. 3 (2007): 309-326.

[13] Holzinger, Andreas, Michael D. Kickmeier-Rust, Sigi Wassertheurer, and Michael Hessinger. "Learning performance with interactive simulations in medical education: Lessons learned from results of learning complex physiological models with the HAEMOdynamics SIMulator." Computers & Education 52, no. 2 (2009): 292-301.

[14] Kay, Alan C. "A personal computer for children of all ages." In Proceedings of the ACM annual conference-Volume 1, p. 1. ACM, 1972.

[15] Vogel, Jennifer J., David S. Vogel, Jan Cannon-Bowers, Clint A. Bowers, Kathryn Muse, and Michelle Wright. "Computer gaming and interactive simulations for learning: A meta-analysis." Journal of Educational Computing Research 34, no. 3 (2006): 229-243.

[16] Andrews, P. V., J. Schwarz, and R. D. Helme. "Students can learn medicine with computers. Evaluation of an interactive computer learning package in geriatric medicine." The Medical Journal of Australia 157, no. 10 (1992): 693-695.

[17] Bangert-Drowns, Robert L., James A. Kulik, and Chen-Lin C. Kulik. "Effectiveness of computer-based education in secondary schools." Journal of computer-based instruction (1985).

[18] Rutten, Nico, Wouter R. van Joolingen, and Jan T. van der Veen. "The learning effects of computer simulations in science education." Computers & Education 58, no. 1 (2012): 136-153.

[19] Miranda, Sergio, Giuseppina Rita Mangione, Francesco Orciuoli, Matteo Gaeta, and Vincenzo Loia. "Automatic generation of assessment objects and Remedial Works for MOOCs." In Information Technology Based Higher Education and Training (ITHET), 2013 International Conference on, pp. 1-8. IEEE, 2013.