A7134: Textbook Open Knowledge Network

Michael Genesereth genesereth@stanford.edu

Richard Baraniuk richb@rice.edu

Mark Musen musen@stanford.edu

Craig Heller hcheller@stanford.edu

Vinay K. Chaudhri vinayc@stanford.edu

Overview

College students today face the challenge of mastering concepts in the new subject areas, relating those concepts across multiple disciplines, and one size fits all nature of textbooks. Intelligent Textbooks (ITB) using Artificial Intelligence (AI) and knowledge graphs (KG) solve these problems by allowing students to dynamically interact with the textbook content, increasing their ability to understand concepts, increasing engagement, and thereby, improving academic performance. ITBs offer students easy access to definitions and descriptions of concepts, make connections across different sections of the syllabus, and allow students to pose their own questions.

Description

Initial trials of ITBs that utilize KGs have been found to improve student grade outcomes by a full letter grade over the control group that was using a conventional textbook. ITBs have been found especially helpful for underperforming students, thus, broadening participation.

The potential of ITBs to facilitate better learning has been extremely difficult to realize without major investments of time, money, and expertise. The reason is that KGs are currently constructed using human subject-matter experts in a process that is extremely expensive and time consuming. Due to the large investment required, publishers and ed tech providers keep their KGs proprietary, eliminating their utility outside of the scope of the project for which they were created.

The primary outcome of our project is an open source Textbook Open Knowledge Network (TOKN) that can be freely used for creating ITBs and a variety of education technology applications. A secondary outcome is a novel process and tools for creating the KGs that combine automatic construction of a KG with validation by humans to ensure high accuracy. We envision a community of educators who would co-create TOKN, and eventually take the ownership for its future development and evolution.

We will use the new approach for KG construction developed in our project for Biology and Psychology undergraduate textbooks that are published by the open source publisher OpenStax. We will eventually integrate the ITB technology into the full OpenStax library of 41 textbooks, which has the potential to influence millions of students. OpenStax has further developed a robust ecosystem of 54 commercial partners (including every major publisher save one) who use OpenStax materials within their educational technologies and a further 30 major college and university systems who are using OpenStax to institutionalize open educational resources (OER) at their campuses.

Our KG construction tools will be integrated with Stanford's Protégé environment. Protégé is one of the most widely used knowledge authoring environments in the world today. In 2019 alone, Protégé was downloaded 145,000 times, and its web version hosts over 70,000 projects with over 50,000 user accounts.

Differentiators

An ITB relies on an explicit representation of knowledge in a textbook that matches human understanding and enables precise reasoning with it. The current narrative in AI is dominated by machine learning and natural language processing that achieve scale by sacrificing either accuracy or expressiveness. Such a compromise is acceptable in applications such as search, recommendation systems, machine translation, etc. In education, our domain models must be accurate and expressive. The textbooks need to be nearly 100% accurate. Any computer-based tools that will eventually be as good as human tutors have to use an explicit model of the knowledge of the domain. These characteristics are also shared by many other applications, for example, income tax calculations and automated enforcements of a contract.

Our team is a world leader in research, application and education of formal knowledge representation methods. We have created strong partnerships with non-profit and for-profit organizations in the field of textbook publishing, educational technology, and academic search, thus, substantially enhancing the probability of success of our Phase II effort.

Road Map

Activity	Milestone
KG Construction Tool	9 months
KG Tool Evaluation for Biology	12 months
KG Tool Evaluation for Psychology	15 Months
ITB incorporating the KG	18 months
ITB Evaluation	24 months

There will be two primary products of our project: a KG construction tool, and TOKN that contains KGs for a few chapters from Biology and Psychology textbooks from OpenStax. We will evaluate the KG construction tool for OpenStax Biology and Psychology textbooks, and measure the time and effort required per chapter. We will evaluate the resulting KGs by incorporating them into an ITB and measure its impact on student learning and engagement.

Partnerships

The foundation of the team is a partnership between Stanford University and OpenStax at Rice University. We bring together strong prior research on ITBs and KGs from Stanford with the vision to reimagine and reinvent textbooks from OpenStax.

To ensure that our innovations impact the commercial textbook industry, we have formed a collaborative relationship with Macmillan Learning. They will evaluate ITBs for two of their textbooks, gather user feedback, and help guide our future development.

To ensure that TOKN can be useful outside the context of an ITB, and for educational technology products, we have teamed with Educational Testing Service (ETS). ETS will specify the requirements of KGs for item generation, and once the KG is developed, evaluate it for item generation and question scoring outside the context of an ITB.

To leverage and contribute to publicly available Big Data, we have formed a collaboration with Microsoft Academic Graph (MAG). In our collaboration with MAG team, we will investigate if their technology can help bootstrap the taxonomy for TOKN, and if TOKN could be incorporated into MAG ensuring its wide usage and contribution to the task of academic search.

Intellectual Property

TOKN, the KG construction platform, and new ITB applications will be released under an open source license. The existing ITB platform, called Inquire, is owned by Vulcan Inc, and Stanford has rights to sublicense it. Automatic Item Generation algorithms are owned by ETS and will be licensed to the Stanford team.

CACCEL PHASE II - Track A1 (Open Knowledge Network):
A7134: Textbook Open Knowledge Network
Stanford University
NSF Convergence Accelerator Pilot (NSF 20-055)

Project Summary

OVERVIEW

The United States faces a crisis in science, technology, engineering, and mathematics (STEM) education. Properly educating the STEM leaders of tomorrow requires moving beyond the centuries-old, ingrained paradigm of education that views the process of learning as a one-way street in which knowledge is transmitted from teacher to learner via paper textbooks and lectures.

Intelligent Textbooks (ITB) using Artificial Intelligence (AI) and knowledge graphs (KG) solve these problems by allowing students to dynamically interact with the textbook content, increasing their ability to understand concepts, increasing engagement, and thereby, improving academic performance. ITBs offer students easy access to definitions and descriptions of concepts, make connections across different sections of the syllabus, and allow students to pose their own questions. Initial trials of ITBs that utilize KGs have been found to improve student grade outcomes by a full letter grade over the control group that was using a conventional textbook. ITBs have been found especially helpful for underperforming students, thus, broadening participation.

The potential of ITBs to facilitate better learning has been difficult to realize without major investments of time, money, and expertise. KGs are currently constructed using human subject-matter experts in a process that is extremely expensive and time consuming. We will create a scalable model of KG construction that has a faster turnaround and a fraction of the costs incurred by existing methods.

INTELLECTUAL MERIT

The Open Knowledge Network (OKN) movement has gained rapid momentum and aims to realize the value of open data to society. Aligned with this movement, we are proposing to launch the Textbook Open Knowledge Network (TOKN), a "vertical" OKN project in the domain of education whose long-term goal is to create a library of ITBs for all educational subject areas. The objectives of our Phase II project are to:

1. Complete the development of tools and processes that leverage machine learning and crowdsourcing for generating high-quality and extensible KGs for ITBs.
2. Evaluate the new tools and processes by developing KGs for a few chapters from Biology and Psychology undergraduate textbooks, and measure the time and effort required.
3. Incorporate the new KGs into an ITB, and evaluate its impact on student learning, engagement and academic outcomes.

BROADER IMPACTS

TOKN has the potential to significantly advance education worldwide. We will eventually integrate the ITB technology into the full OpenStax library of 41 textbooks, which has the potential to influence millions of students. OpenStax has further developed a robust ecosystem of 54 commercial partners (including every major publisher save one) who use OpenStax materials within their educational technologies and a further 30 major college and university systems who are using OpenStax to institutionalize open educational resources (OER) at their campuses. Our KG construction tools will be integrated with Stanford’s Protégé environment. Protégé is one of the most widely used knowledge authoring environments in the world today. In 2019 alone, Protégé was downloaded 145,000 times, and its web version hosts over 70,000 projects with over 50,000 user accounts. Once included in Protégé, our KG tools will benefit its large user community building knowledge graphs across a range of applications spanning biomedicine, engineering, and business.

1 INTRODUCTION

The United States faces a crisis in science, technology, engineering, and mathematics (STEM) education: K-12 students perform below their peers from other industrialized countries on international tests of STEM subjects, fewer students are enrolling in STEM disciplines, and a lower percentage is completing a degree, many graduates are ill-prepared for the workforce because they retain only a fraction of what they were taught, and faculty are pressured to teach an expanding number of students while also covering an increasing volume of material. Furthermore, there is much evidence to suggest that the current STEM education system is not working equally well for all Americans: participation of underrepresented groups is declining, and the retention rate of degree-seeking minority and women learners is only around 40% (Clough, 2005). Our nation is at risk of losing its competitive edge unless fundamental changes are made in the way that we educate the STEM leaders of tomorrow. In summary, revitalizing STEM education at all grade levels is a major priority in the United States (Augustine, 2005).

To properly educate our students, we must move beyond the centuries-old, ingrained paradigm of education that views the process of learning as a one-way street in which knowledge is transmitted from teacher to learner via paper textbooks and lectures. In addition, we need to provide learners with tools to effectively engage in self-regulated learning outside the classroom.

The ongoing *artificial intelligence* (AI) revolution is providing us with the opportunity to make learning a more active and generative process, and thereby fundamentally changing education to help avert this crisis. The machine-automated functionality provided by intelligent textbooks (ITBs) can generate summaries of textbook content, generate useful practice exercises for students, provide interactive dialogues with students. One study found that ITBs provided rich study enhancements for students, and improved student grade outcomes by a full letter grade over control conditions using conventional textbook formats (Chaudhri, et al., 2013).

At the heart of an ITB is a *knowledge graph* (KG) that captures the relationships among the concepts in the domain being taught. A KG is a knowledge representation that imbues ITBs with functionality such as concept summarization, question generation, and question answering. The KG will ultimately enable the development of the next generation of AI-augmented tools for ITBs.

The potential of ITBs to facilitate better learning has been extremely difficult to realize without major investments of time, money, and expertise. Indeed, it is the construction of the KG itself that presents the biggest barrier towards developing extensive libraries of ITBs. The reason is that KGs are currently constructed using human subject matter experts, and this process is extremely expensive and time consuming. As a concrete example, consider the *Inquire Biology* ITB that was created by senior investigator Chaudhri for the popular introductory textbook, Campbell's *Biology* (Reece, et al., 2015). The knowledge engineering for just the first 10 chapters required an effort of 5 person years from biology subject matter experts. Scaling this effort to all 56 chapters of the textbook would have cost over $1.5M. Similarly, while the AI-based *Cognitive Tutor Algebra* course has markedly increased student achievement and is now used in more than 1000 school districts by more than 500,000 students (Koedinger, Anderson, Hadley et al., 1997; Carnegie Learning, 2019), it required over a decade of heavily funded research to develop. These expensive and time-consuming processes must be repeated for each new textbook or subject area, rendering the entire concept unscalable. Compounding matters, due to the large investment required, it is too tempting for publishers and education technology providers to keep their KGs proprietary, eliminating any of their utility outside of the scope of the immediate project for which they were created.

1.1 Intellectual Merit

The opportunities and challenges involved in developing a scalable AI infrastructure for creating KGs needed to support ITBs and revolutionize education resonate strongly with the growing movement that aims to develop an Open Knowledge Network (OKN), which realizes the value of open data to society by providing semantic, machine-readable information regarding how different concepts relate to each other (NITRD, 2018). Construction of an OKN requires both "horizontal" efforts that develop infrastructure and tools that apply to all domains and "vertical" efforts that tackle challenges specific to topical domains.

We will launch the ***Textbook Open Knowledge Network* (TOKN), a vertical OKN project in the domain of education** whose long-term goal is to support a library of ITBs for all educational subject areas. Education is a rich domain that will push the broader OKN research community to consider more expressive KGs than traditional graph architectures like the (entity, relationship, entity) tuple can accommodate. Our objectives in the phase II of the project are as follows:

- **Objective 1: Create a scalable infrastructure for KG construction.** Our Phase I work initiated the development of tools for information extraction algorithms and *crowdsourcing* of KG-relevant knowledge from students. We will complete the development of these tools and extend them providing facilities for KG visualization and verification, and for active learning.
- **Objective 2: Create TOKN for Biology and Psychology:** We will evaluate our KG construction tools by creating KGs for a few chapters from OpenStax Biology (Clark, et. al., 2018) and OpenStax Psychology (Spielman et. al., 2020) textbooks. During this process, we will measure the time and effort required for creating the KGs and assess the coverage and accuracy of the resulting KGs.
- **Objective 3: Evaluate pedagogical effectiveness of TOKN:** We will evaluate the use of TOKN in its ability to improve student learning and engagement. OpenStax will incorporate KGs into new ITB tools and will evaluate their pedagogical effectiveness. Educational Testing Service (ETS) will evaluate the KGs for their ability to support question generation. Macmillan Learning will conduct user focus groups and a beta test of an ITB prototype.

Our objectives build on the preliminary work initiated during Phase I, will lead to several innovative KG construction tools, and launch TOKN, which will be a crucial resource for textbook publishing and educational technology industries.

1.2 Alignment with the Track Objectives and Convergence Research

In the spirit of the NSF C-Accel program, TOKN brings together a multidisciplinary team of researchers from multiple institutions as well as a partnership with the educational non-profit, OpenStax based at Rice University. Rice and Stanford have proven expertise in the areas of KG construction, educational product development, and evaluation.

Stanford PI-Genesereth, Prof. Heller and Dr. Chaudhri have actively collaborated since January 2016. Working together they raised funding for the Intelligent LIFE project from the Wallenberg Foundation in Sweden. Prof. Heller secured the research use license for the LIFE Biology textbook (Sadava et. al., 2016), recruited biology students at Stanford, promoted the intelligent textbook concept among his co-authors, and is engaged in the development of pedagogical materials for the new prototype. Dr. Chaudhri developed the Intelligent LIFE prototype and has investigated automated methods for KB construction.

Stanford co-PI Prof. Musen is the principal investigator of the Protégé project (Musen, 2015), the BioPortal project (Noy, et. al., 2009) and the Center for Expanded Data Annotation and Retrieval (CEDAR) (Musen, et. al., 2015). Protégé is an indispensable open-source resource for an enormous international community of scientists—supporting the development, maintenance, and use of ontologies and knowledge bases by investigators everywhere, and across all domains.

Stanford investigator Prof. Michael Bernstein's research applies a computational lens to helping groups achieve their collective goals. He designs, builds, and studies social computing systems at scales from teams to crowds. Prof Bernstein used crowdsourcing in creating two of the most impactful datasets in AI today: ImageNet Challenge and Visual Genome (Russakovsky et. al., 2015; Krishna, et. al. 2017).

Rice co-PI Baraniuk has significant experience developing machine learning (ML) algorithms for graph-based educational tasks, including learning content analytics (Lan, Waters, Studer et al., 2014), collaboration detection (Waters, Studer & Baraniuk, 2014), peer grading (Waters et al., 2015), and automatic question generation (Wang, Lan, Nie et al., 2018). OpenStax is part of Rice University and is a 501(c)(3) nonprofit charitable corporation. OpenStax is the publisher of 41 free, open-source textbooks used by 3.4M students in 61% of all higher education institutions in the United States. The OpenStax instructor and student community is crucial for the crowdsourcing aspects of this project. OpenStax has significant relationships with numerous course instructors and subject matter experts that will be used to help recruit participants for studies.

Educational Testing Service (ETS) is a leader in educational testing, develops over 60,000 test questions every year, and engages over 600 content specialists. ETS is engaged in assessment development work for K-12 education in multiple states.

The success of TOKN hinges on the multidisciplinary nature of the research partnership. Past efforts to produce KGs have been wildly expensive and impractical. Only by partnering machine learning scientists and cognitive scientists with educational providers, can we affordably produce sizable KGs.

1.3 Partnerships

In addition to the partnerships listed in the previous section, we have worked out collaborative relationships with Macmillan Learning and Microsoft Research. Macmillan will provide us a research use license for their textbooks, conduct several user groups to obtain feedback on the Intelligent LIFE ITB, and organize a beta test of Intelligent LIFE ITB. Microsoft Research will leverage the TOKN created as part of our project in conjunction with their Academic Graph (Wang, et. al., 2020) and facilitate the construction of textbook taxonomies using their existing machine learning infrastructure.

1.4 Sustainability Beyond Phase II

We will pursue five different approaches for sustainability of our work beyond phase II: nonprofit publishers, commercial publishers, open data initiatives, training industry, and related domains.

OpenStax is a representative of a non-profit publishers, and through their leadership in Phase I, and by being a central participant in the Phase II, they have become a central stakeholder in the success of the work proposed here. Our Phase II work will demonstrate the usefulness of ITB concepts for portions of the OpenStax Biology and Psychology textbooks. As a result, incorporating ITB concepts into the product versions of these books, and in the full library of 41 textbooks is on their future road map (see 6.2 Milestones Past Phase II). OpenStax believes that they can pursue this vision by leveraging their ecosystem of 58 educational technology partners who use their content.

Macmillan learning is a representative of a commercial publisher, and through their participation on the advisory board of our project, we will work with them to define a path to creating an ITB product offering. In addition to the LIFE textbook, Macmillan is interested in testing the ITB concept for their Biochemistry textbook (Tymoczko et. al., 2011). During Phase I, we have also initiated conversations to create an ITB for Neurobiology (Luo, 2015) with Taylor & Francis, and for a Bio Informatics textbook (Compeau & Pevzner 2018) with Active Learning publishers. Working with these publishers, we can ensure that our deliverables are responsive to the market needs.

Through a formal collaboration with Microsoft Academic Graph, and with the advisory board participation by Dr. Guha from Google's Data Commons, we will find a place for TOKN among the ongoing open data initiatives. With a well-defined contribution to open data efforts, TOKN will be eligible for support from different government and commercial sources.

Fourth, we will work with other participants in the convergence accelerator to investigate knowledge networks for domain such as training and manufacturing. Some candidates for such collaboration are Team B-6915 (creating training materials for work force development), Team A-7043 (creating a manufacturing network), and several others listed in the supplementary document on track integration.

Finally, we envision using the same tools for enabling formal representations in related domains such as income tax law, contract law, etc. Through Stanford's center for computational law, we will seek out applications that can benefit from semiautomated tools for KG construction.

1.5 Broader Impacts

TOKN has the potential to significantly advance the state of education worldwide.

First, the rich functionality afforded by ITBs have been shown to significantly improve student learning and performance (Chaudhri et al., 2013). We have seen ITBs to be especially useful tools for underperforming students, who may hesitate to ask questions in the classroom, and may be left behind. Multiple classroom experiments have confirmed that ITBs are just as easy to use as traditional electronic books, and yet lead to significant student learning benefits (Koc-Januchta et. al. 2020).

Secondly, the crowdsourcing activities that we propose not only will collect data for KG generation but also will benefit the students as pedagogical tools. Specifically, concept maps have been shown to be effective at promoting *relational processing* (Grimaldi et. al., 2015).

Thirdly, the ML and crowdsourcing activities that we propose will enable efficient and scalable KG generation throughout all subject domains in education, not just biology, and can easily be integrated into any courseware platform outside of the OpenStax ecosystem. Additionally, the tools and techniques that we will develop will be easily applicable to *any* natural language corpora. We envision viable applications of our proposed methodology to training manuals and content sourced from the Internet.

Finally, the KG construction tools will be made available for use on Protégé, a well-established means of creating and maintaining ontologies with an impressively large user base. Protégé (including its Web-based client, WebProtégé) is the most widely used tool in the world for creating ontologies and knowledge bases. Several members of the current convergence accelerator cohort are active users of Protégé. Releasing our KG construction tools will benefit not just our team's work, but over several hundred thousand users in the Protégé community.

2 BACKGROUND ON INTELLIGENT TEXTBOOKS AND KNOWLEDGE GRAPHS

We will begin by discussing our insights from user research, and then discuss a concrete ITB prototype, and how it helps students in learning better.

2.1 Challenges Faced by Students and Opportunities with Intelligent Textbooks

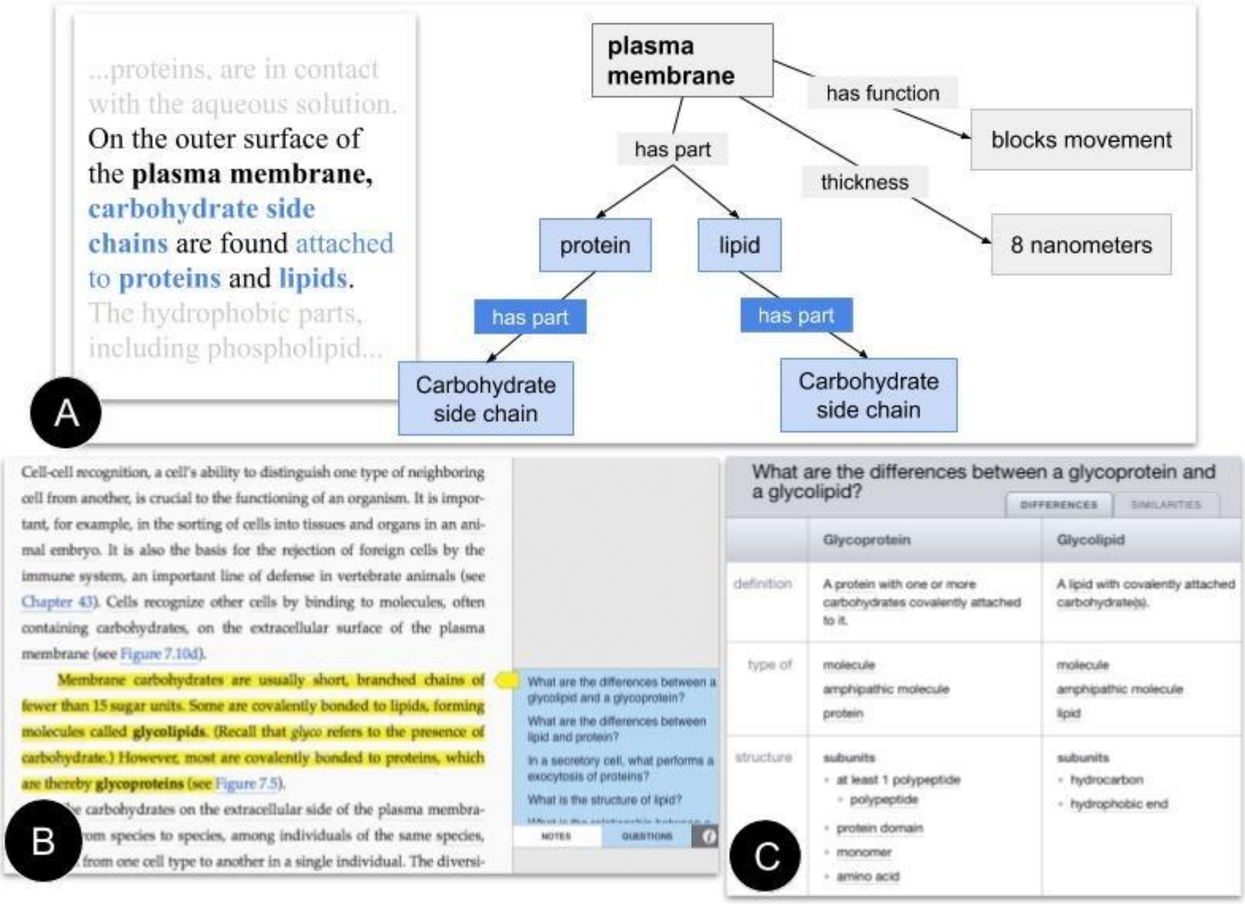
We conducted user research into college learning environments and associated study materials. Three major challenges faced by students in introductory biology courses are: (1) learning a massive new vocabulary and relationships connecting that vocabulary (2) learning four dimensional concepts (e.g., spatial functions through time) and (3) inter-relating different levels of biological organization. Unfortunately, many students struggle to overcome these challenges and traditional textbooks provide very few resources to assist students who struggle. When utilizing traditional textbooks, students often rely on ineffective study strategies such as repeated reading or underlining/highlighting (Miyatsu, Nguyen, & McDaniel, 2018). Our research revealed that successful learning environments contained an extensive *tacit* body of knowledge held by instructors and teaching assistants. This tacit knowledge can span many forms (see Phase I Portfolio, insights P2K1, P2K2, P3K1), but includes unmentioned relationships between ideas covered in a textbook as well as relationships across neighboring disciplines such as Biochemistry and Anatomy. We also found that teaching assistants place emphasis on assistive tools such as diagrams which allow students to interpret course content multi-modally, improving understanding and retention (P1K2).

ITBs can be equipped with powerful techniques drawn from AI, ML, as well as pedagogical techniques from learning science to directly help students overcome these challenges. For example, since information in an ITB is internally represented as a graph, the ITB designers can utilize this graph to represent information in various ways such as flashcards, related concepts, or simply by visualizing the graph itself. Additionally, our research into publishers and instructors informed us of the broad challenges of textbook content creation (e.g. quiz questions and evaluations). By leveraging a knowledge graph, these use cases can be more readily addressed with the graph's machine-interpretable "ground truth".

2.2 Case Study: A Knowledge Graph and Intelligent Textbook

To illustrate the promise and utility of ITBs, we will use the *Inquire Biology* ITB as a case study (Chaudhri et al., 2013). Central to all the functionality present in *Inquire* is the underlying KG, which represents all of the terms and relationships present in biology. The KG used in Inquire Biology expresses existential rules (Mugnier, 2012). The KG needed for *Inquire* was hand-created by subject matter experts in biology and included approximately 6,000 concepts and more than 100,000 relationships between these concepts. We provide a pictorial representation of the *Inquire* KG in Figure 1-A.

2.2.1 Application of a Knowledge Graph in an Intelligent Textbook



...proteins, are in contact with the aqueous solution. On the outer surface of the **plasma membrane, carbohydrate side chains** are found attached to **proteins** and **lipids**. The hydrophobic parts, including phospholipid...

Cell-cell recognition, a cell's ability to distinguish one type of neighboring cell from another, is crucial to the functioning of an organism. It is important, for example, in the sorting of cells into tissues and organs in an animal embryo. It is also the basis for the rejection of foreign cells by the immune system, an important line of defense in vertebrate animals (see Chapter 43). Cells recognize other cells by binding to molecules, often containing carbohydrates, on the extracellular surface of the plasma membrane (see Figure 7.10d).

Membrane carbohydrates are usually short, branched chains of fewer than 15 sugar units. Some are covalently bonded to lipids, forming molecules called **glycolipids**. (Recall that *glyco* refers to the presence of carbohydrate.) However, most are covalently bonded to proteins, which are thereby **glycoproteins** (see Figure 7.5).

e carbohydrates on the extracellular side of the plasma membra-
om species to species, among individuals of the same species,
from one cell type to another in a single individual. The diversi-

What are the differences between a glycolipid and a glycoprotein?

What are the differences between lipid and protein?

In a secretory cell, what performs a exocytosis of proteins?

What is the structure of lipid?

What are the differences between a glycoprotein and a glycolipid?

	Glycoprotein	Glycolipid
definition	A protein with one or more carbohydrates covalently attached to it.	A lipid with covalently attached carbohydrate(s).
type of	molecule amphipathic molecule protein	molecule amphipathic molecule lipid
structure	subunits • at least 1 polypeptide • polypeptide • protein domain • monomer • amino acid	subunits • hydrocarbon • hydrophobic end

Figure 1. (A) A graphical representation of a portion of the *Inquire* textbook and the underlying knowledge network that captures the relationships between terminology. The machine learning algorithms used by *Inquire* use this underlying knowledge graph to provide enhanced learning features for students. (B) An example of automatic question generation in the *Inquire* textbook. A student has highlighted a portion of text, and the textbook has provided a series of related practice questions. (C) An example of automatic question answering in the *Inquire* textbook. The student has input the question "What is the difference between a glycoprotein and a glycolipid" and the textbook has provided a table comparing and contrasting the two.

Inquire uses its KG to support several intelligent textbook features designed to aid students. One such feature provided by *Inquire* is automatic question generation. A large body of research has demonstrated that practicing recall of information is critical to the learning process (e.g., Karpicke & Grimaldi, 2012). However, research has also shown that students are reluctant to engage in retrieval practice on their own (Karpicke, 2009). Students using Inquire can highlight a portion of the text and then receive automatically generated practice questions related to that material. An illustration of the question generation feature is provided in Figure 1-B.

An additional feature of the *Inquire* textbook is the ability for students to engage in active dialog with the book. Students can submit questions to *Inquire* about the content and *Inquire* uses the underlying KG to provide accurate and appropriate answers to the student. Research has shown that asking questions is positive for learning (Chin & Osborne, 2008), but students are often reluctant to ask questions in class, often for fear of embarrassment (Watts & de Jesus, 2005). By offering the students the ability to ask questions in a closed setting and receive immediate feedback, students are more likely to obtain the benefits of question asking. An example of this functionality in *Inquire* is shown on Figure 1-C.

A third application of KGs is monitoring student progress in a domain. By decomposing all knowledge in a textbook into discrete chunks, it becomes much easier to identify gaps in student learning and target those gaps for remediation. This approach is commonly used in intelligent tutoring systems, using a process known as “knowledge tracing” (Koedinger et al., 1997). Intelligent systems using this approach produce positive results on student learning (Steenbergen-Hu & Cooper, 2014).

It is important to note that the features described above represent only a subset of the possible applications of KGs. The value of KGs lies in their power to drive advanced learning features in intelligent textbooks and learning systems. However, most developers and researchers do not have access to KGs due to their proprietary nature. We expect that by opening knowledge graphs to the broader educational research community, it will unlock a great deal of innovation in educational technology and intelligent learning systems. Thus, there is a critical need to produce *open* KGs for use by the community.

2.2.2 Pilot Studies with an Intelligent Textbook

Inquire was evaluated in a pilot experiment with community college students at DeAnza College in Cupertino, California. Study participants were divided at random into three experimental conditions: paper textbook (N=23), electronic textbook (N=25), and a textbook with knowledge representation and intelligent textbook features supported by *Inquire* (N=24). The students in each group engaged in active reading and problem solving (like what students would do during a homework assignment) on the topic of membrane structure and function. Later, the students were given closed-book quizzes to assess their learned knowledge. Students using *Inquire* scored a full letter grade (10%) higher than students in both control groups. This effect was statistically significant below the p=0.01 level.

2.3 The Challenge of Knowledge Graph Construction

The utility and impact of *Inquire Biology* is entirely dependent on the quality of the underlying KG. Notable in the development of the KG used by *Inquire* is the large cost and time involved. At the peak of the project, KG construction utilized over 15 subject matter experts. Further, we estimate that KG construction required roughly 6 person-months/chapter, with an investment of over $1.5M required to complete the entire textbook. The difficulty in making a single KG makes the scalability of this approach infeasible for a large library of textbooks. Worse, the KG is proprietary and is unable to be utilized by the broader education community. Thus, there is a critical need to lower the time and monetary investments necessary to create KGs and produce KGs that are open and freely available.

2.4 Overview of Approach and Research Strategy

An ITB relies on an explicit representation of knowledge in a textbook that matches human understanding and enables precise reasoning with it. The current narrative in AI is dominated by machine learning and natural language processing that achieve scale by sacrificing either accuracy or expressiveness. Such a compromise is acceptable in applications such as search, recommendation systems, machine translation, etc. In education, our domain models must be accurate and expressive. The textbooks need to be nearly 100% accurate. Any computer-based tools that will eventually be as good as human tutors have to use an explicit model of the knowledge of the domain. There does not exist any automated technique that can create an accurate KG with an expressiveness of existential rules that is used in Intelligent LIFE.

The above observations have led us to pose the questions: Could we leverage automation (ie, ML/NLP) in a way that its output corrected through human effort? Could we further save on the cost by leveraging lower cost crowd workers? We recognize that for a KG of accuracy and precision appropriate for textbooks, it will eventually have to be approved by an expert which raises the following question: Could the KG creation be eventually taken over by a community of volunteer educators?

Natural candidates for automation in KG construction are entity extraction and relation extraction which are heavily researched tasks in natural language processing. Typical use cases for these tasks are web corpora and a small number of entity classes such as people, places, companies, organization, etc. There exists some work on relation extraction from textbook (Berant, et. al., 2014; Ling et. al., 2013) but accuracy is nowhere close to 100%. Therefore, during Phase I, we proposed to leverage entity and relation extraction techniques to bootstrap the KGs for textbooks. In combination with the automation, we also proposed to investigate a way to leverage crowd work though a concept mapping task which can yield high quality relationships. Given some preliminary success in these tasks, during Phase II, we will be undertaking a task to create larger KGs that can be validated by domain experts.

We recognize that even with this approach, we will be sacrificing expressiveness as the resulting initial KGs will not be as expressive as existential rules, but our goal is to make no sacrifice on the accuracy. With a convergence of our work with the publishing industry and the student feedback we hope that any further expansion of the expressiveness can be driven by the user-community and the market forces.

3 PRELIMINARY WORK

Our goal for our Phase I work was to lay the groundwork for a set of algorithms and crowdsourcing tools to enable the scalable creation of KGs. We envisaged four major thrusts to meet this goal: (a) automated construction of KGs (b) human-driven construction of KGs (c) fusing human-generated and crowd-sourced KGs (d) planning to scale on the OpenStax platform. Before giving our progress on each of these four thrusts, we give an overview of our strategy, and why chose to pursue these specific tasks.

3.1 Automated Construction of KGs using Machine Learning and Natural Language Processing

Our approach to automated construction of KGs has two key components: term extraction and relation extraction (shown in the figure to the right). We developed deep learning models for these components by adapting BERT pre-trained language models (Devlin et. al., 2018). For term extraction, we use the existing glossaries at the end of each chapter found in multiple OpenStax textbooks to create training data. Once trained, we input new textbook sentences to the extractor, and it predicts the key terms found in each sentence. We train the relation extractor using an existing biology KB called KB Bio 101 that we manually built in a previous project. As the mappings between KB Bio 101 and the textbook sentences are not available, we created the desired training data using distant supervision (Zeng, et. al., 2015). Once trained, we fed a sentence that contains a term pair into the model, and the model predicts the likely relationship between the two terms. The term extractor currently has a precision of 0.73, recall of 0.51, and an F1 measure of 0.61. At the time of writing this proposal, we are still working to improve the accuracy of the relation extractor through weak supervision methods (Ratner, et. al., 2020).

3.2 Human-driven Construction of KGs using Crowdsourcing

The crowdsourcing task is to identify the relationship between two terms given a sentence with those two terms in it. For example, given a sentence: *A nucleus resides inside the cytoplasm*, and the terms “nucleus” and “cytoplasm”, the user must correctly identify that the terms are related by an “is inside” relationship.

Our first step in this process was to choose an appropriate set of relationships that we should ask the crowd workers to choose from. Our relationship choice is based on an upper ontology called Component Library or CLIB (Barker, et. al, 2001). As CLIB was used extensively for constructing KB Bio 101 (Chaudhri, et. al., 2014), we analyzed the most frequently used relationships. We selected the relations for describing the structure and function of entities, structure of processes, and causal relationships between processes. We were informed by the empirical experience of the effectiveness of these relationships in practice as well as more recent work on linguistic analysis of relations (Gisborn & Donaldson, 2019). For example, we replaced some of the confusing relationships from CLIB (such as agent, object, and base) with a general relationship participant. The relationships we currently support include taxonomic relationships for classes and instances; structural relations such has *has part; material,* spatial relationships such as *is inside, is above;* function relationships such as *has function, facilitates;* event structure relationships, such as *subevent, next event;* and causal relationships such as *enables, prevents.* We also allow the possibility that no direct relationship may exist.

In Phase I, we implemented a crowdsourcing tool to capture the selected relationships. The intended user of this tool is a biology student who does not have any formal training in knowledge engineering. During the development of the tool, we repeatedly validated our designs through rapid prototyping with such users. The user is first asked to read a section from the textbook and then undergo a short training on the relationships. We designed the training using simple common-sense examples that new users would find easy to understand. For example, we explain the "is inside" relationship using a visual in which a cat is shown hiding inside a box (Figure 2). We have developed similar illustrations for all the different relationships supported by the tool. After the training, the user completes a series of tasks through an interactive dialog to identify relationships between various textbook terms. All possible relationships between a pair of concepts can be extremely large. Some simple insights make our task tractable namely: most term pairs are not related (i.e., the final graph is sparse), that the terms that are connected are also likely to co-occur closely in the text, and that we can group the relationships into families so that the user first chooses a family of relationship before choosing the actual relationship.

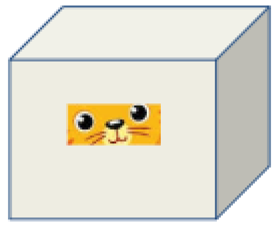
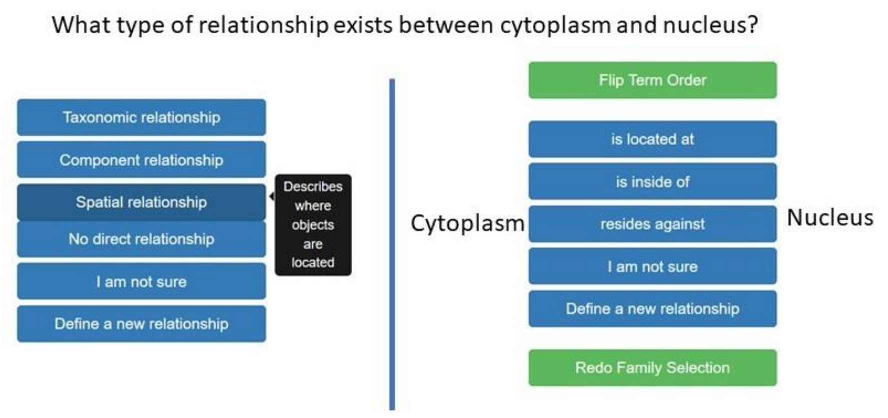


Figure 2: Illustration of *inside* relationship

As a concrete example, consider this two-step selection in the dialog shown in Figure 3 where the user is asked to relate the terms "cytoplasm" and "nucleus". The user first chooses the appropriate relationship family for the terms, including taxonomic, spatial, and component-based relationships. The user further can select that the terms have no relationship between them, that they are unsure of the relationship, or that they would like to define a new relationship to relate the terms via a textbox. In this example, the correct relationships family is a spatial relationship and clicking on this option takes them to a second set of options to specify which spatial family relationship is correct. In this dialog, they have an option to flip the order of the terms to ensure that the chosen relationship applies in the correct direction. Once they flip the order of terms, they can correctly indicate that the nucleus *is inside* cytoplasm.



What type of relationship exists between cytoplasm and nucleus?

Figure 3: The crowd worker first selects the correct family of relationship, followed by the actual relationship

3.3 Fusing Human-Generated and Crowd-sourced KGs

To test the term extraction and crowdsourcing methods in conjunction, we selected Section 10.2 from the OpenStax Biology 2e textbook. (We anticipate completing additional experiments for Section 4.2, and a section derived from a Psychology textbook before the end of Phase I.) We used automated term extraction to identify all the terms in this section. As the precision and recall of the automated method is not perfect, the biologists on the team validated the terms. We then parsed the section into individual sentences and automatically identified all term pairs that existed in each sentence. A sentence that contained N terms would have $\binom{n}{2}$ possible pairings, with each pairing considered to be a single task. Using the tool described above, the tasks were presented to participants on a crowdsourcing platform. By the time of the writing of this proposal, we have collected over 8000 relationship labels across 478 tasks from 100 participants.

Workers have varying degrees of competency. Similarly, we expect tasks to have varying degrees of difficulty. Because of this, as well as having a relatively small amount of labeling data per task, taking a simple majority vote to determine the correct task label can lead to significant estimation errors.

To overcome these issues, we implemented a sophisticated denoising algorithm, which we now describe. In our experiment, we have *N* users who are jointly labeling *T* tasks. We note that each user will, in general, only label a subset of all possible tasks. In our experiment, the average user labeled around 80 of the 470 possible tasks (roughly 17%). We assume a priori that each user *i* has a latent raw ability $\theta_{i\,,}$and that each task *j* has a latent difficulty μ_j. Further, each task is assumed to have a correct label c_i. We use the variable W_{ij} to denote the event that user i knows the correct label for task j and we model this scenario probabilistically via the relationship $P(W_{ij} = 1) = \Phi(\theta_i\text{-}\mu_j)$, where $\Phi(x)$ is the inverse probit link function (normal cumulative distribution function). In words, the user is likely to know the correct label if their latent ability is larger than the difficulty of the labeling task. Because the user is selecting relationships from a discrete set, we further assume the possibility that a user who does not know the correct label can still guess the correct label with a task-specific parameter γ_j. In the end, the probability that the a user will correctly label the given task is given by $\gamma_j + (1\text{-}\gamma_j)*\Phi(\theta_i\text{-}\mu_j)$. We fit our model to data using Markov Chain Monte Carlo (MCMC) using a Gibbs sampler. This method is efficient and enables us to compute posterior distributions for all parameters of interest. To simplify the sampling of user ability and task ability, we augment our chain with a latent variable $Z = \theta_i\text{-}\mu_j$. After fitting our model to data, we obtain estimates of user ability, task difficulty and, most importantly, the correct label for each task considering all other information in the model.

We found through our denoising analysis that the average task difficulty was slightly higher than the average worker ability -- implying that we do need a reasonable number of labels per task in order to denoise correctly. We further found that majority voting did agree with the output of the denoising algorithm for easy to moderate tasks, but that the two have a higher chance of disagreeing on harder tasks. Concretely, for the 478 tasks the two methods agreed 453 times (95%) and disagreed 25 times (5%). The 5% of tasks where there was disagreement correlated with the upper 10% most difficult tasks. Our SMEs examined the tasks where there was disagreement between the majority vote and the denoiser and found that the denoiser produced a more reliable label in over 80% of the cases.

As a final measure of the success of our experiment, we randomly selected 50 tasks with their labels from the denoised output and presented these to two subject matter experts to verify if the labels that we obtained through our experiment were, in fact, correct. For each task we only assume that our method produced the correct label only if both SMEs agreed with the label. Doing this produced an overall accuracy of 86%. If we relax our scoring criteria to only requiring one of the two SMEs to agree with our label, our accuracy improves to 96%. These results show that our methodology can be used to produce accurate relationship data between term pairs and that this task can be carried out by non-specialists.

3.4 Planning to Scale on OpenStax Platform and Beyond

We have completed several tasks necessary to scale this effort on the OpenStax platform. First, we now have tools in place to process OpenStax textbooks that can be fed into the automated KG construction algorithms. Second, we have a backend database schema to store all the crowdsourced data. Our plan is to incorporate the crowdsourcing tools as part of OpenStax Research Labs tools so that the students studying from the Biology 2e textbook could choose to participate in the crowdsourcing task. Finally, we have established a collaboration with Microsoft Research who will process OpenStax textbooks to help bootstrap the concept taxonomy for our project. They will leverage and disseminate any improvements we make to this taxonomy which will be integrated into the Microsoft Academic Graph.

4 PROPOSED RESEARCH

The objectives of our Phase II work are to complete the development of KG construction tools initiated during Phase I, experimentally validate them by creating KGs, use the developed KGs in ITBs, and undertake preliminary evaluation of their pedagogical effectiveness.

4.1 KG Construction Infrastructure

We will consider our plans for completing the tools initiated during Phase I. In addition, we will build tools to verify, visualize, and reason with the resulting graph.

4.1.1 Term and Relation Extractor

Recall from Section 3.1 that the term extractor exhibited a precision of 0.71, recall of 0.51, and an F_1 measure of 0.61. The current performance provides a good starting point to bootstrap the terms to be represented in a KG. We will make two improvements. First, as we use the term extractor on new sections of the textbooks, and validate its output, it will generate additional training data. We will incorporate this training data to improve the performance of the term extractor. Second, the validation of the extracted terms was performed by the biologists on the team. We will make such validation more scalable by an initial verification through crowdsourcing (Deng et. al., 2015). The final verification will be done by an expert biologist. While making these improvements, we will leverage the experience gained in creating Visual Genome dataset, which is a large crowdsourced dataset for computer vision (Krishna, et. al., 2017)

In addition to extraction of raw terms, we will extend our extractor to deal with linguistic knowledge. For example, to recognize singular vs plural forms, to recognize abbreviations, to recognize morphological variations, etc. For our Phase I work, we relied on the lexicon built into the spaCy library, but we have found its coverage to be lacking for specific domains such as Biology.

As reported in Section 3.1, our automated relation extractor is still under development. During our Phase II work, we will investigate two new approaches: weak supervision, and leveraging the data produced by the crowdsourcing app. We will leverage the weak supervision framework provided in Snorkel, which is a tool developed by colleagues at Stanford (Ratner, et. al. 2020). We will write labeling functions that generate noisy labels from weak supervision sources such as lexical pattern heuristics. Snorkel synthesizes and combines these labels into probabilistic labels to train a machine learning model. To get maximum boost in performance through such training, we will leverage prior work on gated instruction which combines an interactive tutorial to correct errors during training (Liu, et. al. 2016; Zhou, et. al. 2019).

4.1.2 Relation Extraction through Crowdsourcing

OpenStax will continue the development of the concept map generation tool from Phase I so that it can be used across new types of knowledge. Recall from Section 3.2 that the first step in the process is to choose a suitable set of relations. We will extend the app to capture information about process regulation (Chaudhri, et. al. 2014b). Once the relations are chosen, the training must be extended so that the crowd workers can be trained on it before they perform the task. We will use the feedback from users to improve its design.

4.1.3 Fusing the Data from Crowdsourcing and Automated Extraction

Our Phase II research for our crowdsourcing task will leverage information derived from the denoising model to help effectively adapt the presentation of tasks to individual users. Recall from Section 3.3, we used our denoising model to accurately estimate the correct label for each task given the relationship data. While this alone is a significant advantage of using the denoiser over simple majority-voting schemes, another advantage is that it enables us to recognize task labels for which we have high confidence as well as tasks for which we have less confidence. We will use this information to *adapt* presentation of tasks to of the labeling app (Weld et. al., 2015). If we identify a task whose label is highly uncertain, we can assign high-performing users who can provide more useful data to help us resolve uncertainty. We can provide tasks to users with an appropriate level of difficulty to avoid overwhelming users with tasks too difficult for them. By adjusting the difficulties of tasks to users we also envision being able to train our users to improve their raw labeling ability over time.

4.1.4 Tools for Visualizing, Reasoning, Querying and Verifying Knowledge Graphs

Using the concept mapping app described in Section 3.2, the crowd workers provide us with relations between a pair of terms. To combine these relationships to form a knowledge graph, we must import them into a knowledge representation system. We will use the Protégé system and its associated tools.

The diagram in Figure 4 depicts a subset of denoised relationships that were collected during Phase I experiment reported in Section 3.2. We generated the diagram by mapping the extracted relationships into an OWL ontology through the following transformations: each *type of* relationship introduces a *subclass of* relationship, each *instance of* relation introduces an *assertion*, every other relationship introduces a *subclass of* relationship with an existential restriction. For example, the input data contains *is inside of* relationship between a *vesicle* and a *cell*, and a *pole* and a *cell*. These relationships created subclasses of *vesicle* and *cell* with an existential restriction that instances of them were *in side of* a cell. The visualization tool in WebProtégé stitched these assertions and many others to produce the diagram. When the relations are imported to form a connected graph, the presentation creates several opportunities and challenges. As an opportunity, the connected graph gives an overview of what we know about each concept, thus, making it possible for an expert to validate and verify it. As a challenge, the overall graph may contain erroneous connections, and it can get too complex and unwieldy to manipulate. A validated and completed graph can serve as a powerful teaching tool for Teaching Assistants who want to present an overview of a content area. We will extend this visualization facility in WebProtégé in the following ways: (a) implement a translator to import the denoised data from the crowdsourcing app, (b) provide visualization controls so that a domain expert can easily navigate the graph, (c) implement tools for checking incompleteness, inconsistency, and under specification in the graph, and provide feedback to experts for improving it, (d) provide a way so that a community of experts can comment on the graph (Kulkarni et. al., 2015; Kotturi et. al., 2015).

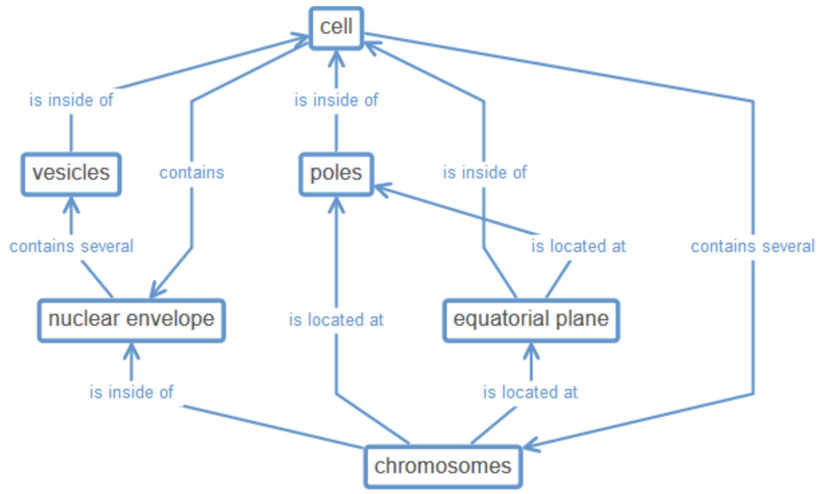


Figure 4: Visualization of denoised crowd data in Web Protégé

4.1.6 Active Learning for End-to-End KG Construction

Individual tools need to form a pipeline that can be used as a single unit. As noted in previous sections, even though we bootstrapped our effort using textbook glossary and a manually built KB, we can do better as we validate the results, and use those results to train our automated methods.

To leverage the human validation effort to iteratively improve the automated term and relation extractors, we will incorporate our tools into an active learning framework (Settles 2009). Active learning interleaves machine inference and human review into a single process. Human reviewers are provided tasks for which the automated process displays low confidence. Human-supplied labels are then fed back as training data for the model, and the process repeats. This approach has the benefit of improving the model's decisions while providing humans with potentially more challenging and interesting tasks.

Incorporating our tools into one active learning pipeline has two clear benefits. First, once sufficient training data has been built up, future textbooks will require significantly less human labeling overall. Therefore, this initial set of textbooks in TOKN will 'seed' the dataset for future endeavors. Second, individual tools will be integrated into a single unit which can be used outside our project.

4.2 TOKN Construction

Our goal in this task is to test whether the new tools developed here are indeed cost effective and general. We will perform this test by creating sample KGs for a Biology and a Psychology textbook. Our approach will be to process the textbook one section at a time. We will use the experience of creating a KG for each section as an opportunity to further test and refine our KG construction tools. We envision that, for our initial KG construction, we will use the workers from a crowdsourcing platform such as Prolific.

4.2.1 Scope of Knowledge Graph Construction

For the OpenStax Biology 2e textbook, we will create KGs for Chapter 4 and 10. Recall that our Phase I work has been focusing on Sections 4.2 and 10.2. For the OpenStax Psychology textbook, we will create KGs for Chapter 14 on Stress, Lifestyle, and Health. We have carefully chosen the chapter so that the content has some overlap with the content in Biology, yet it is different in interesting ways. For example, Section 14.4 is on regulation of stress. The concept of regulation is frequently seen in Biology, but our current concept mapping app does not yet include the ontology to capture knowledge about regulation.

4.2.2 Glossary Construction

We will process the textbook one chapter at a time through our relation extractor. We will use a combination of crowd workers and our team's expert biologists to validate the glossary terms. A well curated list of glossary terms is extremely important. Through the validation process, we will develop guidelines for glossary term selection. Some example guidelines are as follows: always use the singular form of a term, a modifier should not be a term on its own, synonyms should not be defined as separate terms, etc. As an example, our term extractor identified "eukaryotic" as a term, but it is a modifier, for example, "eukaryotic cell". In such cases, even though "eukaryotic" is a valid Biology term, but for KG construction, the use of "eukaryotic cell" is more appropriate. Such subtleties are difficult to capture using automation, and it is only an expert biologist with some training and background can help curate high quality terms.

4.2.3 Taxonomy Construction

We have formed a new partnership with the research team at Microsoft Academic Graph (MAG) (Wang et. al, 2020). The MAG team has created technology to process text to derive a taxonomy. The taxonomy created by MAG is, however, not a true ontology in that it does not satisfy transitivity properties. We will investigate if MAG tools could be used to bootstrap the taxonomy for our project. We will import the taxonomy and improve it through our KG construction process. We will develop an approach through which the improved taxonomy can be exported back to MAG, thus, enhancing their coverage of textbooks. We will begin this exploration by starting from the OpenStax Biology and Psychology textbooks.

4.2.4 Concept Map Data Collection

We will conduct the concept map data collection over crowd sourcing platform called Prolific. Participants on this platform can be easily screened for factors such as languages in which they are fluent, country of origin, and relevant subject matter expertise. During Phase I, only fluent English speakers from the United States, United Kingdom, and Ireland with subject matter expertise in Biology, Biological sciences, and Biochemistry were included. During Phase II, we will re-evaluate and further fine-tune our inclusion criteria to ensure that our participants are adequately representative of our target population, are completing the task satisfactorily, and not resulting in poor data quality (e.g. making only a single response or responding with "I don't know" across all tasks). We will process the concept map data using the denoising algorithm and import them into Protégé. We will recruit expert biologists to review and evaluate the resulting graph.

4.2.5 Evaluation of the KG Infrastructure

We will collect data on the time and effort required to create the KGs for each chapter. We will characterize the time and cost for each stage of KG creation: crowdsourcing of data, its processing by the time, its final review and verification. During the validation and verification of the resulting KG, we will measure its accuracy, and coverage. Accuracy is straightforward to measure as the experts can check if an edge is correct. Measuring coverage requires checking if some of the necessary edges are missing. Some of the missing edges can also be because they are not expressed in the textbook. We will design metrics to characterize these situations to provide insight into the overall quality of the KG.

4.3 Evaluate Pedagogical Effectiveness of TOKN

We will undertake three different evaluations for the pedagogical effectiveness of TOKN. First evaluation to be undertaken by OpenStax will explore ITB tools that will leverage the new KGs developed during the project. Second evaluation to be undertaken by ETS will use the KG developed during the project for automatic item generation. Finally, Macmillan Learning will undertake an evaluation of the existing ITB prototype. Each path to evaluation will offer us different insights and will also maximize the chances of success for our Phase II work.

4.3.1 Evaluation of TOKN derived study-features

Phase I user interviews revealed that KGs are valued across the publishing industry for their potential in developing product enhancing features such as adaptive tutoring, customized lesson plans, and assessment items. Indeed, as an education tool developer, OpenStax recognizes the value of KGs in affording student-facing study features such as concept summarization, question generation, question answering, and even concept mapping - the core data collection task at the start of the KG construction pipeline. Our evaluations of KG-afforded study features will focus on two distinct questions: Have students or faculty established a desire to use such a study feature? Is the study feature pedagogically beneficial to students?

To answer question 1, OpenStax will use low fidelity prototypes to assess the demand for study features derived from the KGs produced in this project. The development teams at OpenStax are seasoned users of UserTesting.com, a service that allows product developers and researchers to target specific user populations (e.g. college-aged students studying biology) and gather feedback on new features and ideas from low-fidelity wireframes all the way through live, fully developed websites. OpenStax will develop the following low-fidelity prototypes: visualizing concept maps at different levels of abstraction as a pedagogical tool to improve student understanding of relationships, generating assessment questions at various levels of difficulty and abstraction, and generating personalized flashcards based on student highlights and annotations. We will pick one “winning” idea for which we will undertake high-fidelity prototyping, validation, and evaluation. Let us refer to the winning idea as an *ITB Tool.*

To answer question 2, we will use OpenStax Labs, an experimental R&D platform to conduct randomized control trials (RCT) with and without a study feature to evaluate its pedagogical efficacy. OpenStax will use its close ties with instructors and teachers across the United States to advertise the opportunity and to directly recruit participants for this project.

Let us consider a sample study design. We will invite Biology 2e students to interact with two qualitatively similar sections (in terms of difficulty, readability, etc.). Students will be randomly assigned to either group A or group B. Students in group A will study the first section using the ITB Tool, and the second section without it. The tasks for students in group B will be flipped. Students in both groups will take the same quiz as critical assessment at the end of each section designed to evaluate both recall of facts, as well as the ability to understand and analyze the concepts. We will compare the student performance across these 4 conditions to make meaningful inferences about the learning effects of the ITB Tool.

4.3.2 TOKN Evaluation for Automatic Item Generation

Automated item generation (AIG) began to emerge in the last quarter of the 20th century (Bejar, 1985) and was deployed by the end of the century (Bejar, 2002). AIG is important to testing companies, like ETS, because it is a mechanism to digitize further and automate the content creation process while maintaining traditionally high-quality standards. AIG also opens the door to the forms of assessment likely to be needed in the 21st century. Notably, AIG enables enhancing intelligent textbooks to include self-monitoring of student learning (Madnani et al., 2019). The feasibility and effectiveness of AIG have been most clearly demonstrated in the assessment of quantitative skills from K-12 to the sort of reasoning measured by the Graduate Record Examination – GRE® (Bejar et al., 2003; Newstead, Bradon, Handley, Dennis, & Evans, 2006). The application of AIG has been especially successful in the K-12 context (Bejar et al., 2018) where the human-in-the-loop-AIG increased efficiency by orders of magnitude compared to traditional methods of test development. Other testing organizations are also implementing AIG, such as the Medical Council of Canada (Pugh, De Champlain, Gierl, Lai, & Touchie, 2016) and the international publisher Elsevier (Cole, Lima-Walton, Brunnert, Vesey, & Raha, 2019).

In short, AIG has the potential to reduce the costs of item development as well as to speed up item and test development. However, the development of AIG for science subjects presents unique requirements. To implement science AIG, a KG is needed, and no suitable KGs are readily available. We will evaluate the KG developed in the project for AIG in the following stages.

Requirement Specification. To help achieve those potential benefits, KG needs to fulfill certain requirements. First, it needs to be accurate. ETS always stresses the focus on quality of items and tests, and AIG is no exception. For practical educational purposes, AIG also needs to be highly configurable – test developers using AIG need to be able to specify a variety of parameters. Flexibility requirements include such aspects as what content domain is required, what coverage is needed, what specific topics need to be addressed, what kinds of questions/items need to be generated, what level of inquiry is requested (e.g. high-school vs college, etc.), what proportion of different items is needed, levels of item difficulty. An additional requirement for AIG is to be able to generate both open-ended questions and multiple-choice items. The latter type is still much more popular and are easier to score, but they require additional precision in crafting the distracter options.

AIG Algorithm Development. ETS has developed a prototype AIG system that generates multiple-choice items for the domains of Biology and Zoology. The system uses a small-scale KG that was curated by subject-matter experts. The KG is stored in RDF-like form which facilitates fast generation of many different items. Assertions from the KG are used for generation of both the stem (question), the key and the distracter options. However, scaling the KG is a serious challenge, and this is where ETS expects to benefit greatly from the TOKN. By adapting our AIG system to use the large-scale TOKN based on OpenStax textbooks will enable leveraging the current prototype system to potentially operational level.

AIG evaluation. Given the timeline of the project, it is not feasible to field test generated questions from the knowledge graph. Instead, we will evaluate the items through subject matter specialists at ETS that develop science items for the multiple science tests produced at ETS. The items will be evaluated in terms of acceptability (is the item acceptable from a grammatical and semantic perspective?); difficulty (is the item in a range of difficulty appropriate to the intended population of students?); discrimination (is the item likely to make it possible to distinguish students that truly know the correct answer?). Three ETS science content specialists will rate the generated questions as Yes/No/Maybe in each dimension. A specialist will rate only one of the dimensions at a time to prevent as much as possible rater biases. The data will be analyzed to determine the level of interrater agreement and the percentage of items that meet all criteria, two criteria, and one criterion. Based on this analysis, the AIG system will be revised to maximize the percentage of acceptable items. The content specialists will be chosen to represent a variety of ethnic and cultural backgrounds.

4.3.3 TOKN Evaluation in the Wild

We have established a collaboration with Macmillan Learning to get direct feedback on ITB concepts from students and teachers using their books. Our advisory board member, Ms. Lisa Lockwood from Macmillan Learning will organize two different activities: user focus groups and a beta test.

The goal of the user focus groups is to get direct feedback on the intelligent textbook concept from students and teachers. The Cell Biology sections of the LIFE textbook supported in the current ITB prototype are used at UC Santa Cruz and UC Davis campuses which are in the same geographical area as Stanford. Ms. Lockwood will organize the focus groups and will be responsible for gathering and analyzing the feedback. Stanford's role will be to support Macmillan by providing necessary information and software that they may require for this purpose. We will conduct the user focus groups during the first year (2020-2021) of the project.

The goal of the beta test is to try out the ITB prototype with a small group of students in a Biology course using the LIFE textbook. We will collaborate with the Macmillan Learning Science team to design the experiment, formulate desired hypotheses, provide them software and support their data collection efforts.

Macmillan Learning Science team will analyze the data and derive the conclusions. We will conduct the beta test during the second year (2021-2022) of the project.

5 CONTRIBUTION TO TRACK SUCCESS

We will contribute to track success by: (a) Disseminating the results through Journal special issues (b) Organizing an integrative workshop on Textbooks of the Future (c) Offering Textbook KGs as a service to other participants (d) Leveraging public data sets (e) Providing KG platform (f) Providing KG education. More details are available in the supplementary documents.

6 TIMELINE AND KEY DELIVERABLES

6.1 Phase II Deliverables

Key Deliverable 1 (month 0): Improved versions of Phase I tools for term extraction, relation extraction, denoising crowd data, KG construction and concept mapping.
Key Deliverable 2 (month 6): We will release a tool for visualization and verification of the knowledge graph. This will be a new capability, and it will be released as part of the Protégé environment.
Key Deliverable 3 (month 9): We will release an active learning tool so that the new training data is used to improve the term and relation extraction tools.
Key Deliverable 4 (month 15): Knowledge graphs for two chapters from the OpenStax Biology and one chapter from the OpenStax Psychology textbooks. Report on time and effort required to create the KG.
Key Deliverable 5 (month 18): We will release new ITB tools that will result from the exploration of KG supported study tools.
Key Deliverable 6 (month 24): Evaluation report from study tools experiment, and automatic item generation evaluation.

6.2 Milestones Past Phase II

Milestone (Phase II + 24 months): Release of Textbook scale TOKN for both Biology and Psychology. Integration of ITB concepts into selected portions of OpenStax Biology and Psychology textbooks.
Milestone (Phase II + 60 months): Expanded coverage of TOKN to include multiple textbooks. Integration of ITB concepts into full length OpenStax Biology and Psychology textbooks.
Milestone (Phase II + 120 months): TOKN with a coverage of all textbooks in the OpenStax textbook library. Integration of ITB concepts into the full OpenStax library.

7. RESULTS FROM PRIOR NSF SUPPORT

Accelerating STEM Learning Through Large-Scale Data Science, Baraniuk (PI), Grant #IUSE-1842378, $5.2M, 10/19-09/22. Intellectual Merit: Carefully designed, large-scale learning studies involving large numbers of students and teachers coupled with modern data science tools can provide insights into learning factors and their interactions with learning environments unprecedented clarity and precision. Broader Impact: Armed with a new understanding of human learning and the new learning tools, educators and institutions can respond to these inequities with new programs and pedagogies that act as a great equalizer by providing each student the knowledge, skills, and support they need to succeed in today's rapidly changing world and tomorrow's workforce. This award has not yet produced any publications.

RAISE: C-Accel Pilot - Track A1: Open Knowledge Network: An Infrastructure for Intelligent Textbooks, Baraniuk, Genesereth M. (co-PIs) NSF 1937134, $1,000,000, 2019-20. Intellectual Merit: Enabling intelligent textbooks through the development of cheap and scalable knowledge graphs created through crowd sourced data. Broader Impacts: We will ultimately release developed algorithms and knowledge graphs to the public for use in AI-based applications. Through this project, we have created tools for automated term extraction, and crowdsourced relation acquisition that can be used for creating a knowledge network for textbooks. At the time of this writing we are exploring how these networks can be used in educational applications.

CACCEL PHASE II - Track A1 (Open Knowledge Network):
A7134: Textbook Open Knowledge Network
Stanford University
NSF Convergence Accelerator Pilot (NSF 20-055)

Phase I Portfolio

Our technical goal for the Phase I work was to lay the groundwork for a set of algorithms and crowdsourcing tools to enable the scalable creation of KGs useful for creating Intelligent Textbooks (ITB). We begin with an overview of our participation in the Phase I meetings and new opportunities created as a result. We undertook 12 user research interviews and summarize our results. Finally, we give an overview of the creative products from our phase I work.

Participation in Phase I Meetings

Our team was represented at all the Phase I meetings and webinars. Dr. Vinay K. Chaudhri attended all the face-to-face meetings and webinars except for the December meeting which he had to miss because of a family obligation. Attendance at these meetings create several new opportunities summarized below.

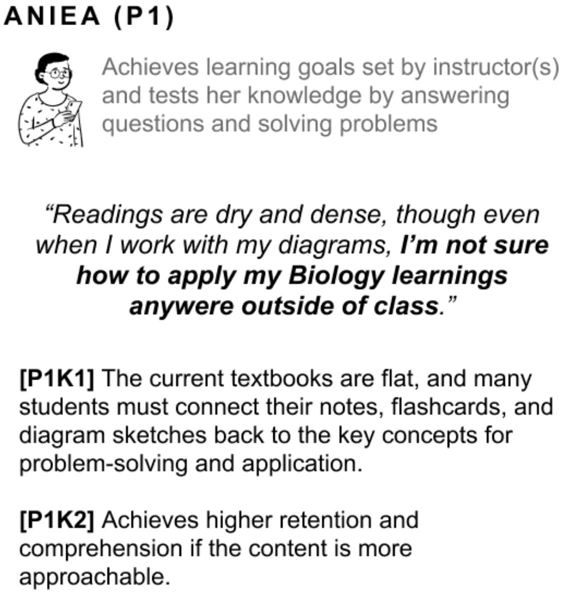
- At the February meeting, we met with Mr. Tyler S. Benster and held a brainstorming meeting with him on the Stanford campus. Mr. Benster proposed applying the ITB concept to a Neurobiology textbook authored by Prof. Linqun Luo. Prof. Luo examined the ITB concept, and immediately expressed his consent to see it applied to his textbook. He introduced us to his publisher Taylor & Francis who agreed to grant us a license to his textbook so that we could demonstrate the ITB concept for it.
- Dr. Chaitanya K. Baru introduced us to Pevel Pevzner who has been investigating the concept of Massively Adaptive and Interactive Textbook (MAIT). The concept of MAIT is orthogonal to ITBs, but synergistic in the sense that it recognizes the investment of human effort necessary to create good textbooks. In our follow up conversations with Prof. Pevzner, he expressed an interest to see the ITB concept applied to his textbook. We viewed this as an interesting new opportunity for leveraging ITBs in the context of a massive open online course (MOOC). Prof. Pevzner agreed to collaborate with us to explore this idea further.
- Dr. Kuansan Wang and Dr. RV Guha, who were the invitees at the December meeting, agreed to participate on our advisory board. With Dr. Wang's team at Microsoft Academic Graph, we plan to investigate techniques for automated taxonomy construction for OpenStax textbooks, and potentially create a new application for the KGs that our team will create. With Dr. Guha we hope to be able to identify new partners who could benefit from and contribute to our work.

Through our participation in the meetings, and in the process of discussing our research, we have formed the following new collaborations. (1) In collaboration with Prof. Beverly Woolf, we will investigate the use of knowledge graphs for training manuals with the goal of supporting an intelligent tutoring system. (2) In collaboration with Prof. Binil Starly, we will investigate the use of knowledge graphs for a manufacturing textbook in support of a query answering system they are building. (3) We will leverage the expertise of Prof. Caferella as a consultant to our team to help advise on entity extraction and relation extraction techniques. (4) We will assist Prof. Nora El-Gohari in leveraging our new KG construction tools.

User Research

During Phase I, we conducted 12 interviews, spanning skillsets, job descriptions, and use cases for Intelligent Textbooks. We have synthesized key user needs in six distinct personas: college student, instructor, teaching assistant, content developer, publisher, and author. Individually, each persona consists of motivation, needs described as key insights (indexed as P1K1, etc.), areas of opportunities for Phase II (indexed as P1O1, etc.). To illustrate our findings, we discuss four of these personas, and organize them into two categories of ITB use: ITBs in an academic setting (ITB content consumer) versus a publishing setting (ITB content creators).

Academia - ITB Content Consumers



ANIEA (P1)

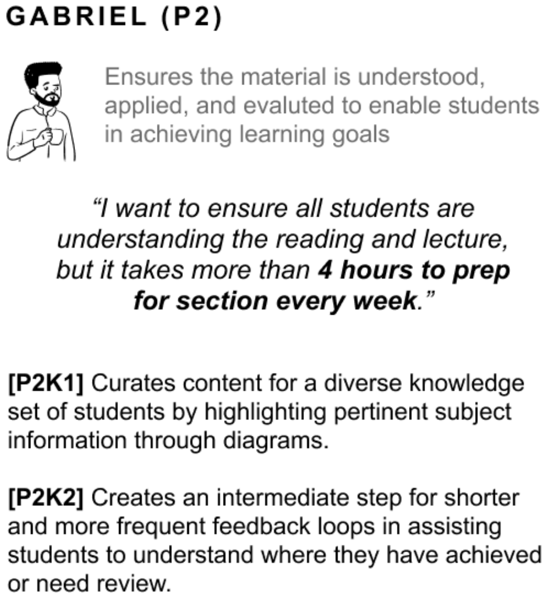
Achieves learning goals set by instructor(s) and tests her knowledge by answering questions and solving problems

*"Readings are dry and dense, though even when I work with my diagrams, **I'm not sure how to apply my Biology learnings anywere outside of class**."*

[P1K1] The current textbooks are flat, and many students must connect their notes, flashcards, and diagram sketches back to the key concepts for problem-solving and application.

[P1K2] Achieves higher retention and comprehension if the content is more approachable.

College student persona and key insights



GABRIEL (P2)

Ensures the material is understood, applied, and evaluted to enable students in achieving learning goals

*"I want to ensure all students are understanding the reading and lecture, but it takes more than **4 hours to prep for section every week**."*

[P2K1] Curates content for a diverse knowledge set of students by highlighting pertinent subject information through diagrams.

[P2K2] Creates an intermediate step for shorter and more frequent feedback loops in assisting students to understand where they have achieved or need review.

Teaching assistant persona and key insights

ITBs are tools for students in strengthening their STEM knowledge and can also benefit teaching teams supporting achievements in their learning goals. A college student, like Aniea, not only wants to do well, but wants to understand how and why she is doing well through questions and problem sets (Persona 1), yet current textbooks are flat, and disconnected from her study materials, like notes, flashcards, and diagram sketches [P1K1]. This is where Gabriel, her TA, enables Aniea's learning as it is imperative that the material is understood, applied, and evaluated to ensure students are achieving the learning goals of the course (Persona 2). Gabriel and his colleagues create short and more frequent feedback sessions for Aniea and her classmates. However, the time it takes to understand where each student's level of understanding of the material, while also creating short review labs for all, present time-intensive work. Through our Phase II work, we will continue to obtain better understanding of the intricacies of ITB for an academic setting and will undertake further rapid prototyping and iterative testing with students and teaching assistants. But how could ITB start to address these intricacies by design? This is where publishing tools provided through ITB to our other two personas are key in creating a holistic, engaging experience for all.

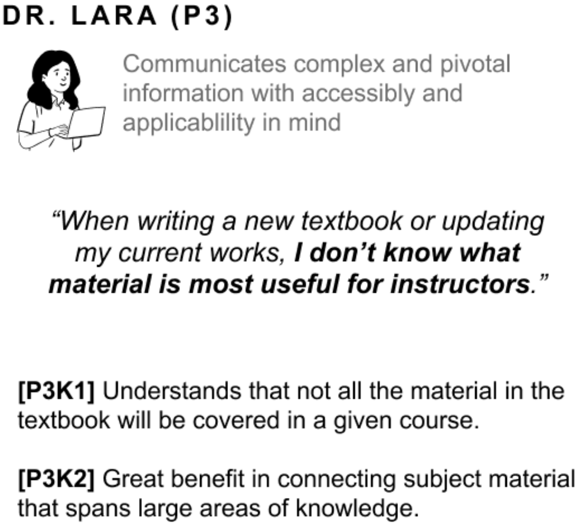
Dr. Lara and other textbook authors create complex textbook material in an accessible and applicable manner (Persona 3). However, there is little to no guidance to ensure that topics and diagrams help connect to other subject materials in STEM [P3K2]. This is where publishers can assist when creating product suites to aid authors as they create textbook content. While Bailey is equipped to ensure that the content is digitally accessible and high quality (Persona 4), she is currently unable create her version of the ITB due to current product features and lack of expertise [P4K1]. When authors and publishers have a lack of insight as to how students and teaching assistants are currently using their tools, our partnership and collaborative creation of ITB will be pave the way for the next wave of innovation for their product suite and content curation toolset.

Lesson Learned and Adaptations from User Research

We can organize the lessons learned from user research into several categories: confirmation, new software tool ideas, and new partnerships explained in detail.

Confirmation. P1K1 confirms our vision that there is a hunger and need to go beyond the current generation of textbooks. P4K1 and P4K2 reinforce the market reality that there are numerous innovation options which makes it challenging for publishers to make choices despite their commitment to bringing in new technologies to the market.

Publishing - ITB Content Creators



DR. LARA (P3)

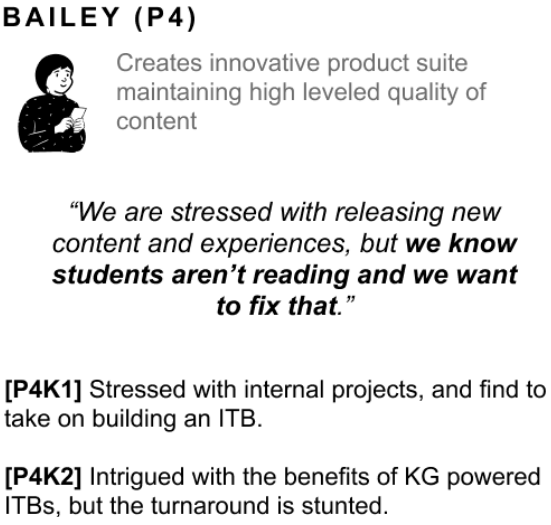
Communicates complex and pivotal information with accessibly and applicablility in mind

*"When writing a new textbook or updating my current works, **I don't know what material is most useful for instructors**."*

[P3K1] Understands that not all the material in the textbook will be covered in a given course.

[P3K2] Great benefit in connecting subject material that spans large areas of knowledge.

Author persona and key insights



BAILEY (P4)

Creates innovative product suite maintaining high leveled quality of content

*"We are stressed with releasing new content and experiences, but **we know students aren't reading and we want to fix that**."*

[P4K1] Stressed with internal projects, and find to take on building an ITB.

[P4K2] Intrigued with the benefits of KG powered ITBs, but the turnaround is stunted.

Publisher persona and key insights

New software tool ideas: The use of diagrams by TAs as suggested in P2K1 is a powerful insight and suggests exploration of tools that could leverage KGs for summarizing and synthesizing information. The authors' interest in new tools as in P3K1 for their books suggests the possibility of creating a vocabulary tool that can perform term analysis and help an author in being rigorous with their writing.

New Collaborations: Several new opportunities for collaboration resulted from our user research. During our interview with Prof. Scott Dixon, who teaches an introductory Biology class at Stanford, expressed a strong interest in trying out an ITB for his class. We interviewed Educational Testing Service (ETS) as a representative of an educational technology company, and they have now become a member of the team. Macmillan Learning offered to organize user focus groups to help us obtain more feedback on existing ITBs. Finally, we have developed a closer cooperation with the product team at OpenStax with the intention of creating tools that could be incorporated into their learning platform.

Phase I Creative Products

Towards the goal of creating KGs for ITBs, we have produced the initial prototypes for the following tools: (1) Automated term extractor: This tool can automatically process a book, and identify key terms with a precision of 0.73, thus significantly reducing the human effort. (2) Concept mapping tool: This tool first teaches a crowd worker the task of concept mapping through common sense examples. It then presents textbook content to the worker one sentence at a time and asks for the relation between two key terms to be identified. (3) Denoising algorithms: We developed algorithms that can take the crowd data and denoise them to identify consensus opinion, consider worker ability, and task difficulty. (4) A knowledge graph for a section of the Biology textbook. (5) An ITB prototype for an OpenStax Biology textbook.

Ours is a brand-new team, and the researchers from Rice, Stanford, and OpenStax had never worked together before. Through the work products created during Phase I, we have also established a collaborative relationship that sets us up to succeed during Phase II.

Contributions to Track Success

We will undertake the following integrative activities to ensure the overall success of the program.

PROMOTION THROUGH CONFERENCES AND JOURNALS

We will promote the work being done in the Convergence Accelerator at leading academic venues. Dr. Chaudhri is on the editorial board of the AI Magazine, and will propose a special issue devoted to showcasing the work being done in the program. Prof. Mark Musen will propose a workshop on knowledge graphs (KG) to be held in conjunction with the International Semantic Web Conference.

VISION WORKSHOP ON TEXTBOOK OF THE FUTURE

In 2006, National Science Foundation sponsored a workshop on the theme of *Textbook Reconsidered* (Bierman et. al., 2006). The genesis of the workshop was the observation that for more than a century, the printed textbook has shaped the curriculum in most science, math, and technology disciplines. As the world wide web has given many students 24/7 access to information, interactive exercises, and dynamic simulations, in the face of this new reality, the workshop organizers hoped to answer the question: Will the reign of the textbook continue or is it time for printed texts to retire, going the way of the abacus and the slide rule?

Since the time of this workshop, much has changed in the education landscape. We have had the advent of Massive Open Online Courses or MOOCs (Pappano, 2012), electronic textbook readers have become common place, and learning management systems abound. And yet, the textbooks are an integral component of the modern education system. Indeed, the textbook authors have the role of curating the knowledge of humanity and packaging it in a form through which we can preserve our knowledge as well as pass it on to the future generations. Our belief is that the textbooks will remain part of our education system for a very long time, and the right question to ask is: how can we revolutionize the textbook using modern technology?

The vision of Intelligent Textbooks (ITB) pursued relies on the formal representation of the knowledge in a book. This representation should enable one to answer deep questions on the subject matter of the book at the same level as the author, not just give replies to superficial questions. At least in principle, it should allow one to build an "expert system" that can do the same. The representation would not replace textbook but provide a foundation on which to develop pedagogy. A complete formal representation can enable much more. It would allow the textbook to adapt to the student. For example, the intelligent text could go over portions of the material again if the student were to answer exercises incorrectly. It could design new examples. It could provide a simulation environment in which the student could test out his understanding.

We will organize a workshop on the theme of *Textbook of the Future* through which we will present what we have achieved with an Intelligent Textbook, and yet challenge the participants to think beyond the incremental improvements which are the norm today. As the concept of a textbook is naturally integrative for academics, it would be easy to involve the participants within the track, our commercial and noncommercial publishing partners, and leading thinkers from the broader academic community. The proposed Workshop will socialize the achievements of our project as well as bring a change in the mindset on how we design, author, create, and use textbooks. It may also inspire a new initiative to continue the transformation of textbooks.

TEXTBOOK KGs AS A SERVICE

Through our participation in the Phase I meetings, and in discussing our research with other participants, a general model of collaboration has emerged. Given exists a body of circumscribed knowledge documented in a textbook like resource, we will design and create a KG. The resulting KG enables a downstream application such as an intelligent tutor, a training system, a query answering system, a question generation system, etc. We envision several collaborations described below in which our KG expertise could enhance the success of our C-ACCEL colleagues and their Phase II projects.

Prof. Beverly Woolf (team B-6915) is developing a framework for diagnosis, recommendation, and training in continuous workforce development. They are interested in developing an intelligent tutoring component to help workers operating robots to acquire new skills. As the skills are explained in training manuals, the knowledge networks created from such manuals would enable construction of more powerful intelligent tutoring systems than would be possible otherwise. (A letter of collaboration is attached.)

Prof. Binil Starly (team A-7043) is creating a KG for product design. They are very interested in incorporating a KG from a manufacturing textbook. We will work with them to understand the requirements of their application, design a KG, and provide them an initial prototype that they could evaluate. (A letter of collaboration is attached.)

We have had interest from several instructors to test out the ITB concept in conjunction with their textbooks: Prof. Pevel Pevzner, UC San Diego for his textbook on bioinformatics; Prof. Scott Dixon, Stanford University, for a textbook on Biochemistry; Prof. Linqun Luo, Stanford University, for a textbook on Neurobiology. We will collaborate with these interested partners to formulate what should be provided in *Textbook KG as a Service,* so that they can leverage ITBs in the context of their textbooks.

LEVERAGING SOURCES OF PUBLIC DATA

The core of the effort in our project relies on creating new KGs. We will evaluate other sources of public data that may be available to us. We have identified two such specific sources. First, there exist public ontologies for anatomy, for example, the Foundational Model of Anatomy, and Uber Anatomy Ontology (Uberon), that could be leveraged in the context of anatomy textbooks. Second, we will work with the Microsoft Academic Graph team to investigate if they could create and provide us with an initial version of taxonomy for the textbooks that our effort could leverage.

KNOWLEDGE GRAPH EDITING, REASONING AND DISSEMINATION PLATFORM

Our team has developed important community infrastructure for KGs that includes: Protégé and Web Protégé for browsing and editing ontologies, BioPortal for accessing a library of biomedical ontologies, and CEDAR (Center for Expanded Data Annotation and Retrieval) to create a computational ecosystem for development, evaluation, use, and refinement of biomedical metadata.

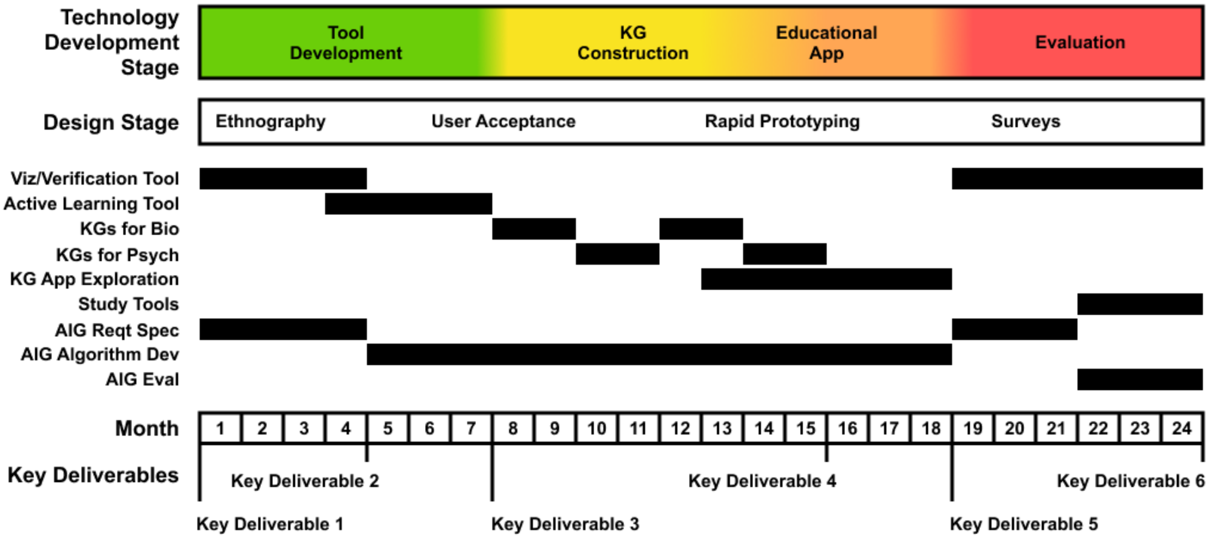
Several teams within the Convergence accelerator are already using Protégé. Prof. Hitzler on team A-6677 is developing a spatial open knowledge network. One component of their work is to develop a Protégé plugin for modular knowledge engineering. Prof. Nora El-Gohary (team A-7115) has been using Protégé for creating ontologies for her project on Civil Infrastructure. Dr. Lisa Kerr (team A-6950) uses Protégé for her work on ocean decision sciences.

Our team will provide our ontology design expertise, and training support to use Protégé, and related tools. We will support the program participants to publish the knowledge graphs they create through Bio Portal. We will advise them in creating meta data for their projects based on the best practices developed in our CEDAR project.

KNOWLEDGE GRAPH EDUCATION

We will create a course and a textbook on knowledge graphs. Towards this goal, we organized a seminar on knowledge graphs during Spring 2020. The seminar received enthusiastic response with over 1300 sign ups, and the course videos have already received thousands of views. We are planning to offer the seminar again during Spring 2021. The seminar will evolve into a course, textbook, and exercises resulting from the projects in the convergence accelerator.

Timeline



PHASE II DELIVERABLES

Key Deliverable 1 (month 0): Improved versions of Phase I tools for term extraction, relation extraction, denoising crowd data, KG construction and concept mapping.

Key Deliverable 2 (month 6): We will release a tool for visualization and verification of the knowledge graph. This will be a new capability, and it will be released as part of the Protege environment.

Key Deliverable 3 (month 9): We will release an active learning tool so that the new training data is used to improve the term and relation extraction tools.

Key Deliverable 4 (month 15): Knowledge graphs for two chapters from the OpenStax Biology and one chapter from the OpenStax Psychology textbooks. Report on time and effort required to create the KG.

Key Deliverable 5 (month 18): We will release new ITB tools that will result from the exploration of KG supported study tools.

Key Deliverable 6 (month 24): Evaluation report from study tools experiment, and automatic item generation evaluation.

DELIVERABLES PAST PHASE II

Milestone (Phase II + 24 months): Release of Textbook scale TOKN for both Biology and Psychology. Integration of ITB concepts into selected portions of OpenStax Biology and Psychology textbooks at OpenStax.

Milestone (Phase II + 60 months): Expanded coverage of TOKN to include multiple textbooks. Integration of ITB concepts into full length OpenStax Biology and Psychology textbooks.

Milestone (Phase II + 120 months): TOKN with a coverage of all textbooks in the OpenStax textbook library. Integration of ITB concepts into the full OpenStax library.

CACCEL PHASE II - Track A1 (Open Knowledge Network):
A7134: Textbook Open Knowledge Network
Stanford University
NSF Convergence Accelerator Pilot (NSF 20-055)

References

Augustine, N. R. (2007) Committee on Prospering in the Global Economy of the 21st Century. *Rising above the gathering storm: Energizing and employing America for a brighter economic future.* Washington, DC: National Academies Press.

Barker, K., Porter, B., & Clark, P. (2001, October). *A library of generic concepts for composing knowledge bases*. In Proceedings of the 1st international conference on Knowledge capture (pp. 14-21).

Bejar, I. I. (1985). Speculations on the future of test design In S. E. Embretson (Ed.), *Test design: Contribution from education and psychometrics (pp 279-294)*. New York Academic Press.

Bejar, I. I. (2002). Generative testing: From conception to implementation. In S. H. Irvine & P. C. Kyllonen (Eds.), *Item generation for test development* (pp. 199-218). Mahwah, NJ: Earlbaum.

Bejar, I. I., Flor, M., Deane, P., McCaffrey, D., Holtzman, S., & Bruno, J. E. (2018). *The feasibility of generating vocabulary items for the Word Knowledge ASVAB subtest* (Contract No: GS09T12BHD0012, Task Orders 36 and 69). Retrieved from

Bejar, I. I., Lawless, R. R., Morley, M. E., Wagner, M. E., Bennett, R. E., & Revuelta, J. (2003). A feasibility study of on-the-fly item generation in adaptive testing. *Journal of Technology, Learning, and Assessment*, *2*(3), Available from http://www.jtla.org.

Bennett, B., Chaudhri, V., & Dinesh, N. (2013, September). *A vocabulary of topological and containment relations for a practical biological ontology*. In International Conference on Spatial Information Theory (pp. 418-437). Springer, Cham.

Berant, J., Srikumar, V., Chen, P. C., Vander Linden, A., Harding, B., Huang, B., ... & Manning, C. D. (2014, October). *Modeling biological processes for reading comprehension.* In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP) (pp. 1499-1510).

Zhou, S., Gordon, M., Krishna, R., Narcomey, A., Fei-Fei, L. F., & Bernstein, M. (2019). *HYPE: A Benchmark for Human eYe Perceptual Evaluation of Generative Models*. In Advances in Neural Information Processing Systems (pp. 3444-3456).

Bierman, P., Massey, C., & Manduca, C. (2006). Reconsidering the textbook. *Eos, Transactions American Geophysical Union*, *87*(31), 306-306.

Carnegie Learning. (2019), *Math Curriculum & Software Solutions.* Retrieved from http://www.carnegielearning.com

Chaudhri, V. K., Cheng, B., Overtholtzer, A., Roschelle, J., Spaulding, A., Clark, P. & Gunning, D. (2013). *Inquire Biology: A textbook that answers questions.* AI Magazine, 34(3), 55-72.

Chaudhri, V. K., & Dinesh, N. (2013b). Conceptual models of structure and function. In Proceedings of the Second Annual Conference on Advances in Cognitive Systems ACS (Vol. 255, p. 271).

Chaudhri, V. K., Elenius, D., Hinojoza, S., & Wessel, M. A. (2014). *KB_Bio_101: Content and Challenges.* In FOIS (pp. 415-420).

Chaudhri, V. K., Dinesh, N., & Heymans, S. (2014b). *Conceptual Models of Energy Transfer and Regulation.* In FOIS (pp. 263-276).

Chin, C., & Osborne, J. (2008). Students' questions: *A potential resource for teaching and learning science.* Studies in Science Education, 44(1), 1–39.

Clark, M.A., Douglas, M., & Choi J. (2018). *Biology 2e.* OpenStax. Houston, TX. Retrieved from https://openstax.org/details/books/biology-2e.

Clough, G. (2005). Educating the engineer of 2020: Adapting engineering education to the new century. *National Academy of Engineering*, Washington, DC.

Compeau, P., & Pevzner, P. A. (2018). *Bioinformatics Algorithms: An Active Learning Approach. La Jolla*. CA: Active Learning Publishers.

Cole, B., Lima-Walton, E., Brunnert, K., Vesey, W., & Raha, K. (2019). Taming the Firehose: Unsupervised Machine Learning for Syntactic Partitioning of Large Volumes of Automatically Generated Items to Assist Automated Test Assembly. *Journal of Applied Testing Technology, 21*(1), 1-11. doi:http://jattjournal.com/index.php/atp/article/view/146483

Deng, J., Russakovsky, O., Krause, J., Bernstein, M. S., Berg, A., & Fei-Fei, L. (2014, April). Scalable multi-label annotation. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 3099-3102).

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). *BERT: Pre-training of deep bidirectional transformers for language understanding.* arXiv preprint arXiv:1810.04805.

Fernandez, T., Godwin, A., Doyle, J., Verdin, D., Boone, H., Kirn, A., … Potvin, G. (n.d.). More Comprehensive and Inclusive Approaches to Demographic Data Collection. 2016 ASEE Annual Conference & Exposition Proceedings. doi: 10.18260/p.25751

Gisborne, N., & Donaldson, J. (2019) Thematic Roles and Events. In *The Oxford Handbook of Event Structure*. Oxford University Press,

Grimaldi, P. J., Poston, L., & Karpicke, J. D. (2015). How does creating a concept map affect item-specific encoding? Journal of Experimental Psychology: Learning, Memory, and Cognition, 41(4), 1049.

Karpicke, J. D. (2009). *Metacognitive Control and Strategy Selection: Deciding to Practice Retrieval During Learning.* Journal of Experimental Psychology: General, 138(4), 469–486.

Karpicke, J. D., & Grimaldi, P. J. (2012). Retrieval-Based Learning: A Perspective for Enhancing Meaningful Learning. Educational Psychology Review, 24(3).

Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. Science, 331(6018), 772-775.

Koedinger, K. R., Anderson, J. R., Hadley, W. H., & Mark, M. A. (1997). *Intelligent Tutoring Goes To School in the Big City*. International Journal of Artificial Intelligence in Education, 8, 30–43.

Kotturi, Y., Kulkarni, C. E., Bernstein, M. S., & Klemmer, S. (2015, March). *Structure and messaging techniques for online peer learning systems that increase stickiness.* In Proceedings of the Second (2015) ACM Conference on Learning@ Scale (pp. 31-38).

Krishna, R., Zhu, Y., Groth, O., Johnson, J., Hata, K., Kravitz, J., ... & Bernstein, M. S. (2017). *Visual genome: Connecting language and vision using crowdsourced dense image annotations.* International Journal of Computer Vision, *123*(1), 32-73.

Kulkarni, C. E., Bernstein, M. S., & Klemmer, S. R. (2015, March). *PeerStudio: rapid peer feedback emphasizes revision and improves performance.* In Proceedings of the second (2015) ACM conference on learning@ scale (pp. 75-84).

Lan, A. S., Waters, A. E., Studer, C., & Baraniuk, R. G. (2014). Sparse factor analysis for learning and content analytics. The Journal of Machine Learning Research, 15(1), 1959-2008.

Ling, X., Clark, P., & Weld, D. S. (2013, October). *Extracting meronyms for a biology knowledge base using distant supervision.* In Proceedings of the 2013 workshop on Automated knowledge base construction (pp. 7-12).

Liu, A., Soderland, S., Bragg, J., Lin, C. H., Ling, X., & Weld, D. S. (2016, June). *Effective crowd annotation for relation extraction.* In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (pp. 897-906).

Luo, L. (2015). *Principles of neurobiology*. Garland Science.

Madnani, N., Beigman Klebanov, B., Loukina, A., Gyawali, B., Lange, P., Sabatini, J., & Flor, M. (2019, jul). *My Turn To Read: An Interleaved E-book Reading Tool for Developing and Struggling Readers.* Paper presented at the Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: System Demonstrations, Florence, Italy. https://www.aclweb.org/anthology/P19-3024

Miyatsu, T., Nguyen, K., & McDaniel, M. A. (2018). *Five Popular Study Strategies: Their Pitfalls and Optimal Implementations.* Perspectives on Psychological Science, 13(3), 390–407.

Mugnier, M. L. (2012, September). Existential rules: a graph-based view. In International Datalog 2.0 Workshop (pp. 21-26). Springer, Berlin, Heidelberg.

Musen, M.A. (2015). *The Protégé project: A look back and a look forward. AI Matters*. Association of Computing Machinery Specific Interest Group in Artificial Intelligence, 1(4).

Musen MA, Bean CA, Cheung K-H, Dumontier M, Durante KA, Gevaert O, Gonzalez-Beltran A, Khatri P, Kleinstein SH, O’Connor MJ et al.. 2015. The Center for Expanded Data Annotation and Retrieval. Journal of the American Medical Informatics Association, JAMIA. 22 (6) 1148-1152.

Newstead, S. E., Bradon, P., Handley, S. J., Dennis, I., & Evans, J. S. B. T. (2006). Predicting the difficulty of complex logical reasoning problems. *Thinking & Reasoning, 12*(1), 62-90.

NITRD, N. (2018). *Open Knowledge Network: Summary of the Big Data IWG Workshop. [*online] Available : https://www.nitrd.gov/news/Open-Knowledge-Network-Workshop-Report-2018.aspx

Noy, N. F., Shah, N. H., Whetzel, P. L., Dai, B., Dorf, M., Griffith, N., & Musen, M. A. (2009). *BioPortal: ontologies and integrated data resources at the click of a mouse.* Nucleic acids research, 37(suppl_2), W170-W173.

OpenStax. (2019). OpenStax Tutor. https:/tutor.openstax.org

Pappano, L. (2012). The Year of the MOOC. *The New York Times*, *2*(12), 2012.

Pugh, D., De Champlain, A., Gierl, M., Lai, H., & Touchie, C. (2016). Using cognitive models to develop quality multiple-choice questions. *Medical Teacher, 38*(8), 838-843. doi:10.3109/0142159X.2016.1150989.

Ratner, A., Bach, S. H., Ehrenberg, H., Fries, J., Wu, S., & Ré, C. (2020). *Snorkel: Rapid training data creation with distant supervision.* The VLDB Journal, 29(2), 709-730.

Reece, J. B., Urry, L. A., Cain, M. L., Wasserman, S. A., Minorsky, P. V., & Jackson, R. B. (2014). *Campbell biology* (No. s 1309). Boston, MA: Pearson.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., ... & Berg, A. C. (2015). Imagenet large scale visual recognition challenge. *International journal of computer vision*, *115*(3), 211-252.

Sadava, D. E., Hillis, D. M., & Heller, H. C. & Hacker, S. (2016). *Life: the science of biology*, 11th Edition, Macmillan.

Settles, B. (2009). *Active learning literature survey*. University of Wisconsin-Madison Department of Computer Sciences.

Spielman, R. M., Jenkins, W. J., & Lovett, M. D.(2020). *Psychology 2e*. OpenStax. Houston, TX. Retrieved from https://openstax.org/details/books/psychology-2e

Steenbergen-Hu, S., & Cooper, H. (2014). *A meta-analysis of the effectiveness of intelligent tutoring systems on college students' academic learning.* Journal of Educational Psychology, 106(2), 331–347.

Tymoczko, J. L., Berg, J. M., & Stryer, L. (2011). *Biochemistry: a short course*. Macmillan.

Wang, Z., Lan, A. S., Nie, W., Waters, A. E., Grimaldi, P. J., & Baraniuk, R. G. (2018, June). QG-net: a data-driven question generation model for educational content. In Proceedings of the Fifth Annual ACM Conference on Learning at Scale (p. 7). ACM.

Wang, K., Shen, Z., Huang, C., Wu, C. H., Dong, Y., & Kanakia, A. (2020). Microsoft Academic Graph: When experts are not enough. Quantitative Science Studies, 1(1), 396-413.

Waters, A. E., Studer, C., & Baraniuk, R. G. (2014). Collaboration-type identification in educational datasets. Journal of Educational Data Mining, 6(1), 28-52.

Waters, A. E., Tinapple, D., & Baraniuk, R. G. (2015, March). BayesRank: A bayesian approach to ranked peer grading. In Proceedings of the Second (2015) ACM Conference on Learning@ Scale (pp. 177-183).

Watts, M., & de Jesus, H. P. (2005). *The cause and affect of asking questions: Reflective case studies from undergraduate sciences.* Canadian Journal of Science, Mathematics and Technology Education, 5(4), 437–452.

Weld, D. S., Lin, C. H., & Bragg, J. (2015*). Artificial intelligence and collective intelligence.* Handbook of Collective Intelligence, 89-114.

Zeng, D., Liu, K., Chen, Y., & Zhao, J. (2015, September). *Distant supervision for relation extraction via piecewise convolutional neural networks.* In Proceedings of the 2015 conference on empirical methods in natural language processing (pp. 1753-1762).