Abstract

In this paper, we give an overview of Project Halo, and how its technology could be applied to address the needs of an Internet Classroom. In particular, we describe how the creation of large-scale, socially-authored knowledge bases for classroom subjects, coupled with systems for answering sophisticated questions based on the information in these knowledge bases, can provide a revolutionary new method for students to interact with the subject matter of the curriculum.

Overview of Project Halo

The long-term goal of Project Halo is to build an application called Digital Aristotle that can answer questions on a variety of science topics and provide user and domain appropriate explanations.

As a near-term goal, we have been developing a suite of knowledge acquisition and question answering tools. Specifically,

- A system called AURA to capture knowledge from science textbooks and answer questions from an Advanced Placement Exam
- An extension to Semantic MediaWiki to embed semantic markup in online text documents
- An ontology mapping tool to import the marked up content from Semantic Media Wiki into AURA so that it could be used for deductive question answering

The AURA system enables graduate students in science to encode knowledge from a science textbook in such a way that high school students can receive correct answers to Advanced Placement Exam questions concerning that knowledge (Chaudhri, John et al. 2007), (Clark, Chaw et al. 2007). We have been focusing on three particular science textbooks for the domains of physics, chemistry, and biology (Campbell and Reece 2001), (Giancoli 2004), (Brown, LeMay et al. 2003). The AURA system can currently support knowledge capture and question answering such that it can correctly answer 50% of the questions on a suite of AP questions. An independent evaluation of this system on unseen questions is planned for the Fall of 2008.

Using the Halo extension of Semantic MediaWiki, a user can annotate a scientific page that has been copied from the normal English-language Wikipedia into a separate environment based on the Semantic MediaWiki platform (Krötzsch et al., 2006). This
extension of the MediaWiki platform (which is used by Wikipedia) supports typed links, each of which encodes a property (relation or attribute) of the entity described by the page in question. We employ a graphical schema-mapping tool called OntoMap® to link the logical knowledge in the Semantic MediaWiki with knowledge in AURA so that AURA can use the combined knowledge for question answering. OntoMap® is a schema-mapping tool that enables the specification of relations between different schemas in a declarative way (De Brujin, J., et. al. 2004).

The current focus in Project Halo is limited to a small portion of the Advanced Placement syllabus in the domains of Physics, Chemistry, and Biology. We anticipate that in the future phases of Project Halo, we will expand its scope to full syllabus in these three domains, and also expand into new Advanced Placement domains such as Micro Economics, Environmental Sciences, and US Government and Politics.

**Internet Classroom and Project Halo**

We envision that a key feature of the Internet Classroom will be to provide free access to a variety of curricula, supported by AI and Web 2.0 technologies. A precondition to this feature is access to high-quality repositories of computer-processable domain knowledge that can be used for different pedagogical purposes, from computer-assisted tutoring to student supplementary material to student assessment. AURA, augmented by our extensions to Semantic MediaWiki, can provide these knowledge bases and the associated question answering systems. More importantly, though, AURA’s focus on student-based knowledge curation allows students to be directly involved in the design, construction, and validation of these knowledge bases. We see five main areas where AURA and other Project Halo technologies can help make the Internet Classroom a reality.

First, the emphasis on student authorship and debugging of sophisticated academic knowledge bases is one of the powerful features that makes AURA unique. AURA is designed to be used by students as an easy-to-use “calculator” for concepts, allowing students to create, play with, and explore the consequences that come from different ways of formulating the underlying conceptual structures in the AP science disciplines. In this way, AURA brings a new and powerful tool to the classroom – one which has previously only been available to professional knowledge engineers. With AURA, we can leverage the power of computers (for logical calculation) and the power of the Internet (for wiki-based dynamic social authoring and validation) into a new capability for the Internet Classroom which isn't available with traditional classroom-based pedagogical methods.

Second, and beyond the direct educational benefits that can accrue from students jointly constructing and interacting with Halo-cultivated knowledge bases, the resultant knowledge bases themselves are a key technology for the Internet Classroom. Tutoring systems that leverage this core subject knowledge with pedagogical techniques such as explanation, prompts, hints, and formative assessments can deliver one-on-one tutoring that can adapt to a student’s needs. Furthermore, the reasoning engine that produces answers to questions in AURA could be used in novel ways. For example, a tutor could use AURA to produce explanations of increasingly complex concepts that build on prior
questions. An AURA knowledge base could allow a teacher to select increasingly sophisticated questions that students should pursue in more depth, and rapidly zero in on areas of student strength and weakness. With better, more targeted data from formative or standardized assessments, K-12 teachers in the Internet Classroom will be able to precisely identify specific areas for study for individual students.

A third area of application for AURA is as a sophisticated concept mapping tool. The usefulness of concept maps is well-known in education, but at present, concept mapping tools are typically domain independent and require the student to build, from scratch, all of the requisite nodes and links to capture complex structures or processes. Using AURA as a concept mapping tool, students can start from an exploration of clear declarative semantics with the concept maps that can serve as a basis for answering questions. Alternatively, student-generated concept maps can be compared with teacher-provided ‘ground truth’ maps in AURA, and tutoring, coaching, or self-directed learning can help students clear up discrepancies. Placing a concept mapping application into an Internet Classroom opens up the possibility of students collaboratively exploring science content using shared concept map overlays on the AURA knowledge. These overlays could represent students’ developing understanding of a particular set of related ideas. Students could assess each other’s overlays (i.e., understanding) by tracing the maps to generate explanations and seeing where the maps fall short.

Fourth, AURA and its accompanying Semantic MediaWiki provides a clear way to construct freely accessible knowledge bases, and link them to the emerging freely-accessible educational materials that will be characteristic of the Internet Classroom. MediaWiki can easily serve as a platform for supplementary material to textbooks; in fact, there are a whole host of textbooks and supplementary material that are being developed as part of the Curriki effort and others. This structured content could be linked with AURA’s concept maps and knowledge bases to provide additional tutorial and exercises based on the question answering ability of AURA.

Finally, the development of an Internet Classroom could provide an orienting technical challenge that could be used to promote interest in computer science education – especially computer science with an interdisciplinary focus. Aura is a powerful tool to learn about and experiment with different styles of conceptual modeling. Given this, joint teams of computer scientists and science domain experts could collaborate to create freely accessible and highly usable knowledge repositories. In AURA, they could work in a bounded programming “sandbox”, programming a piece of knowledge, then testing and debugging it until the target question is answered correctly by AURA. These design and testing skills are core engineering skill cultivated in computer science programs, and could entice students earlier in their careers to pursue computer science. Also, programming science knowledge provides a bridge between computation and science, furthering the development of students interested in computational science from the perspective of their own disciplines.

The Internet Classroom cannot be realized without provision for computer-based representations of science knowledge that can be using in reasoning, tutoring, warranting,
and learning about programming. The Halo software could provide a strong starting point for this need.

Acknowledgment

The research discussed in this paper is being funded by Vulcan Inc. in the context of their Project Halo.


