Journal of Intelligent Information Systems, 18, 5–14, 2002

Editorial—A Survey of Research Questions for Intelligent Information Systems in Education

JOHN C. CHERNIAVSKY jchernia@nsf.gov
National Science Foundation, 4201 Wilson Blvd, Room 855, Arlington, VA 22230

ELLIOT SOLOWAY soloway@eecs.umich.edu
Department of EECS, University of Michigan, 1101 Beal Avenue, Ann Arbor, MI

Received August 9, 2001; Accepted October 9, 2001

Abstract. Education is an application domain in which many research questions from Intelligent Information Systems may prove their worth. We discuss three themes in this editorial: distributed education and learner modeling, semantic analysis of text, and intelligent information management.

Keywords: education, intelligent information systems, research questions

Introduction

We present a challenge to the IIS research community to take on projects that will improve the human condition. Because of our interests, we believe that the most lasting and important application area for these projects is education and learning: for our children for their future, for many for their employment, and for all for the betterment of life.

In this survey we discuss many of the research areas represented by the IIS community that are relevant to education and learning. Information and intelligent systems can be interpreted very broadly. Humans interact with their environment through the intelligent use and manipulation of information and so a broad interpretation covers all human activities. We narrow the interpretation to that used by the JIIS so that information means data bases or other digitized and organized data sources and intelligence refers to computer based intelligence which ranges from AI tools and techniques to mathematical approaches to computer learning and inference.

The existing environment for learning consists of formal schooling; employment related learning (this includes training); and learning for life (learning for your own interests). Schooling occurs in primary, secondary, university, Ph.D., and postgraduate settings. Self motivation for learning increases with educational level. The primary, secondary, and university settings are typically classroom settings and are characterized by lectures, student problem solving, and assessments. Ph.D. and postgraduate settings are mostly apprenticeships—that is learning by watching, emulating, and doing under the supervision of mentors. Employment related learning occurs both on the job through apprenticeships and through training either in classroom settings or self-learning settings through the internet, CD-ROM, or written materials. Learning for work is usually motivated by economic considerations. Learning for life can be learning in any setting and is fully self motivated. The technology that is currently used in these learning setting is uninspiring. The face to face classroom lecture is replaced by the internet lecture; the student/teacher face to face interaction is replaced by email interactions; and the paper based materials are replaced by web pages. As has been noted many times before, this situation in educational technologies is similar to early film making when the films faithfully recorded live stage performances. It was only later that the new affordances of film were utilized to tell compelling stories and to fundamentally change entertainment.

The general state of schooling is especially grim in U.S. primary and secondary schools. Our schooling model dates from the early 20th Century as an efficient model for educating the manufacturing workers of the day. The workers of today are more likely to be knowledge workers in the service sector than workers on the shop or factory floor. This implies a change in what students' should be learning—a change towards learning how to learn. Learning how to learn has always been characteristic of scholarly work and the knowledge worker increasingly is doing scholarly work both in the depth of analysis and in the intellectual flexibility required to be current in an environment with rapid technology changes. We envision a more distributed schooling model where collaboration and self study take on an increasingly important role. The role of the teacher will change from the classroom authority to a mentor and a coach. The information technology to support the change in schooling will draw heavily on research in intelligent information systems. An example of the sorts of learning environments that are likely in the future is explored in Roger Schank's vision of virtual learning environments (Schank, 2001). In this paper we also discuss IIS research necessary for technology enhanced learning for work and learning for life.

Distributed pedagogy and learner modeling

Below we present a scenario arising from a vision of distributed education mediated by technology. Imagine a school of tomorrow that is distributed across space and time. Imagine easy access to information resources, mentors, and fellow students. Imagine telepresence with so high a quality so that the virtual is nearly real with gestures, facial expressions, and body movements fully communicated along with voice. These affordances can fundamentally change the quality of learning and prepare the learner for life in a world of information.

We speculate on how these affordances might be used in a distributed schooling environment that supports student learning and then identify research questions appropriate for the IIS community.

Scenario—the distributed learner

Peter begins his day by checking his overnight messages.

"Hmm," he thinks. "There is a lot of activity in the pollution study group. I had better process the micro climate data from the school's instruments and incorporate it into the town pollution model."

After incorporating data from the pollution monitors he ran a visualization of the state of his town Milford. Noting missing data from his friends Alan and Joy, he sent messages reminding them to update their data. He noted that his prediction on turbidity in the local streams was confirmed but that his prediction on decreased air pollution levels was wrong. He went back to his causal model to see what factors he had missed.

“Where is the missing pollution coming from?,” he thought. “I’ll bring it up in class.”

Peter then started his mathematics homework. The problem of the day dealt with the mathematics of diffusion. After running several simulations to develop a feel for the mathematics, Peter went to diffusion databases on water and air pollution to see how accurately the models predicted real world diffusion. Then he went on to the equations of change and studied the boundary conditions of the real world problem given him. It was now time for science class.

Peter turned on his virtual presence cameras and projectors and his friends Joy from Middletown and Alan from Derby appeared along with the pollution group’s mentor, Phil, from New Haven.

“OK guys, I don’t know what went wrong with my air pollution prediction. According to the model, the arrival of the cold front should have reduced the Sulfur and particulates but they increased instead!,” said Peter.

“Yeah,” said Alan, “I also got an increase when I should have gotten a decrease—mine’s even bigger than yours!. There must be a new pollution source west of us and I’m probably closer.”

“My predictions were right on.”, said Joy. “This means it’s not North of us.”

“Why don’t you use the wind data to see if you can track the source of the pollution.”, said Phil.

The group did this by querying the weather service database online. They then calculated possible areas for a new pollution source.

“Hmm,” thought Peter, “maybe my diffusion studies would give us a distance estimate.”

“Look guys,” said Peter, “we can use the difference between Alan’s and my readings to get an estimate of the diffusion rate. If we can get a boundary condition, then we could estimate the distance from Alan that the pollution is coming from.”

“Why don’t you query the web to see if you can find information sources about typical pollution outputs from factories or power plants?”, said Phil.

After searching the EPA web site, the group found references to typical power plant outputs. Feeding these into their model and using Peter’s diffusion models, they were able to narrow the search down to three possible sources in New Jersey. The group reported their results to the EPA who found one power plant whose scrubbers weren’t working properly.

In order to make the above scenario real, a number of IT support systems must be in place. The students need access to virtual presence hardware, high bandwidth networks, scientific networks of sensors, extensive data bases, and a rich digital library of educational resources. There is student modeling underlying both the level of difficulty of the material presented and in the design of the activity. There is middleware that mediates information access and fusion to produce appropriate views for the students. Implicit in the scenario is the assessment of student progress (part of the student modeling) and the management of resources that match mentors, students, and projects with a view to optimizing learning while containing costs.

These support system needs define education related IIS research to support the student modeling and distributed pedagogy. The desired digital library and scientific data base capabilities require new information system capabilities. Both the databases and the digital library need intelligent mappings from a model of a learner's knowledge to the information so that the information that is presented to the learner is at the right granularity and cognitive level. This motivates research in database and digital library views (Thuraisingham, 2001) but now coupled with changing needs driven by students' learning. An IIS research question is how to support this mapping? A start could be to take a model such as Anderson's ACT-R (Anderson et al., 1997) used in automated tutoring systems, then work on the details of the cognitive model of learning underlying the domain (a good opportunity for new research on domain ontology and epistemology), and develop the dynamic maps that support student learning using the underlying student model. The "far out" research question is to automate this process so that both the cognitive models and the maps to the domain can be more easily constructed.

Once the maps between the epistemologies imposed on the digital libraries and scientific databases and the cognitive models that represent student knowing have been built, there is the additional question of how best to represent the information given the existence of a rich multimedia environment. Finally, there should also be an intelligent natural language interface to the information so that the interactions between the learner and the needed information are seamless.

Going back to the scenario, we can see where good research can lead to the desired system affordances. Peter wakes up, collects data, incorporates it into a model, and then visualizes the results. This is collaborative learning, so he needs the data from his co-investigators. The data collection must be at the right level—Peter needs to understand what he's collecting. The model must be at the right cognitive level. No learning takes place by just plugging in numbers. It is also best that this be authentic—thus getting over the difficulties of physical system modeling using simulation can be useful, but affective factors favor the doing of real science. The visualization must be pegged at Peter's level of expertise. Peter must be able to use the visualization as a scientist would—to identify an underlying phenomenon quickly-but at the appropriate cognitive level. In going to his mathematics class, he has available a rich resource of complex materials. The diffusion equations would be appropriately simplified, the mathematics of change needed to understand the diffusion equations might have been learned using principles similar to those exhibited in middle school calculus curriculum (Roschelle et al., 2000), the boundary value settings and the understanding of the equations could come out of simulation and visualization software. To be able to do this naturally and individualized to each student's capability requires fundamental advances IIS.

Does this model of distributed learning and learner modeling really make sense? We think it does and we draw on current changes in how science and business is being done. Due to the global nature of many companies and the shortage of skilled personnel, the use of the distributed R&D teams is increasingly common. Typically, the team is both geographically and temporally distributed yet must perform a learning function—that is solving an R&D problem—using distributed collaboration. The current IIS to support this activity are the typical CSCW support—white boards, threaded discussions, asynchronous shared workspaces, and the like. No attempt is made to model expertise, no attempt is made to record and model the learning of team individuals or overall team learning. In addition, these are frequently multi-national efforts and neither linguistic nor cultural factors are incorporated into the systems. Current research on collaboration (Cassell, 2001) has shown the universality of multi-modal communication and the very powerful affective nature of the multi-modalities of communication (voice, gesture, and facial expression) can have in effective collaborations—even when mediated by avatars representing models of the collaborators. Where business goes, we feel the rest of society will follow and that is the responsibility of a good applied researcher to anticipate this and have the research ready for the application.

To summarize, the following are IIS research areas supporting distributed pedagogy and learner modeling:

- Rich modeling of learner knowledge in an epistemological sense
- Rich modeling of underlying human cognitive structures as information systems
- Intelligent views of epistemologies tied to the cognitive structures
- Effective information presentation dynamically tied to learner knowledge and cognitive structure
- Effective modeling of learner interactions using the underlying knowledge and learner cognitive levels

The above topics suggest both theoretical studies drawing on AI, Philosophy, Logic, Cognitive Science, Neuroscience, Social Sciences, and Knowledge Management and more applied research in building systems according to the theories and testing the systems for effectiveness.

Semantic analysis of text

A holy grail of the artificial intelligence community has been the development of a system that understands natural language. Recent results in information retrieval suggests that much of what is wanted from a full natural language understanding system can be gotten from statistical approaches to classifying text—especially newer approaches that may provide explanations for language understanding and acquisition. These systems are exemplified by recent work in Latent Semantic Analysis (Landauer et al., 1998) but also include the algorithms that underlie systems such as Google (Brin and Page, 1998) and Clever (Kleinberg, 1999). What can these systems do? In the education domain they are already used as a component of an automated grading tool for SAT essays. This use could clearly be expanded to large on-line English composition classes either at the university level or in K-12. LSA essay systems give an assessment of the semantic content of an essay based solely on statistical properties of word occurrences in texts that is robust even across different languages and which agrees remarkably well with semantic content assessment by human readers (Foltz et al., 1998). The extent to which LSA can be applied in other contexts is an active area of research and the topic of this special issue's two papers. But much more is possible and the possibilities are particularly intriguing in the use of digital libraries for learning, the next generation of automated tutoring systems, and in applications such as learning on demand.

We illustrate with some examples intended to illustrate the use of IIS in providing access to content for learning. The first example explores a student's use of tools in completing a history assignment. The second example explores the use of a tool that captures the learning that occurs in a group project. The third example explores some of the necessary components of using a digital library not only for content, but also for best practices in instruction.

Example 1—history lesson

Peter is writing an original essay on the history of Connecticut concentrating in particular on the Charter Oak incident and how is presaged the United States of America rebellion against the British. He has his technique for original research down to a routine. Having briefly heard the story of the Charter Oak where a document called the Connecticut Charter was hidden from an agent of James II in a White Oak tree, he determined his essay structure. First discuss the history of the charter in context of colonial America and the English government, then link to ideas behind the rebellion of the colonists. Having written this outline, he pressed a search button and links to original sources and interpretations were generated. He had a copy of the charter, an annotated history of English royalty at the time, a brief colonial history, and interpretations of the causes of the American revolution. Using this information, he proceeded to draft an essay outline. The Writing Tutor critiqued the outline and Peter used the comments to improve the outline and to write his first draft. The Writing Tutor critiqued that draft as to the plausibility of Peter's argument, the structure of the essay, the quality of the writing, and the adherence to his outline. At the end of an iterative process, after Peter felt he refined the essay sufficiently, the Writing tutor incorporated the references that Peter used and automatically corrected grammatical errors. After approval, Peter submitted the essay to his instructor for assessment along with the comments of the Writing Tutor and a description of the writing and re-writing of the essay. This work and his instructor's assessment became part of Peter's writing portfolio that would be assessed periodically, by the Writing Tutor, to give Peter and his parents feedback on Peter's progress.

Example 2—team learning affordance

The Project Coordinator is a program that is responsible for coordinating, interpreting, and analyzing the communications between distributed project teams on engineering projects. It collects all project related documents—emails, interactive meetings, project planning documents, and annotations of artifacts such as code documentation, requirements statements, etc. The program takes this information and analyzes it for consistency and completeness. The program notifies the project participants if it detects inconsistent or incomplete project components such as requirements, component functionality, and time lines for task completion. The Project Coordinator also summarizes the state of the project from the information it has and provides this information to the project participants and estimates the time to project completion. When the engineering project is complete, the Project Coordinator program, its data, and its interaction with the project team, is an object that captures the project design and history and can be used as a regression test for system changes or upgrades. It can also be incorporated into a knowledge repository that can be used to extract best practices, to trouble shoot problems, and to distill process knowledge. The project coordinator becomes an object, along with appropriate knowledge extraction tools, that enables engineering team communication and learning.

Example 3—digital library helper

Sara, a teacher, is interested in using the National Science, Mathematics, Engineering, and Technology (SMET) Digital Library (NSDL) to construct a lesson in Newtonian Mechanics. She logs into the NSDL Lesson Builder (which has a profile of Sara's use of the digital library and her preferences in pedagogy as exhibited by her use of the libraries resources) and types in an outline of her lesson plan. Since Sara wants to follow the State requirements on Physics teaching, she has informed the Lesson Builder that she wants her lessons to be coordinated according to the State's science standards. The Lesson Builder compares her lesson plan with previous lessons she has constructed and checks consistency with the State standards. The Lesson Builder advises Sara if she is missing a concept or element of the State standards and, if she wishes, will provide sample lessons. Further, the Lesson Builder has profiles of the assessments of Sara's students and is able to advise Sara of what concepts the students' need help with and suggests individually tailored lessons for those students. The Lesson Builder also analyzes the student profiles and advises Sara if the class as a whole is missing concepts that Sara thought had been mastered. This provides feedback to Sara as to what to review and whether she should consider a different approach to presenting the concept. Finally the Lesson Builder suggests simulations and laboratory experiments that would aid in understanding the concepts that she wants to convey.

After building and reviewing her lesson plan, Sara decides to review what others have done with similar concepts. The Digital Library contains videos of exemplary teaching of similar topics in a variety of classroom settings. Sara reviews those that are closest to her lesson plan and her school—an urban school system in a low SES neighborhood. Once again the Lesson Builder comes to her aid using the constructed lesson plan, her classroom demographics, her profile, and the student profiles to give a good match to an exemplary teacher of similar style, teaching a class on a similar topic, in a similar environment. It provides a contextual localization unavailable from previous systems providing exemplary teaching libraries.

None of these examples involve a full natural language understanding system. They do require the construction of associations that can be accomplished through existing link structures (Web or Digital Library) or through information clustering (using LSA or other text clustering techniques). Also critically important in construction such systems is an understanding of their usability. The systems should be unobtrusive and a natural aid to the user. This requires a full understanding of the task domain and a full understanding of the process of knowledge management in that domain. The presentation of information is not discussed, but clearly visualization and summarization of information must be supported and it must be as transparent as possible and as easy to use as possible since the above tools are in support of higher level learning and interaction and mustn't get in the way of the higher level activities.

Progress in the IIS research areas to support these learning environments include all of the areas mentioned in the first section and the following:

- Advances in LSA and associated text clustering algorithms
- Advances in automated knowledge linking algorithms
- Advances in summarization technology

The far out research is in the automated understanding of task domains and thus in the automated construction of appropriate visualizations and summarizations.

Intelligent information management

Intelligent information management is at the heart of IIS. It permeates much of group learning (such as in the Project Coordinator example above) and drives much of the industrial interest in IIS. We illustrate with a scenario of such use in learning environments—in this case learning on demand—but we could present many other applications.

Scenario—experience repository

“We got another system install gone bad,” said Kit to Rob.

“Yeah,” replied Rob, “this is the third mess we’ve been asked to clean up this year. When is management going to wise up and hire some competent installers to go with their sales staff.”

Rob and Kit proceeded to troubleshoot the problem. This was an installation of a new generation switching system using all optical switches, petabit/sec data communications, and a novel routing algorithm using anticipatory caching at company remote sites to improve network response time. Neither Rob nor Kit had worked with this particular system—the switches, the interfaces to the communications network, the trouble shooting software, inter-connections, and algorithms were all new to Rob and Kit. Rob and Kit were trained senior customer support engineers for the mega communications giant CNL formed from mergers of three troubled telecommunications equipment firms. They each were the head of a group of engineers—Rob’s group was responsible for customer support on switches and computers and Kit’s for customer support on network and other system software. During this merger, the information systems of all three companies were integrated into a single information system using an intelligent federated model.

Kit brought up the high level description of this particular installation and proceeded to run the logs of the system’s activity during the previous 24 hours against a visualization of the high level description. He observed a visualization of the customer’s complaint—there was an excessive overall network delay. He asked the system to find bottlenecks.

“Rob, look here. We have significant network activity at the New York and the Chicago sites, but much less at the New Jersey and Los Angeles sites. That’s strange since our records for this customer’s previous system indicated about the same level of activity at all 4 locations. Could you do a more detailed check on the engineering work at those four sites?”

"Sure Kit. Any idea of what we're looking for?"

"I'm not sure. What I'd like to do is to see if the experience repository has information on similar problems in the past. The repository doesn't have information on this configuration, but it includes all the experiences of all of the companies on earlier generations of this system—maybe something will jump out at us."

Kit then sets up a data mining process. He parameterizes the pattern that is being visualized from equal activity to anomalous activity—this is easily done with a visualization capture tool, and then queries the experience database to show similar patterns of activity from previous trouble shoots from the historical experience repositories. Sure enough, a number of candidates appear from all three of the companies. Unfortunately, it turns out that none of these scenarios provide information on the existing problem which turns out, after further analysis, to be due to incompatibilities in legacy systems. This entire trouble shoot process is automatically captured and stored in the experience repository.

The above scenario doesn't seem to be about education and learning yet it is since much of real life work of a knowledge worker is learning using knowledge tools. The learning that Kit and Rob engage in is sometimes referred to as learning on demand or just in time learning. They did not need to learn about legacy system failures of a certain type until a failure of the existing system occurred. Then they just needed knowledge of similar failures, not the entire history of all failures.

The tools needed to support knowledge workers are intelligent information management tools. They need to work with data bases that incorporate process, history, patterns, visualizations, summarizations, and other forms of data related to knowledge and its acquisition. These tools apply not only to experience repositories of the type described above, but also to applications suitably adapted to K-12 education (as described in the first scenario) or university education. Since such tools have a clear economic payoff, they will most likely be first developed for the workplace and then migrate to education and government applications—though perhaps early versions will first appear in military applications.

Examples of research in IIS needed to support these applications include:

- Research in process and design in all of its forms so it can be captured in repositories
- Research in visualizations of process information
- Research in effective indexing and search on repositories

The far out research is once again in automating the process of building the repositories in any knowledge domain.

We have discussed several IIS research areas and how they apply to learning and education. We feel strongly that the application area of learning and education is both challenging and extremely rewarding. We are beginning to see applications of IIS that have an impact on schooling, but there is substantial opportunity to do more.

Discussion of papers

In this issue are the two papers that were accepted from a call for papers on IIS and Education.

The paper by Zampa and Lemaire describes an application of LSA to the modeling of students in a tutoring system. The system uses cognitive theories that have shown that LSA, viewed as language learning in a high dimensional space, mimics human learning. Thus modeling both the students and the texts that make up the knowledge domain in the same high dimensional space is logical. The degree of student understanding of a concept can now be modeled as the distance between the student's vector and the knowledge domain's vector and an automated tutor can use the distance information in strategies of presenting information to the student to facilitate learning.

The paper by Newby explores the use of 3D visualization in understanding and learning information spaces. He reviews three decades of information visualization, identifying research trends, and runs an experiment comparing 3D visualization to 2D visualization. The results of the experiments are mixed and lead Newby to propose extensions of the experiments.

References

Anderson, J.R., Matessa, M., and Lebiere, C. (1997). ACT-R: A Theory of Higher Level Cognition and Its Relation to Visual Attention. *Human Computer Interaction*, 12, 439–462.

Brin, S. and Page, L. (1998). The Anatomy of a Large-Scale Hypertextual Web Search Engine. In *Proceedings of 7th World Wide Web Conference*, Brisbane.

Cassell, J. (2001). Embodied Conversational Agents: Representation and Intelligence in User Interface. *AI Magazine*, 22(4).

Foltz, P.W., Kintsch, W., and Landauer, T.K. (1998). The Measurement of Textual Coherence with Latent Semantic Analysis. *Discourse Processes*, 25, 285–307.

Kleinberg, J. (1999). Authoritative Sources in a Hyperlinked Environment, *Journal of the ACM*, 46(5), 604–632.

Landauer, T.K., Foltz, P.W., and Laham, D. (1998). Introduction to Latent Semantic Analysis. *Discourse Processes*, 25, 259–284.

Roschelle, J.M., Pea, R.D., Hoadley, C.M., Gordin, D.N., and Means, B.M. (2000). Changing How and What Children Learn in School with Computer-Based Technologies. *Children and Computer Technology*, 10(3), 76–101.

Schank, R.C. (2001). The Computer isn't the Medium, It's the Message. *Communications of the ACM*, 44(3), 142–143.

Thuraisingham, B. (2001). *Managing and Mining Multimedia Databases.* Boca Raton, FL: CRC Press.