

## **Knowledge Extraction for Educational Planning**

Christopher Eliot

Beverly Woolf

Victor Lessor

*Department of Computer Science*

*University of Massachusetts*

*Amherst, Mass 01002*

*eliot@cs.umass.edu*

**Abstract.** Digital libraries promise to provide rapid access to immense knowledge resources, but the quality and organization will vary. We will implement and evaluate software to construct high quality courseware using resources from a Digital Library. We propose to use context sensitive planning methods to create educational plans using digital library resources. These plans will be customized for individual students, using a student model, and optimized, using machine learning methods to evaluate the learning value of digital resources.

### **1. Educational Use of Digital Libraries**

Digital Libraries have organized thousands of instructional objects that vary from intelligent tutors [Woolf & Hall, 1995], to online lectures and papers, Table 1. More than 27,000 college-level courses were delivered over the Internet and more than 1.6 million students enrolled in a distance education course in 1997-1998 [Boettcher, 2000]. As these numbers increase serious problems of efficiency will develop unless novel mechanisms are implemented to manage the resources and interactions. Search engines can find these resources but no technology exists to assist students in organizing and scheduling online learning.

We are developing a system to select and customize educational resources from a Digital Library in order to support the needs of individual students and teachers. Digital resources will be assembled into courses designed to enable students to achieve specific learning goals, subject to constraints on pedagogy, quality, cost or duration. The system will also support teachers in the development of curriculum plans. The system will exploit a combination of manual and automatic evaluation mechanisms to optimize the course plans.

This research will address the following questions:

- How much planning knowledge can be extracted from an existing digital library?
- Can software automatically determine the content and approach of a teaching resource?
- What evaluation data is useful for optimizing course assembly?
- Can automatically constructed course plans compare with human authored plans?

### **2. State of the Art**

Conventional search engines do not organize results. The number of specialized collections is growing so rapidly that manually created directories (e.g. Yahoo) cannot keep up. Many high-quality sites are extremely specialized: searchable indices to the complete works of Darwin, nutritional information for foods, historical hydrological information in Western Washington, etc.

### 3. Proposed Approach

The core of our approach is the use of context sensitive planning methods to develop student specific courses optimized on the basis of feedback from prior student usage. Machine learning and statistical methods will be used to evaluate learning resources. Evaluation data will help refine, prioritize and schedule the selected resources. Information retrieval methods will be used to construct formal representations of available learning resources.

Resources will be managed by three conceptual modules: *Course Assembly Agents* will be responsible for resource selection and planning; and *Evaluation Agents* will monitor instruction and assist with course optimization and *Ontology Agents* will extract planning knowledge from the Digital Library

Table 1. Existing Instructional Repositories
<p><u>COURSES:</u>            E-College, <a href="http://www.ecollege.com">www.ecollege.com</a>, thousands of courses, one hundred degree programs            California Virtual University, <a href="http://www.cvc.edu">www.cvc.edu</a>, 1569 courses.            Western Governor's University, <a href="http://www.wgu.edu">www.wgu.edu</a>, 275 courses.            Southern Regional Education, <a href="http://www.-srec.sreb.org">www.-srec.sreb.org</a>, 300 courses,</p>
<p><u>OBJECTS:</u>            Educational Object Economy, <a href="http://www.eoe.org">www.eoe.org</a>, 2600 learning objects            NEEDS engineering database, <a href="http://www.needs.org">www.needs.org</a>, 863 Modules</p>
<p><u>LIBRARIES:</u>            Chemistry, <a href="http://www.chem.ucla.edu/chempointers.html">www.chem.ucla.edu/chempointers.html</a>            Mathematics, <a href="http://www.forum.swarthmore.edu">www.forum.swarthmore.edu</a></p>
<p><u>DATABASES</u>            NASA, <a href="http://www.nasa.gov/gallery/index.html">www.nasa.gov/gallery/index.html</a>            Human Genome, <a href="http://www.ncbi.nlm.nih.gov/genemap99">www.ncbi.nlm.nih.gov/genemap99</a></p>
<p><u>CLEARING HOUSES, PORTALS, CHANNELS</u>            American Distance Education Consortium, <a href="http://www.deal.unl.edu">www.deal.unl.edu</a>            The Gateway to Educational Materials, <a href="http://www.thegateway.org">www.thegateway.org</a>,            Ask-ERIC, <a href="http://www.askeric.org">www.askeric.org</a>            Advanced Distributed Learning <a href="http://www.adlnet.org">www.adlnet.org</a></p>

#### 3.1 Course Assembly Agents

A student<sup>1</sup> presents a query and the system determines which educational resource, or combination of resources are relevant. *Course Assembly Agents* will use information about the student, resources and topics, to construct an optimized course. When a consistent and complete plan is found, it will be presented to the student for approval. Further search may yield a more optimal plan, by exploring more options or as a result of unexpected changes in the availability or quality of learning resources.

The course planner will use well known planning techniques. Each learning resource will be described using an IF-THEN rule<sup>2</sup>. In Figure 1, RULE-1 describes a Calculus course and RULE-2 describes a physics course. (In practice, the course assembly system will use many rules that describe learning resources much smaller than a course.) As shown, a goal to know about forces can be achieved by taking a course on Calculus followed by a course on Physics, assuming the student already knows Algebra. A student who didn't already know Algebra would first have to take that course, while a student who already knows Calculus only needs the Physics course. Our system will create curriculum plans by finding a chain of planning rules that connect the intended final learning goal with the student's prior knowledge.

After planning has selected learning resources and partially determined their order, a specific "best" solution will be extracted, using TÆMS, a heuristic scheduling system [Wagner et al., 1999]. TÆMS produces a comprehensive linear instantiation of one possible solution to the problem, based on constraints, such as preferred time, quality or cost of the teaching materials. TÆMS will generate an initial solution and then enable the student to alter the parameters to retrieve a second solution. Given the constraints supplied by the student, TÆMS will offer a variety of solutions. We believe it is important to present plans with all variables fully instantiated since many people have difficulty reasoning about abstractions.

<sup>1</sup> Teachers and instructional designers, will develop courses for entire classes and job trainers might use the system to develop personalized courses for employees. For simplicity we call all of these uses *student queries*.

<sup>2</sup> This paper describes IF-THEN rules for simplicity; in practice, more general planning operators will be supported.

### 3.2 Evaluation Agents

Once a specific plan has been approved and scheduled, the student will be given access to the selected resources, one at a time, interacting via a browser or other interface. As the student finishes using each resource, the student model is modified to track learning progress, until the student reaches the intended learning goal, or abandons it as a goal. Difficulties encountered along the way are handled by re-planning.

The student model will record the completion of each learning resource and any grades given by the resource. Different resources may potentially (claim to) teach the same topic. Evaluation must determine the relative effectiveness of different educational resources.

A key observation is that the planning knowledge defines producer consumer relationships among learning resources. Resources that assume students have satisfied a precondition are consumers of the students who are produced by resources teaching the prerequisite knowledge. Performance of students learning from an advanced educational resource can be correlated with the resources used to satisfy the preconditions. If prerequisite knowledge (as measured by producers) does not correlate with success in a consumer resource (e.g., time spent and mastery) then the precondition can be dropped. If the average performance of students from one producer is different than the performance of students from another producer, then the two producer resources can be given a relative evaluation.

### 3.3 Ontology Agents

The *Ontology Agent* is designed to extract planning knowledge from existing repositories using information retrieval methods. Many sources of data can be searched to obtain planning information, including university course catalogs, online textbooks and technical ontologies. Specialized *Ontology Agents* will be developed for specialized sources of information.

This knowledge will be extracted in two parts. First, the terms in the prerequisites and postconditions will be isolated and organized into a collection of ontologies. Second, the specific descriptions of learning resources will be created, using the ontologies as a target language. Planning rules are abstracted from the descriptions of learning resources.

The terms used as preconditions and postconditions can be found in existing taxonomy databases and other texts. The Web has over 30 taxonomies of teaching topics varying from stable and mature taxonomies, such as the Library of Congress classification (<http://lcweb.loc.gov>) and Dewey Decimal Classification (<http://www.oclc.org/oclc/fp/>) to homegrown and unstable taxonomies. Within these taxonomies, higher education topics are broken into sub-topics. For example, general chemistry includes chemistry of mixtures, solutions, reactions, titrations etc. In the case of new or evolving subjects, ontologies may be abstracted from structured documents. Figure 2 shows two ontologies, one derived from the Department of Biomedical Engineering at the USC<sup>3</sup> and the other from the Annals of Biomedical Engineering,<sup>4</sup> using a table of contents keywords and abstracts from the six most recent papers in the journal. These ontologies will be extracted and merged for use during course planning.

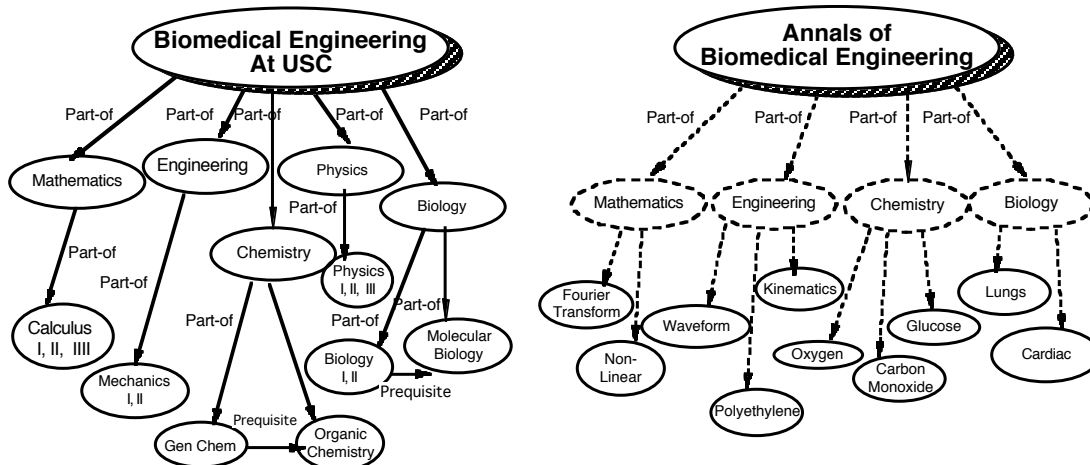
The use of multiple ontologies requires consideration of merging similar terms. We have explored this issue in terms of the ontologies implicitly defined by college course catalogs. We

RULE-1:	<b>IF</b>	Knows(Algebra) AND Passes(Calculus)
	<b>THEN</b>	Knows(Integration)
RULE-2:	<b>IF</b>	Knows(Integration) AND Passes(Physics)
	<b>THEN</b>	Knows(Forces)
Goal:		Knows(Forces)
Solution:		Passes(Calculus), Passes(Physics)

**Figure 1. A Trivial Course Planning Problem.**

<sup>3</sup> <http://www.usc.edu/dept/biomed/>

<sup>4</sup> <http://nsr.bioeng.washington.edu/ABME/annals.html>



**Figure 2. Two Biomedical Engineering Ontologies**

believe that enough information is available online so that these ontologies can be merged. Such information includes course descriptions, syllabus pages and textbook indexes. Information retrieval methods that measure the similarity of documents are used by Internet search engines and can be used to compare course descriptions. Courses that use the same textbook are probably similar. Different textbooks used by the same course are probably similar. Matching the precondition structure of departments from different universities will provide one method of detecting that two courses are similar. Another method for determining that course are similar is to inspect the subtopics of both courses: if the subtopics of one course are similar to another, then the courses are similar. Similarity of subtopics can be determined by examining the syllabus of a course or the table of contents of a textbook. Applying multiple indications of similarity will allow us to form equivalence classes among resources, such as courses, textbooks and other learning resources.

Once the terms used in document metadata are organized into ontologies, our system will convert resource descriptions into planning rules. In its simplest form, our curriculum planner requires IF-THEN rules that model the effects of learning resources. These rules indicate that IF a student knows some prerequisite concepts AND completes a specified learning activity THEN the student will obtain specified postcondition knowledge. For example, IF a student knows Calculus AND the student takes (and passes) Physics, THEN the student will know about Forces, Figure 1. The precondition and postconditions of learning resources will be extracted using information retrieval methods to search meta data tags and other descriptions of the resources.

Planning knowledge will be formalized using terms from the ontology, by analyzing meta data descriptions of learning resources. We analyzed the University of Massachusetts online catalog and extracted planning information from the listing of mathematics courses, Figure 3. Online catalogs from other universities are structurally similar. Established natural language and information retrieval techniques can transform the descriptions from English into formal computer terms. Developers of educational material will not be required to provide complex formal representations of their work but will only have to provide natural language descriptions of each resource. Since educators are comfortable writing course descriptions, we believe they will be able to write similar descriptions of online learning resources.

These automated techniques are primarily intended as a temporary mechanism for quickly incorporating knowledge from existing repositories. Eventually, we expect ontologies will be standardized and authors will use tools that automatically ensure ontological consistency of document metadata.

**131--Calculus I (R2) 4 credits**

Elementary techniques of differentiation and integration of algebraic and trigonometric functions, elementary differential equations. Applications physics, chemistry, and engineering. Students will use calculators on homework and exams. *Prerequisites: high school algebra, plane geometry, trigonometry, and analytic geometry.*

**IF** Knows (High-school-algebra) AND  
Knows (Plane-geometry) AND  
Knows (Trigonometry) AND  
Knows (Analytic-geometry) AND  
Passes (MATH-131)

**THEN** Knows (Differentiation) AND  
Knows (Integration of functions) AND  
Knows (Elementary-differential-equations)

**Figure 3. Formalized Course Catalog Entry**

Much or all of this information is already required by existing meta data standards, such as the Dublin Core<sup>5</sup>, which require compliant learning resources to contain a descriptive tag. Digital Libraries make similar requirements. We will search existing meta data tags to locate descriptions that can be translated into planning knowledge. We will encourage authors to explicitly provide the preconditions and postconditions of all learning resources, so reliance on heuristic methods can decrease.

The long-range goal is to develop methods that can maintain the ontology knowledge autonomously when needed or with human supervision when desired. Automatic ontology maintenance involves matching subtopics and determining their relationships. Eventually, complete libraries will be searched and analyzed to obtain ontologies and planning knowledge.

#### 4. Summary and Claims

Planning methods will be used to combine online educational resources into high quality courseware. Statistical methods will be used to determine which resources best prepare students for learning more advanced topics; the planner will favor resources which have been most effective in the past. Structured and unstructured sources of information will be used to automatically construct ontologies and planning representations of available resources.

The epistemological theory behind our representation includes three significant hypotheses:

- All planning terms appear as preconditions or postconditions;
- Preconditions apply to *learning* rather than *knowing*; and
- Course planning knowledge does not include negative postconditions.

1) All course planning terms are preconditions or postconditions of some learning resource. An ontology is complete if all preconditions and postconditions can be represented. The planner requires no other kinds of terms and some complex mechanisms can be avoided as a result.

2) Ontology items have no precondition links. However, it is possible to, compute the percentage of resources that require a specific precondition before teaching about an ontology item.

3) Nothing that we teach causes a student not to know something. Course planning requires no negative postconditions, hence no backtracking. There are two exceptions. First, certain

---

<sup>5</sup> <http://purl.oclc.org/docs/core/documents/rec-dces-19990702.htm>

demonstrations are most effective when they are surprising to students, but these students already have the knowledge. Second, lessons like laboratory experiments involve manipulating real world objects, such as chemicals. We exclude laboratory experiments from the domain of course planning, since this would require general reasoning about everything in the universe.

We will develop methods for reasoning and planning with open knowledge bases of Digital Library resources. One goal is to use of evaluation methods to guide the selection and planning processes. The feedback cycle between planning and evaluation is designed to ensure a steady improvement in the quality of instruction provided by the system. In order to maximize the ability of our system to take advantage of existing resources, we have developed methods for automatically extracting ontologies and planning knowledge from English text such as course descriptions.

## **References**

- Woolf, B., and Hall, W. (1995) *Multimedia Pedagogues: Interactive Multimedia Systems for Teaching and Learning*, IEEE Computer, 28(5), pp. 74-80.
- Boettcher, J. (2000) *The state of distance education in the U.S.: Surprising realities*. *Syllabus*, 13(7), pp. 36-40.
- Wagner, T. & Lesser, V. (1999) *Relating Quantified Motivations for Organizationally Situated Agents*, From *Intelligent Agents VI: Agent Theories, Architectures, and Languages*, April, Springer.