THE ORGANIZATION OF KNOWLEDGE SPACES FOR A VIRTUAL LEARNING ENVIRONMENT SUPPORTED BY A DIGITAL LIBRARY

Terence R. Smith, Olga Agapova, Michael Freeston, Alex Ushakov*
University of California, Santa Barbara, CA93106 USA
smithtr@cs.ucsb.edu, olga@geog.ucsb.edu, freeston@alexandria.ucsb.edu, ushakov@alexandria.ucsb.edu

* This paper represents the combined work of the whole of the ADEPT Knowledge Organization Team. Other members are: Olha Buchel, Jim Frew, Linda Hill, Richard Mayer, Jian Qin, Laura Smart, and Tim Tierney.

The Alexandria Digital Library Project is developing a concept-based learning environment for the sciences. In this paper, we briefly discuss: the rationale for the approach; the structure of the concept model and correlative relationships between concepts; the components and associated services of the concept-based learning environment; and planned and potential applications of the learning environment.

1. Introduction

In addition to services and collections that provide access to geospatially referenced information, the Alexandria Digital Library (ADL) Project is developing concept-based learning environments for the sciences. Such learning environments form an important component of the Alexandria Digital Earth Prototype [ADEPT 2002] and are integrated into the system's basic middleware search services [Janee and Frew, 2002]. The principal reasons for developing these digital library (DL) services and collections is to help students acquire a ``deeper'' understanding of science.

The approach of ADEPT to concept-based learning rests on premises that:

(1) scientific activity, understanding, and learning are based on a large core of scientific concepts and their interrelations which are used for representing scientific methods, abstract representational systems, and the phenomena under investigation.

(2) the syntactic and semantic information associated with any scientific concept may be represented in terms of a strongly-structured model (SSM) of the concept.

(3) a deeper understanding of any domain of science is promoted when learning environments are based explicitly on the use of SSMs to represent important sets of concepts and their interrelationships [Smith et. al. 2002a, 2002b.]

We have developed a concept-based learning environment that embodies such principles in the context of Web-based services. This environment has three main components and nine associated services (see figure 1.) The first component consists of one or more knowledge bases (KBs) containing SSMs of the concepts needed for representing the scientific knowledge about a given domain. In the next section we outline our current SSM of concepts. Three services associated with such KBs include:

(a) a service for the input and editing of SSMs of concepts;

(b) a service for searching the KB; and

(c) a service for graphic representations of the concepts and their interrelationships.

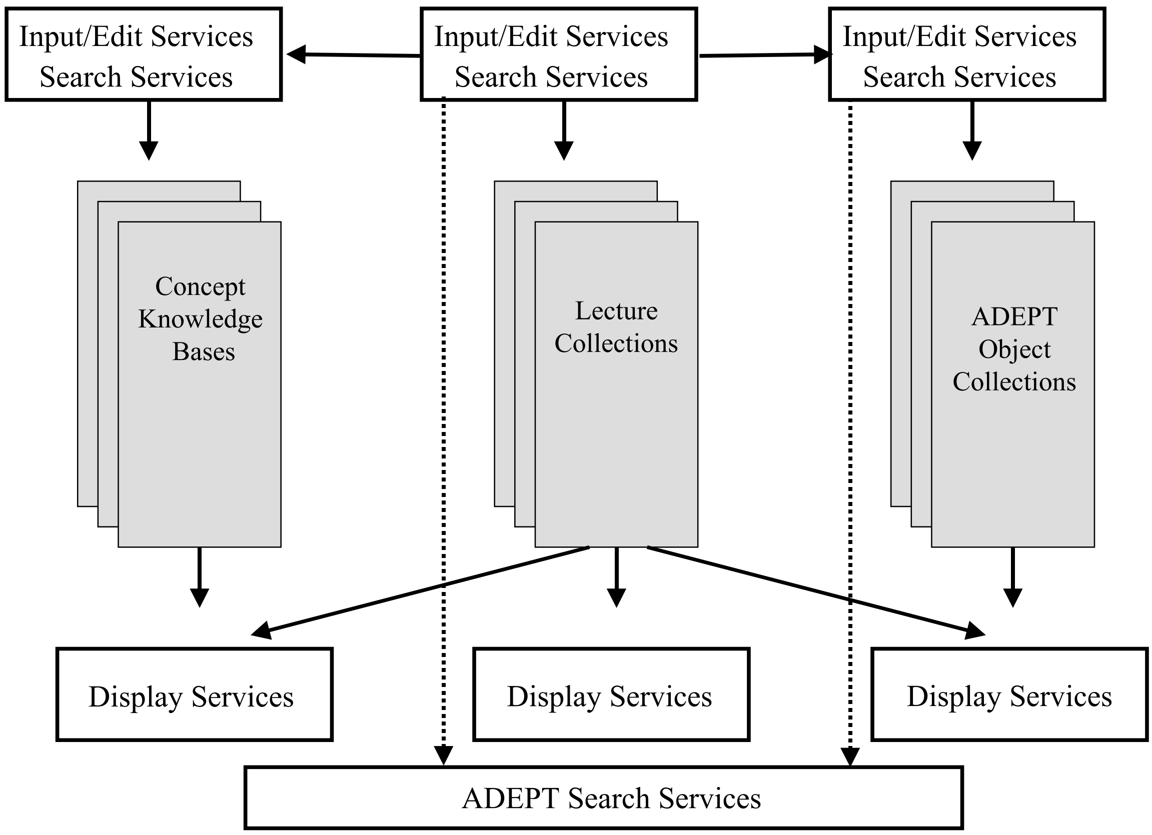

Figure 1: Components and associated services of a concept-based learning environment

The second component comprises one or more (heterogeneous) collections of items for use in illustrating elements of the SSMs of a given set of concepts. Three services associated with this collection include:

(a) a service for the input and editing of metadata records associated with such items;

(b) a service for searching the collections, particularly in terms of the scientific concepts to which they refer; and

(c) a service for graphical representations of the items.

The third component is a collection of lectures, self-learning modules, or laboratory sessions that provides access to the information in the KB and collection in a manner that supports efficient learning. Services associated with this collection include:

(a) graphic services for inputting and editing lectures, self-paced learning modules, or laboratory sessions;

(b) a service for searching the lecture collections; and

(c) a service for graphic displays of the contents of the collection.

In figure 2, we indicate how such components and services may be employed in supporting lecture-based learning environments. Other applications include self-paced learning and laboratory environments. In lecture-based learning, for example, an instructor may employ three distinct graphics displays for presenting concept-based learning materials to students:

1. a Knowledge Window for representing information about the SSMs of concepts in the KBs;
2. a Collection Window for displaying DL objects from ADEPT collections that illustrate various aspects of concepts;
3. a Lecture Window for displaying the lecture notes that organize the presentation of information about a topic and its associated concepts.

Figure 2: A three-window display of concept-based learning materials

We have applied these ideas in teaching introductory courses in Physical Geography and have constructed KBs, Object Collections, and Lecture Collections and implemented the three corresponding sets of services. We currently support the creation of lecture materials by an instructor with the various input, edit, search, and display services integrated into the Lecture Window. This permits the instructor to search the KBs, Object Collections, and Lecture Collections and to organize the results of such search for presentation in the Knowledge and Collection Windows during the course of the lecture.

The rest of the paper is structured as follows. We first describe the SSM of concepts that forms a foundation for our learning environments. We then describe the three services *input*, *search* and *display* for each of the three components *concept*, *collection* and *lecture* of the concept-based learning environment. Finally we discuss planned and potential applications of this learning environment.

2. Strongly-Structured Models of Concepts

We briefly discuss the strongly-structured concept model (SSM) and the services for creating, editing, searching, and displaying the KBs. A basic premise, noted above, is that concepts and their interrelationships are the fundamental building blocks for representing the phenomena of science. The value of any scientific concept rests heavily on the degree to which it may be represented in a manner that: 1. is objectively-communicable; 2. supports the the derivation of scientifically-useful information; and 3. possesses a well-defined (operational) semantics. Such attributes permit scientists to communicate unambiguously about how to represent, use, and interpret scientific concepts and to derive useful information from representations of phenomena that are based on the concepts.

Such requirements suggest the feasibility of constructing ``strongly structured models'' (SSMs) of scientific concepts that incorporate the scientific semantics of a concept. For example, one may view the semantics of a concept relating to the measurable properties of some phenomenon (e.g. the Velocity or Depth of a Water Flow) in terms of objectively communicable and reproducible procedures for the measurement of such attributes. We note that this idea goes well beyond the typically thesauri-like definitions of concepts that have traditionally been used in library environments. Analogous ideas have been recognized and employed, albeit implicitly, for some time by various scientific groups including, for example:

1. The MatML Working Group of the National Institute of Standards and Technology (NIST), which constructed a model for representing concepts relating to substances of Materials Science and their properties [NIST 2001]. In particular, this group has created an abstract DTD for representing such concepts as XML records.

2. The Chemical Abstracts Service [CAS 2001] Registry [Weisgerber, 1997], associated with the American Chemical Society, has developed an SSM

of chemical substances that includes systematic chemical names, their various representations (molecular structure diagrams, molecular formula, index of ring systems), and information about the properties and interrelations of substances.

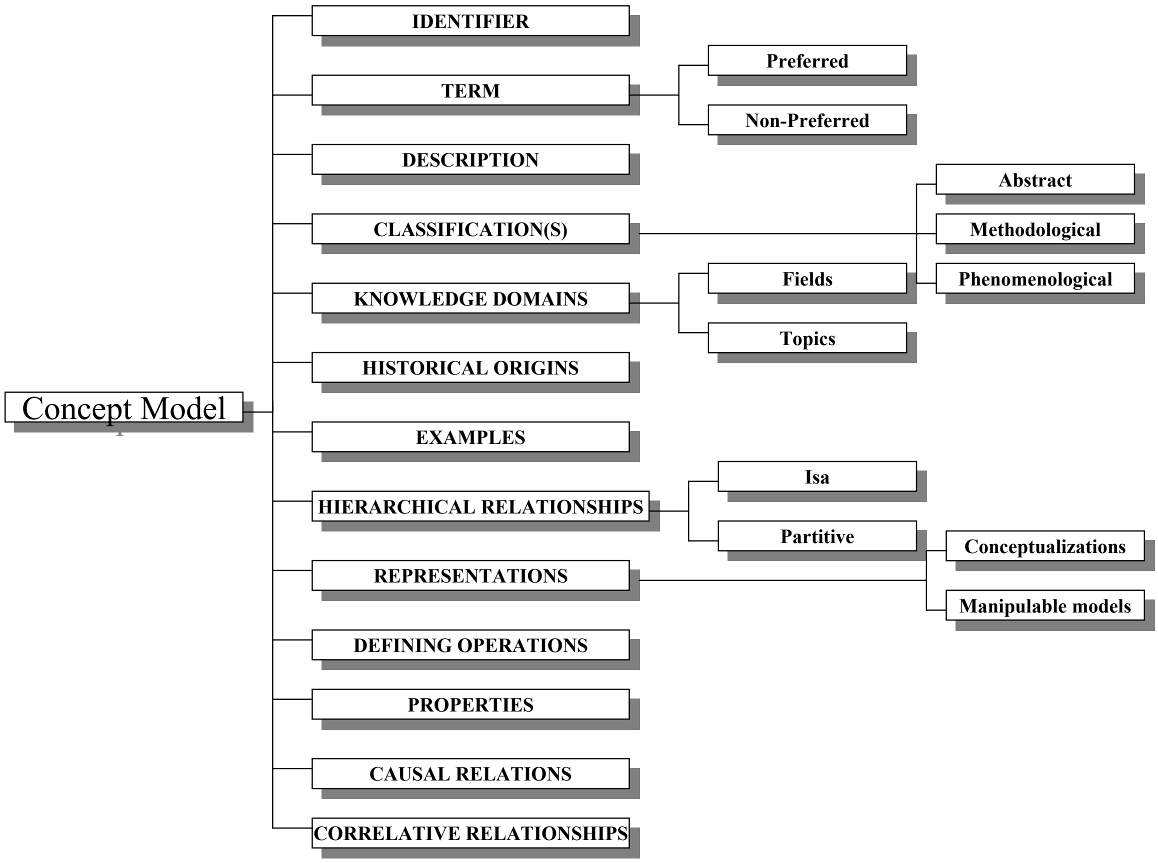

Figure 3: Abstract model (SSM) of a concept and correlative relationships

Based on such examples of SSMs and previous work, ADEPT has developed an SSM of concepts for scientific domains in terms of a frame-based knowledge representation system with slots and attribute-value fillers. Since the SSM has been described in a previous document [Smith et. al. 2002a, 2002b] we provide only a summary outline at this point. Our current SSM of a scientific concept takes the form shown in figure 3.

We comment briefly on the various elements of this model.

1. DESCRIPTION and HISTORICAL ORIGINS are natural language representations of the concept. While they indicate the scientific semantics associated with the concept, this semantics is typically imprecise and ambiguous.

2. CLASSIFICATIONS include, for example an ADEPT classification of concepts into abstract, methodological, and phenomenological (or concrete) concepts [Smith et.al., 2002b] as well as various subclasses, while KNOWL-

EDGE DOMAINS list (some of) the academic FIELDS and TOPICS in which the concept is used.

3. TERMS are simple (linguistic) expressions denoting the concept, while the HIERARCHICAL RELATIONSHIPS are specific, thesaurus-like relations between terms. While such terms provide bases for inference (for example: a ISA b AND b ISA c IMPLIES a ISA c) these terms and relationships are essentially without any scientific semantics.

4. REPRESENTATIONS, DEFINING OPERATIONS, PROPERTIES and CAUSAL RELATIONS, together with the HIERARCHICAL RELATIONSHIPS, relate to relationships among concepts. The REPRESENTATIONS, DEFINING OPERATIONS, PROPERTIES and CAUSAL RELATIONS, however, constitute the heart of the concept model by providing the scientific semantics of a concept. The Defining Operations provide scientifically objective descriptions of activities that define the concept. For example: abstract concepts are defined in terms of syntactic manipulations that are permitted on various representations of such a concept (e.g. as in the manipulation of Algebraic Equations); methodological concepts may be described in terms of recipes for activities carried out by machines and scientists, such as measurement activities; phenomenological concepts may be defined using phenomenological interpretations of the TERM representation of a concept. PROPERTIES (such as the Area property of the concept of a Polygon) are represented not only using the TERMS denoting the property, but also the activity for computing the property from a given representation.

3. Concept input, search, and display services

A knowledge base (KB) of concepts for a given domain is defined in terms of a collection of SSMs for a set of concepts. We have developed a form-based input tool to support the construction of KBs of concepts. The input tool is structured to reflect the abstract model of a concept presented above. Parts of the input process may be automated, as in the case of using authoritative glossaries in electronic form to provide input of natural language DESCRIPTIONS of the concepts. Other elements of the SSM for a concept, however, must currently be entered by knowledge domain specialists.

The current operational version of the learning environment employs RDBMS technology (and MySQL in particular) for storing and accessing the collections of SSMs of concepts. The concept input form is based on this instantiation. We have also developed support for XML databases of concept SSMs using an XML schema to represent the abstract concept model. We plan to convert from relational to semi-structured (XML) versions in the near future.

We have developed visualization tools for concepts and their interrelationships. A current visualization tool, SPROING [Ancona, 2002], supports the visualization of concept maps of subsets of concepts and their interrelationships that may be derived from the KB. This tool employs an MsWindow general purpose C++ application that perform OpenGL-based visualization with a simple binary force springs algorithm. The data for visualization are currently provided by a stand-alone extraction routine that runs over the KB and extracts peer-to-peer pairs of relationships. These are represented as a pair of nodes connected with a an edge. The colors and styles for the presentation are stored in a parameter file in XML format. A simple search function is implemented on the presentation space. Nodes and edges, found by a given search, may then be stored as a visualization subset, which is essentially a substructure of the presentational graph. Extensions of SPROING support the hand-editing of the graphic outputs for the purpose of in class presentation as well as the visualization of such aspects of concepts as DESCRIPTIONS and Mathematical Representations.

4. Collection input, search, and display services

We may associate with any domain-specific KB one or more collections of DL Objects that illustrate various aspects of the concepts modeled in the KB. In the case of a collection of Objects that illustrate the concepts in Hydrology, for example, we may incorporate figures, photographs, or natural language instructions describing measurement techniques for determining the rate at which groundwater flows through some Aquifer. Such collections generally consist of heterogeneous and multimedia Objects, ranging from text to video with soundtrack.

A requirement of such collections is that they be searchable, not only in terms of the usual criteria of spatio-temporal footprint, subject matter, etc., but also in terms of the concepts and concept models represented in a given KB. Hence, for example, an instructor or student may search for all image items that explicitly contain information relating to the concepts of Floodwater Waves from Rivers with a footprint in the Central Valley of California.

Access to such collections within the ADEPT environment and their items is supported by:

1. metadata descriptions of each Object that contain explicit references to the domain-based concepts to which they refer;
2. standard ADEPT middleware search services [Janee and Frew, 2002];
3. ADEPT collection-level metadata for each collection.

We may access the Objects themselves using pointers obtained from the metadata records. Each metadata record takes the form prescribed by the ADEPT/DLESE/NASA (ADN) metadata content standard [ADEPT/DELESE 2001] with an extension defined in terms of an additional element containing TERM representations of a concept. Current collections are implemented as relational databases. This permits the easy construction of ADEPT search buckets using ADEPT's generic Bucket-99 driver.

A web-based metadata entry form has been constructed to support the entry of item metadata records by non-technical users. This entry tool is currently implemented as a series of PHP-driven input forms, developed to accommodate metadata entry into the respective data fields.

The display component of the collection window is functionally controlled from the lecture window. It allows not only the visualization of items selected and linked by the instructor, but also real-time (i.e., lecture-time) search for ADEPT collection items through the use of an ADEPT Browse window.

As an example of one of our collections, we have cataloged and stored each of the more than 400 illustrations, available in electronic form, that come with the textbook that is being used to support a physical geography course that will be taught using the learning services.

5. Lecture input, search, and display Services

In terms of concept-based learning in science, we may characterize a course of instructional material as a trajectory through a space of concepts. This characterization is general, as almost any scientific lecture or scientific textbook will prove. Instructors as well as texts largely differ from each other in terms of the order in which concepts are introduced and the different emphases with which they are modeled and related. A major strength of a concept-based approach, therefore, is its support for the construction of flexible, personalized trajectories through concept spaces from the KBs and Object Collections. Another strength is that instructors may re-use/repurpose trajectories that have been constructed previously by other instructors.

We have developed a set of Lecture Window services that may be used to support the creation, and recreation, of structured lecture notes that integrate links to the KBs and Object collections. In particular, an instructor may employ the Lecture Window for:

1. creating (structured) lecture notes;
2. searching over the KBs and Object collections;
3. embedding links from the lecture notes to displays generated by the KB and collection search services;
4. driving displays on the Knowledge, Lecture, and Collection Windows during a presentation.

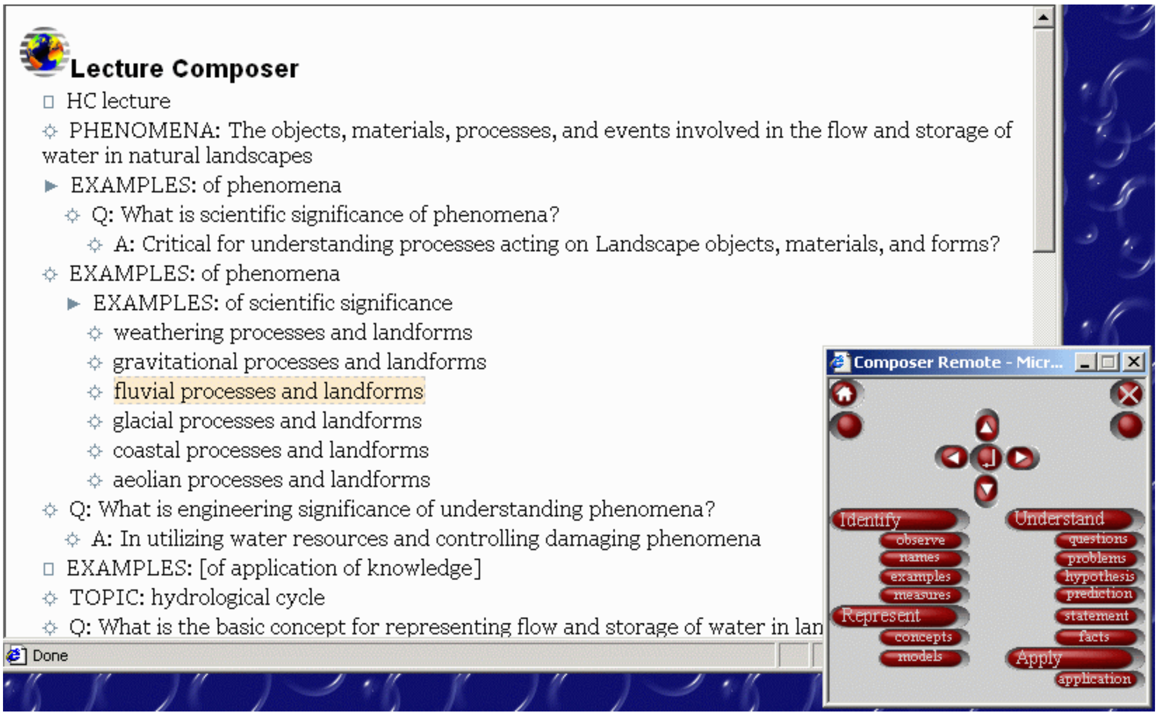

Figure 4: The lecture input and editing tool

We have developed a lecture input and editing tool that allows an instructor to create a series of lecture notes. At the most basic level, these tools allow an instructor to create lecture materials in terms of rich textual outline (figure 4). They support the usual creation and editing operations, including linking concept terms to illustrative materials from the KBs and Object Collections. The links attached to an item may be represented in terms of thumbnails or icons immediately following the highlighted item. We have additionally developed services that allow an instructor to create more personalized trajectories concepts through concept space with the use of various organizing subheads. Our current lecture creation service, for example, supports the quasi-automated structuring of lecture materials in terms not only of scientific concepts but also in terms of basic sets of scientific activities, such as:

1. identifying, observing, and characterizing phenomena;
2. representing phenomena; and
3. understanding phenomena.

The current lecture input tool allows the instructor to choose, with the help of a control panel, lecture note headings from a relatively small subset of terms that characterize each of these sets of activities. Under the category of representing phenomena, for example, we currently support:

1. the creation of subheads for TOPICS(S), CONCEPTS, and MODELS;
2. the use of TOPICS (and SUBTOPICS and SUBSUBTOPICS) subheads for the automated creation of high-level views of lecture materials;

3. the introduction of CONCEPT subheads for explicit description of SSMs of scientific concepts.

The high-level views of lecture materials are presented in a small window to provide context for both instructors and students in ways analogous to a collapsible folder tree. The results of lecture creation are saved in a Lecture Collection. The internal format of lectures follows the enriched Outline Processor Markup Language (OPML) format, which is essentially a subset of XML. Hence lectures are stored in the collection as XML records, supporting interoperability and persistence. The collection itself is implemented as an indexed file structure.

Future developments include putting the lecture collections into the format of Xindice XML databases to support input and editing inside a common browser (IE 5.5+ or NS6+ on a variety of platforms such as Win, Mac, Unix, etc.) Onscreen editing is accomplished on the client side with JavaScript DOM features, matched by Java servlet/JSP components on the server side.

6. Applications of the Learning Environment

The current ADEPT learning environment is being deployed in teaching an introductory course in physical geography to about 100 students in the Fall of 2002 and again in the Spring of 2003. Such applications of KBs of SSMs of concepts in learning environments build on recent studies in educational psychology (see Mayer, Smith, and Borgman, 2002.) These studies indicate that concept maps, defined in this context as ``all terms used in a lesson with labeled links among them'', play significant roles in promoting student learning and assessing student knowledge.

Educational research relating to rhetorical structures, or the ways in which materials can be organized into coherent structures for learning, suggests how the KBs and collections may be used in furthering student understanding of some domain of knowledge. In process rhetorical structures, for example, cause-and-effect systems can be represented by flow charts and maps. This corresponds to a view of the KB that extracts Causal and Hierarchical Partitive relations between a given set of concepts. Other rhetorical structures associated with different views of the KB include problem solution (problem based views), classification (hierarchical views), and compare (matrix-based views in which the attributes of different concepts are compared.) Recent research [Mayer et al, 2002] indicates that greater depths of learning occur when rhetorical structures are used to organize learning materials. We note that our services supporting access to, and use of, the resources in the KBs and associated collections enable the construction of views corresponding both to different rhetorical structures and the personal preferences of instructors.

7. Conclusion

A key issue in the development of ADEPT learning services based on SSMs of scientific concepts is an evaluation of the degree to which they deepen a student's understanding of both the conceptual basis of science and scientific knowledge in specific domains. We are currently instrumenting the teaching of courses in order to develop such evaluations. It will take time to assess the pedagogic value of using KBs of strongly-structured representation of scientific concepts, together with the associated collections and services, in learning environments. Nevertheless, recent developments in both text publishing and in educational psychology indicate their potential value.

Acknowledgments

The work described herein was partially supported by the NSF-DARPA-NASA Digital Libraries Initiative and the NSF NSDL initiative under NSF IR94-11330, NSF IIS-9817432, DUE-0121578, and UCAR S02-36644.

References

ADEPT. (2002) http://www.alexandria.ucsb.edu

ADEPT/DLESE. (2001) *ADN Joint Metadata Content Model*. http://www.dlese.org/Metadata/.

Ancona, D. (2002). *Visual Explorations for the Alexandria Digital Earth Prototype*. Second International Workshop on Visual Interfaces to Digital Libraries, JCDL, Portland, Oregon, 2002.

CAS. (2001). CAS *Chemical Abstracts Service Homepage*. http://www.cas.org.

Janée, G. and Frew, J., (2002). *The ADEPT Digital Library Architecture.* Proceedings of the Second ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL '02), Portland, OR, July 14-18, 2002.

Mayer, R.E. (1991). *The Promise of Educational Psychology: Learning in the Content Areas*. Merrill Prentice Hall, Upper Saddle River, N.J.

Mayer, R., Smith, T.R., and C.L.Borgman, (2002). *Digital Libraries as Instructional Aids for Knowledge Construction.* Educational Technology.

NIST. (2001). *MatML: eXtensible Markup Language for Materials Property Data*. MatML DTD Version 2.0.,2001. Prepared by E.F. Begley on behalf of the MatML Working Group.

http://www.ceramics.nist.gov/matml/matml.htm

Smith, T.R. Zeng, M.L., and ADEPT Knowledge Team, (2002a). *Structured models of scientific concepts as a basis for organizing, accessing, and using learning materials*, UCSB Department of CS Technical Report 2002-04.

Smith, T.R., Zeng, M.L., and ADEPT Knowledge Team, (2002b). *Structured models of scientific concepts for organizing learning materials*, Proceedings of the Seventh International Conference on Knowledge Organization, Granada, Spain.

Weisgerber,D.W., (1997). Chemical Abstracts Service Chemical Registry System: history, scope, and impacts. *Journal of the American Society for Information Science,* 148(4):349-360.