Structured models of scientific concepts as a basis for organizing, accessing, and using learning materials

Terence R. Smith
Department of Computer Science, UCSB
Santa Barbara, CA 93105
smithtr@cs.ucsb.edu

Marcia L. Zeng
SLIS, Kent State University
Kent, Ohio
mzeng@kent.edu

ADEPT Knowledge Organization Team *
Alexandria Digital Library Project, UCSB
Santa Barbara, CA 93106

ABSTRACT

We describe a DL-based environment designed to support learning in domains of scientific knowledge with the use of: (1) a knowledge base (KB) of strongly-structured concept representations [1]; (2) a DL collection of items exemplifying facets of each concept; and (3) a set of services. These services include editing the KB and collection as well as providing various textual and graphic views and supporting their use. We describe the implementation of this environment as an extension of the Alexandria Digital Earth Prototype (ADEPT). We describe the *strongly-structured model for concepts* that underlies the KB and collection metadata, and discuss related models of concepts developed in various scientific disciplines. We specify the architecture of the system and present examples of its use, such as the creation of personalized collections of learning materials that take the form of trajectories through the KB of concepts. The ADEPT learning environment will be used and its efficacy evaluated in an introductory undergraduate course in physical geography being taught in the Fall of 2002.

Categories and Subject Descriptors

H.4.m [**Information Systems**]: Miscellaneous

General Terms

Strongly-structured models of scientific concepts, knowledge representation systems

1. INTRODUCTION

A strength of digital library (DL) technology is its ability to incorporate services supporting content-based organization of, access to, and use of its collections in learning applications. NSF's long-term program for constructing a national DL supporting education in the areas of science, engineering, mathematics, and technology (NSDL [12]), exemplifies this promise.

A significant portion of DL development activities that support learning, however, is focused on providing access to scientific information in terms of traditional "information containers", such as electronic versions of books, journal articles, images, and videos. This approach is an electronic extension of traditional library (cataloging) practises, with metadata schemas supporting retrieval focused on information containers rather than on information content. Good reasons exist for employing this strategy because of the: (1) large heritage of traditional information containers; (2) need to provide electronic access to such learning materials in a timely manner; and (3) significant cost and difficulty in providing access to information by content.

For NSDL to provide near-term access to educationally useful collections, for example, one of its initial foci is developing a Core Integration System that supports access to distributed, and essentially independent, collections of items. Many of the items in such collections are typically electronic analogs of traditional information containers. A representation of every collection item in NSDL will be represented in Dublin Core format at a central metadata repository (and optionally in some native metadata format at the collection site [11].)

The idea of access to DL collections by content has been repeatedly raised, but no consensus has developed concerning appropriate levels of granularity for such access. Hence traditional containers remain the standard level of granularity in organizing knowledge. There is little evidence, however, that this level of granularity is the most effective mode for accessing library collections for learning purposes. The question therefore arises as to whether there exist well-defined and generally-acceptable levels of granularity for organizing,

*Members of the ADEPT Knowledge Organization Team:
Olga Agapova (olga@geog.ucsb.edu);
Olha Buchel (obuchel@hotmail.com);
Michael Freeston (freeston@alexandria.ucsb.edu);
Jim Frew (frew@icess.ucsb.edu);
Linda Hill (lhill@alexandria.ucsb.edu);
Laura Smart (lsmart@ucla.edu);
Tim Tierney (tierney@alexandria.ucsb.edu);
Alex Ushakov (ushakov@alexandria.ucsb.edu).

[1]By "strongly-structured" we mean that the model of concepts possesses a significant degree of structure, but **not** that it is an inflexible structure.

accessing, and using scientific knowledge that are implementationally and economically feasible.

The approach taken in this paper to answering this question is based on the premise that "scientific concepts" and "relationships between concepts" provide a powerful, and perhaps the only, level of granularity with which to support effective access and use for learning. The notion of concepts and their interrelationships is central to science. The sets of concepts and interrelationships developed over many centuries by scientists and mathematicians constitute nothing less than the fundamental building blocks from which useful representations of reality are created, applied, evaluated, and modified.

Understanding the knowledge in some domain of science requires students to understand how sets of concepts and their interrelationships are developed and applied in representing the phenomena of the domain. A large literature has developed in various fields, including psychology, education, and cognitive science, concerning the development, use, and evaluation of the conceptual bases for various domains of scientific knowledge. Most approaches to teaching and learning, however, deal with concepts in relatively implicit ways. For example, students are rarely presented with strongly-structured models of concepts (SSMCs) and their interrelationships. As a result, they too often emerge from a course of study with: (1) limited notions of the nature of concepts; (2) large, but poorly-structured, memorized sets of *terms* denoting concepts; (3) incompletely structured associations of relationships between concepts; and (4) partial knowledge of how to use concepts in creating representations of phenomena or how to create new concepts.

We believe it both feasible and valuable to organize, access, and use scientific knowledge explicitly in terms of the *sets of concepts that underlie some domain of scientific knowledge* rather than in terms of the containing information objects. As we show below, it is possible with current DL technology to develop:

- SSMCs for representing concepts and their interrelationships;
- domain-specific knowledge bases (KBs) of such representations;
- associated DL collections of "illustrative materials" concerning different aspects and attributes of the concepts;
- services supporting the creation, modification, viewing, and use of concepts for various purposes in learning contexts.

In particular, KBs of concepts and associated collections may be employed, together with associated services, in creating tailored courses of instruction that take the form of a "trajectory" through the space of concepts underlying some domain of knowledge. Not only do such KBs of concepts facilitate the creation of different ways of organizing knowledge about some domain in terms of its basis of concepts and relationships, but support a greater focus on critical aspects of concepts. In the case of the *measurable concepts* of science (see below), for example, they provide more explicit emphasis on fundamental aspects of a concept, such as its dimensions, measurement, representation, and interrelationships with other concepts.

We believe that such an explicitly concept-based support for learning leads to a deeper understanding on the part of students of the: (1) nature, structure, and classes of the concepts that, together with the interrelationships between the concepts, provide a basis for scientific development in specific domains of knowledge; (2) scientific roles of various classes of concepts (see section 4.1) across the spectrum of scientific activities; and (3) global structure of some domain of scientific knowledge in terms of the underlying framework of concepts. Advantages that may accrue to instructors include the efficient re-use and re-purposing of the KBs of concepts and the associated collections and services in creating instructional support materials. The recent interest in *concept and topic maps* [5] may well reflect the advantages of a concept-based approach.

In this paper, we outline the details of this approach as it is being implemented within the context of ADEPT [17]. We first discuss the importance of concepts and their interrelationships for learning and practising science, yet note that access to knowledge in terms of concepts is still largely based on *weakly-structured models of concepts*. We then develop the idea of SSMCs, noting some of the SSMCs that have been developed in various disciplines. We describe both a general SSMC, based in part on these examples, that is currently being used in the ADEPT Project, and the architecture of an extended ADEPT that will be employed in the Fall of 2002 at UCSB for teaching an introductory undergraduate course in physical geography. In addition to basic DL search services, this system includes a KB of strongly-structured concepts, associated collections, and services. We conclude with the possible applications of this system in support of learning.

2. CONCEPTS, SCIENCE, AND LEARNING

Scientific activities are based on using a principled, objective [2] approach to developing useful representations of observable phenomena. It is difficult to justify any viewpoint other than ones in which *concepts and their interrelationships provide the fundamental knowledge granules for constructing scientific representations* [3]. Even a limited knowledge of mathematics, for example, indicates the degree to which mathematical knowledge is based on the development and use of hierarchical structures of concepts and interrelationships.

Within the context of human knowledge in general, Binwal and Lalhmachhuana [6] define *"knowledge representation as a systematic way of codifying human knowledge"* and affirm that any knowledge representation system (KRS) requires an *ontological commitment* as to what set of concepts are to be used in representing some aspect of the world. They further argue that *"a central part of KR consists of elaborating: (1) a set of abstract objects; (2) concepts and other entities; and (3) the relations that may hold between them."*

[2] Our use of the term "objective" in this paper typically denotes that knowledge can be communicated unambiguously between scientists, allowing all scientists to obtain similar results when applying the knowledge.

[3] We note that the development of concepts and their representations is also a major focus of scientific activities.

In a related point of view Gärdenfors [14] argues that *"the central question for any theory of knowledge representation (KR) is how concepts should be modeled"*.

In relation to learning, there is a growing consensus that science education should be a meaningful activity in which students learn to think like scientists rather than solely remember information. This suggests strongly the importance of students developing a deep understanding of scientific concepts and their importance in the scientific approach, including their representation, creation, use, and evaluation. The *National Science Education Standards* [10], which makes frequent reference to concepts, calls for students to engage in scientific reasoning in which they "become familiar with modes of scientific inquiry, rules of evidence, ways of formulating questions, and ways of proposing explanations". The success of such activities depends on the depth of understanding of the concepts underlying the practice of science.

We may interpret three generally-accepted principles concerning students' acquisition of deep understanding of some domain of scientific knowledge in terms of a deep understanding of the basis of concepts underlying the domain of knowledge [15, 8, 7]. These principles, derived from cognitive science research into learning and education, are:

- **Domain-specific learning:** *Students learn best when cognitive skills are taught in the context of a specific domain of knowledge rather than in more general contexts.* An interpretation of this principle in terms of concepts suggests the importance of explicitly supporting students' acquisition of a deep understanding of the concepts, concept interrelationships, and reasoning skills related to developing and using scientific representations of phenomena in any domain of scientific knowledge.
- **Case-based learning:** *Students learn best when cognitive skills are learned in the process of solving authentic problems rather than when pieces of information are presented as isolated facts to be learned.* A concept-based interpretation suggests the importance of constructing learning environments in which specific problems are presented in ways to motivate the use of sets of concepts and interrelationships germane to the problems.
- **Scaffolded learning:** *Students learn best when the task difficulty is adjusted to meet their capabilities.* The interpretation suggests providing students with subsets of concepts and concept representations appropriate to their knowledge levels.

These principles and interpretations indicate the value of constructing learning environments based on, and making explicit use of, a KRS containing structured models of concepts and their interrelationships. Pedagogic advantages arising from such an approach include concept-based access to scientific knowledge and the efficient re-use and repurposing of KBs of structured concept representations.

2.1 Weakly-structured models of concepts

Library and DL technology currently employs a range of KRS that represent spaces of scientific concepts and their interrelationships to support the organization of, access to, and use of scientific knowledge. These include (see, for example, [4], [6]):

- **term lists**, such as authority files, glossaries, gazetteers, and dictionaries;
- **classification and categorization schemes**, such as subject headings, hierarchies, taxonomies, paradigms, faceted analysis, and categorization schemes;
- **relationship groups** such as thesauri, semantic networks, frames, ontologies, concept maps, and topic maps;
- **metadata content standards**, and particularly those parts dealing with knowledge representation;
- **general knowledge representation languages**, such as first order predicate calculus, description logics, conceptual graphs, general frame systems, and object-oriented modeling languages.

The first three of these KRS are the ones most generally used in such applications, in which they typically take the form of structured sets of terms representing concepts and their interrelationships as, for example, in the case of thesauri. Graphical representations of concepts and interrelationships derivable from such KRS (which we may term *concept maps*) typically take the simple form of a set of named nodes connected by named links.

A weakness of such applications of concept-based KRS within DL technology arises from *the relatively weak models of concepts* which they typically employ These representations almost invariably take the form of *linguistic terms*, with no semantics explicitly represented in the KRS, that denote concepts and their interrelationships. Such weakly-structured models of concepts are of little value in providing the "deep" organization of, and access to, scientific knowledge that is important for learning. Their importance is largely restricted to accessing knowledge in terms of the traditional information containers such as books and other documents, whose contents may, for example, be accessed by such term-based concept representations as *subject headings.*

3. STRONGLY-STRUCTURED MODELS

Such weaknesses in current applications of KRS in DL environments have been recognized in a variety of scientific disciplines. As a result, various disciplines have developed their own SSMCs, some of which we now briefly review. We first describe a general theory of concept representation developed by Gärdenfors [14], in the area of cognitive science, that we may interpret as an underlying theory for various SSMCs. We then present examples of SSMCs that may be interpreted as special cases of Gärdenfors' theory. The goal of this discussion is to motivate a general SSMC that we describe below.

3.1 Basis for concept models

Gärdenfors [4], in constructing a convincing theory of *cognitive* (or non-objective) representations of concepts, explicitly lays the ground for a convincing theory of *objective* (including scientific) representations of concepts and their interrelationships. He describes three interrelated classes of theories concerning the *cognitive* representation of concepts, namely:

[4] Gärdenfors accepts the notion that a "*a concept is an idea that characterizes a set of categories*" (see [14]) and that a category is a rule for classifying *objects*.

(1) low-level *connectionist/associationist/procedural* approaches; (2) high-level *symbolic* approaches; and (3) bridging-level *conceptual* approaches.

Gärdenfors gives priority to this third class of conceptual representations, for he develops a representational theory based on geometrical structures in well-defined *conceptual spaces*. This contrasts with representations based on either associations or symbols. His focus on the conceptual level reflects the well-known weaknesses of representations based on associations or symbols [5]. He implies, however, that all three classes combine in forming an integrated and powerful KRS.

In the context of the conceptual level, he defines a *natural concept* as being represented by "*a set of regions in a number of domains together with an assignment of salience weights to the domains and information about how the regions in different domains are correlated.*" A *domain* is a multidimensional space, whose *quality dimensions* may take the form of (sometimes measurable) single dimensions or integral dimensions comprised of several interrelated dimensions (such as the three dimensions of color representation), or combinations of these two types. Hence Gärdenfors is able to define: (1) *primitive concepts* as relating to a single dimension; (2) *properties* as concepts defined by an integral dimension; (3) *simple concepts* as defined by a single domain; (4) *complex concepts* as involving several domains; and (5) *objects* as points in a conceptual (representation) space.

Gärdenfors' main contribution to a theory of *objective* representations of concepts and interrelationships are apparent on noting that: (1) the three interrelated classes of theories of cognitive representations of concepts have natural and direct counterparts in theories of scientific (or objective) representation of concepts; (2) his representational theory of concepts at the conceptual level translates almost verbatim to the language of scientific representations; and (3) the intermediate class of SSMCs at the conceptual level are highly expressive, semantically-rich, and very useful in applications. Binwal and Lalhmachhuana's review of objective KRS supports this idea from a different perspective. They note that *structured object systems*, such as semantic nets and frames, provide bases for especially useful KRS characterized by attribute-value pairs containing both procedural and declarative representations. They lend themselves well to model-fitting approaches [6].

3.2 SSMCs in various domains of science

We briefly review three SSMCs developed for concept representation in the areas of materials science, chemistry, and geography.

3.2.1 SSMCs in Materials Science

The MatML Working Group of the National Institute of Standards and Technology (NIST) has constructed a standard for representing concepts about the materials studied within the knowledge domain of materials science [9]. In particular, they have created an abstract DTD for representing concepts in this domain as XML records. At the highest level, the basic KRS for representing materials concepts takes the form:

```
<!ELEMENT Material (BulkDetails,
                    ComponentDetails?,
                    Terms?,
                    Graphs?)          >
```

This statement specifies the basic content model for *Material*. In relation to its component elements:

- *BulkDetails:* contains a description of the bulk material.
- *ComponentDetails:* contains a description of the components that comprise the bulk material. ComponentDetails might be used, for example, to describe the base metal, the heat affected zone, and the weld metal of a welded material or the whiskers, fibers, and matrix of a fiber-reinforced composite material.
- *Terms:* contains descriptions of the material and property terms used in the document.
- *Graphs:* contains two-dimensional graphical images.

The content standard takes the form of a hierarchical structure. For example, BulkDetails has the following structure:

- BulkDetails
  - Name
  - Class
  - Subclass
  - Specification
  - Source
  - Form
  - Processing
    - Name, Parameters, Result, Notes
  - Geometry
    - Shape, Dimensions, Orientation, Notes
  - Characterization
    - Formula, ChemicalComposition, PhaseComposition, DimensionalDetails, Notes
  - Properties
    - PropertyDetails, Value, Qualifier, Parameters
  - Notes

We note here that ComponentDetails additionally contains:

- Associations
  - Associate, Relationship, Notes

3.2.2 SSMCs in Chemistry

Recording, storing, and retrieving information on chemical substances have been critical in the progress of chemistry [18]. The discipline of chemistry has developed an authoritative KRS for representing *concepts relating to chemical substances* [2]. Chemical compounds in terms of their syntheses, properties, and applications, lie at the core of chemistry. The Chemical Abstracts Service (CAS), a division of the American Chemical Society, produces the largest and most comprehensive databases of chemical information, including CHEMICAL ABSTRACTS (CA) and the REGISTRY. CA contains over 18 million document records from the chem-

[5] The semantics of concept representations based on the connectionist approach are encoded implicitly in terms of procedures; those of concept representations based on the symbolic approach are bound up with essentially syntactic operations on the symbols.

[6] They indicate a preference for frame-based KRSs.

ical journal and patent literature. The CAS REGISTRY, which contains records of more than 19 million organic and inorganic substances and more than 16 million sequences, represents a wide variety of *chemical substance concepts*, including the world's largest collection of concepts relating to such substance subclasses as: *Organic compounds, Inorganic compounds, Metals, Alloys, Minerals, Coordination compounds, Organometallics, Elements, Isotopes, Nuclear Particles, Proteins and Nucleic Acids, Polymers, Nonstructurable materials (UVCBs).*

These activities have led to the development of a structured model of chemical substances. CAS extended nomenclature principles of the International Union of Pure and Applied Chemistry (IUPAC) to develop unique names of any substance. Such CA index names are built around (a) a *parent*, which is generally the name of a structural skeleton plus a suffix denoting the principal functional group; (b) substituents on the parent structure; (c) possibly a *modification* which identifies a specific derivative of the principal indexed substance; and (d) stereochemical information. When a chemical substance is first encountered in the literature processed by CAS, its molecular structure diagram, systematic chemical name, molecular formula, and other identifying information are added to the CAS Chemical Registry and assigned a unique CAS Registry Number. Since 1967 many non-CAS publications have adopted CAS Registry Numbers to identify chemical substances.

CAS also developed an SSMC for chemical substances that includes:

- information about generic, proprietary, and trade names of a substance;
- three distinct, and well-defined, *representations* of any chemical substance concept including:
  - *Illustrative structural diagrams:* identifying the molecular skeleton of a ring system or stereoparent for convenience in interpreting associated entries in the Chemical Substance Index. Such diagrams indicate:
    - · the chemical elements present;
    - · the connection among atoms in a molecule of the substance;
    - · the types of connecting bonds;
    - · the arrangements of the atoms in space.
  - *Molecular formulas:* representing invariant properties of a chemical substance and derivable from its molecular structure.
  - *Index of Ring Systems* (IRS): listing names of cyclic skeletons contained in organic chemical compounds in the order determined by their ring analysis.
- *relationships* between chemical substance concepts.

The General Subject Index of CA includes general concepts relating to chemical substances, such as classes of substances, physical representations, chemical concepts, and phenomena, biochemical and biological subjects (other than specific biochemicals), and concepts relating to animals and plants (especially their common and scientific names.) In section 4.4 below, we present more detailed descriptions of this SSMC in terms its mapping into our general model.

3.2.3 SSMCs in Geographical Sciences

The Geography Markup Language (GML), developed by the Open GIS Consortium, is a structured representation of concepts used to model geographic features, both natural and man-made. The GML, which is an XML encoding for the transport and storage of geographic information including both the spatial and non-spatial properties of geographic features, is laid out in an OpenGIS Implementation Specification [13].

As noted in the Specification:

The underlying model of a geographic feature is essentially a named list of properties. Some or all of these properties may be geospatial, describing the position and shape of the feature. Each feature has a type, which is equivalent to a class in object modeling terminology, such that the class-definition prescribes the named properties that a particular feature of that type is required to have. A road, for example, might be defined to have a name, a surface-construction, a destination, and a centreline. The properties themselves are modeled in UML as associations, or as attributes, of the feature class. The feature property type is given by the role-name from an association, or by the attribute name. The values of the properties are themselves also instances of defined classes or types. So the road name is a text-string, the surface-construction might be a text token selected from an enumerated list, the destination is another feature of type Town, and the centreLine is a LineString, which is a geometry property.

4. SSMCS FOR LEARNING ENVIRONMENTS

We now describe a relatively general SSMC that we are using to represent concepts for pedagogical purposes. This model has been informed, in part, by the above examples of SSMCs, by [16], and by the work of Gärdenfors [14]. We first describe a simple classification of scientific concepts based on their roles within scientific activities. This may be used for describing slot values in our general concept model. We informally exemplify the model in terms of a simple example and present a mapping of the CA model of a chemical substance concept into our model to indicate its generality [7].

4.1 A use-based classification of concepts

Our use-based classification of scientific concepts is informally based on a National Research Council publication on scientific education [10]. This classification provides one basis for characterizing concepts in terms of objective scientific operations specifying an operational semantics. For example, an *abstract mathematical* concept, such as an *arithmetic equation*, may be defined in part by the objective syntactic operations that may be carried out on a symbolic representation of an *equation*; or a *measurable* concept, such as the *average velocity of a river*, may be defined in terms of the objective operations that may be used to determine the average velocity of a river. We term concepts that are interpretable within the span of this basis *operationally interpretable* con-

[7]We note that better mappings might be constructed by domain experts.

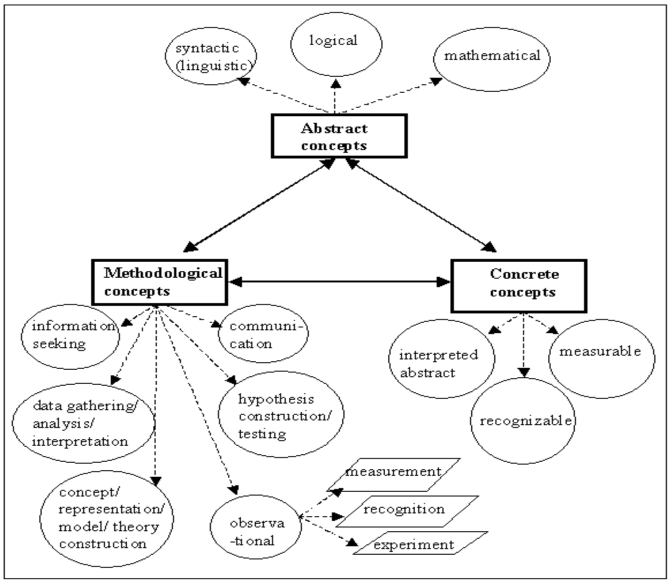

Figure 1: Classification of Scientific Concepts

cepts [8].

The basic form of our classification of scientific concepts is shown in Figure 1. The classification recognizes that the concepts employed by scientists cover a broad range of contexts, as illustrated, for example, by such concepts as *polygon, experiment, dataset, multiple linear regression, hypothesis, momentum, hydraulic geometry,* and *heat diffusion equation.* We believe it important for students to understand the main contexts of scientific activity in which a given concept finds application. Hence one application of the classification lies in providing students with a "model" of scientific activities.

At its highest level, the classification represents operationally interpretable concepts as belonging to one of three classes:

- **Abstract concepts:** have operational semantics defined in terms of syntactic (or computational) manipulations of symbolic representations. Three possible subclasses include, but are not limited to:
  - syntactic (linguistic) concepts;
  - logical concepts;
  - mathematical concepts.
- **Methodological concepts:** have a semantics defined in terms of various classes of scientifically well-defined operations that may be carried in relation to them. Possible subclasses of this class include, but are not limited to, concepts relating to procedures for:
  - scientific observation including, for example,:
    - · measurement;
    - · recognition;
    - · experiment.
  - scientific information seeking;
  - scientific data gathering, analysis, and interpretation;
  - scientific hypothesis construction and testing;
  - concept, representation, model, and theory construction;
  - scientific communication.
- **Concrete concepts:** have a semantics defined in terms of scientifically well-defined operations that provide the concept with an *interpretation* defined phenomena of the domain of scientific knowledge. The concept of *river discharge*, for example, has a characterizing set of operations that defines the concept in terms of various sets of measurement procedures that may be carried out in real-world contexts. Such procedures determine the amount of water passing across some *cross-section* of a *river* during a given *interval of time.* These concepts are, by and large, the class of concepts used in model and theory construction and include the important subclasses of:
  - measurable concepts;
  - recognizable concepts;
  - interpreted abstract concepts.

Figure 1 indicates the existence of non-trivial relationships between the three top-level classes. For example, the classical *heat diffusion equation* is a concrete concept in the sense that it is an (implicit) representation of the heat diffusion process using variables that are *measurable.* It may, however, also be viewed as a certain *mathematical concept* (namely a specific *partial differential equation* whose terms have been given an interpretation in terms of measurable concepts, such as *temperature* and *spatial location.*

We note in particular that this classification of concepts permits the representation of relatively high-level concepts, such as topics, that generalize over various (sub)classes of concepts.

4.2 Strongly-structured model of concepts

Our model of concepts may be viewed as having the form of a frame with six top-level slots (or facets):

- **Unique identifier:** for the concept for computational purposes.
- **Preferred term and synonymous term(s):** for the concept, with authorities and pointers.
- **Description(s):** of the concept in free text, with authorities and pointers.
- **Origin(s):** of the concept in free text, with authorities and pointers.
- **Relationships with other concepts:** including:
  - *Knowledge domain relationships:* with concepts relating to academic disciplines and academic topics within disciplines.
  - *Uses relationships:* with concepts relating to use(s) and role(s) of the concept in scientific activity, with authorities and pointers.
  - *Representational relationships:* with (alternative) sets of concepts that may be used to represent the concept, together with authorities and pointers. Representations have a fourfold classification as:
    - · explicit, full or partial;
    - · implicit, full or partial.
  - *Operational semantic relationships:* with concepts relating to the scientific operations that define the concept, such as the appropriate measurement concepts

[8] It follows that there are many concepts that cannot be so classified including such uninterpretable concepts as the *"way"* of Lao Tzu.

for a given measurable concept.
  - *Property relationships:* property relationships with concepts that represent properties (or characteristics) of the concept and that may be computed from a representation of the concept, together with authorities and pointers.
  - *Hierarchical relationships:* with concepts that are in some hierarchical relationship, including:
    - isa relationships;
    - partitive relationships;
    - other aggregate structures having hierarchical form.
  - *Causal Relations:* with concepts that may be in some (scientifically-accepted) causal relationship with the concept.
- **Instances (or examples):** of concepts, together with authorities and pointers.

4.3 Example: a polygon

We clarify this model with comments on various of its possible facet values for the case of the concept of a *two-dimensional polygon without holes* [9]. For simplicity, we omit references to authorities and pointers, but note that pointers may include "hotlinks".

- We assume that the term *two-dimensional polygon without holes* is the preferred term, and give *two-dimensional n-gon without holes* as a synonymous term.
- In relation to the value of both the **description(s)** and **origin(s)**, it seems of value to quote descriptions from accepted texts or articles.
- Concerning the **relationships to other concepts**:
  - For *knowledge domain relationships* we may note that *two-dimensional polygons without holes* are studied within the discipline of *mathematics*, and within the topic of *Euclidean geometry.*
  - for *uses relationships*, we may note that a *two-dimensional polygon without holes* is a concept within the mathematical subclass of abstract concepts, and indicate its various uses.
  - For *representational relationships*, we may note that a *two-dimensional polygon without holes* can be explicitly and fully represented as a *tuple* of *two-dimensional points without holes* $< p_1, ..., p_n >$ with appropriate constraints on the points; explicitly and fully as a *tuple* of *two-dimensional line-segments*; or implicitly and fully in terms of the intersection of a set of *two-dimensional half-planes* $< l_1, ..., l_n >$. The term "implicit" implies that additional operations (in this case intersection operations) must be carried out to obtain a representation of a *two-dimensional polygon without holes.*
  - For *operational semantic relationships*, one might specify the syntactic (mathematical) manipulations that may be carried out on the various representations of a *polygon.*
  - For *property relationships*, one might list the various concepts characterizing a polygon, such as *area, circumference*, or *centroid*, together with descriptions of the procedures or pointers to algorithms for their computation from a given representation. For example, the algorithm for deriving the circumference of a *two-dimensional polygon without holes* from a *line-segment* representation may be described by:
$$circumference = \sum l_i.$$
  - For *hierarchical relationships*, under the *isa* class on relationships, we may note that the concepts of *triangle, quadrilateral, pentagon, ...* form subclasses of *two-dimensional polygons without holes,* while *connected, closed, bounded, two-dimensional pointsets without holes* form superclasses. Analogous cases of *partitive* relations to other concepts may be given.
  - The class of *causal relationships* is not applicable for *two-dimensional polygons without holes.*
- Instances of *two-dimensional polygons without holes* may be represented in terms of symbolic or graphical representations of various specific examples of the concept.

[9] Representations of abstract concepts are typically easier to construct than representations of the other two classes of concepts.

4.4 CAS model mapped to general the model

To indicate the degree to which our general model of concepts is consistent with the SSMCs described in section 3.2, we map the CAS model for chemistry into our general model. It is possible to make similar "into" mappings from the other SSMCs presented above.

- **Unique identifier:** The CAS Registry Number designates the concept relating to a specific substance so far as its structure has been elucidated and can be defined in terms of atoms, valence bonds, and stereochemistry.
- **Preferred terms and synonymous term(s):** including
  - CA index names, each of which identifies a specific chemical substance;
  - common names;
  - trade names;
  - standard codes (e.g., one- and three-letter codes for searching proteins and nucleic acid sequences).
- **Description(s) of chemical substances:**
  - name modification(s) provide information pertaining to the index heading;
  - stereochemical descriptor adds three dimensional detail to the two-dimensional graph of a chemical structure;
  - substitutents;
  - homograph definition differentiates between identical or closely similar index headings possessing distinct meanings;
  - text modifications provide information relating the index heading to a particular referenced study (e.g., *adsorption of*);
  - indexing notes alert the index user to policies employed in the selection of index headings such as heading content, nomenclature, and indexing assumption;
  - qualifiers (e.g., radiation qualifiers, organ and tissue qualifiers) divide the heading into separate areas of study according to the nature of the topic discussed in the original document;
  - categories appear at subdivided compound and compound-class headings, with each category dealing with a different type of chemical derivative of the heading components.
- **Relationships with other concepts:** Various classes of our inter-concept classification are recognizable in this

model, including:

- *Knowledge domain relationships:* involving, for example, biochemical and biological subjects.
- *Uses relationships:* the CA contains uses and applications of chemical substances in its General Subject Index.
- *Representation relationships*, including:
  · the illustrative structural diagram, which identifies the molecular skeleton of a ring system or stereoparent.
- *Property relationships*, including:
  · the molecular formula, which provides a simple summation of the kinds and numbers of atoms in a molecule;
  · ring analysis, which leads to the information of the number, size, and atomic contents of the component ring of their skeleton.
- *Hierarchical relationships*, including:
  · relationships to concepts concerning chemical functional subdivisions or categories, alloy categories, and classes of substances.
- *Operational semantic relationships:* including, for example, engineering and industrial apparatus and processes.

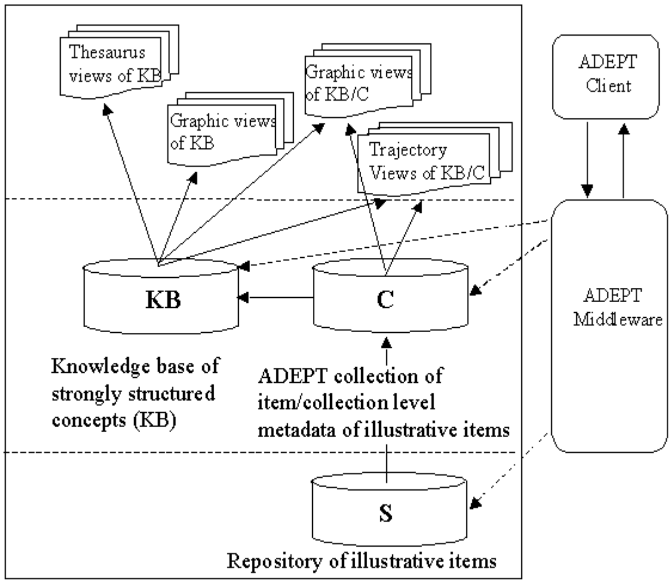

Figure 2: The Extension of the ADEPT Architecture for supporting the KB and attendant collection and services

5. EXTENDING ADEPT WITH A KRS

We are extending the Alexandria Digital Earth Prototype (ADEPT) [17] with a KRS that includes a KB of concepts, a collection of associated materials, and a set of services for use in supporting an introductory course in physical geography at the UCSB in Fall, 2002. The KB is based on our model of concepts. We first describe the general architecture of such extensions to ADEPT, then describe our development activities to support learning in the domain physical geography. We also provide a scenario illustrating some of its applications in support of learning.

5.1 Architecture for integrated DL-KRS

In Figure 2, we present the KRS-augmented architecture of ADEPT system. The instantiation of this architecture involves:

- **Constructing a KB of concepts:** Given the existence of a set of concepts for some domain, we construct a KB of these concepts by taking the abstract SSMC presented in section 4.2 above as the basis for an XML schema from this model, and creating an XML record for each concept. We are currently organizing these records as XML files, although we are exploring options that involve database and knowledge base technologies.
- **Constructing a collection of information illustrating specific facets of the concepts:** Our current approach to creating a DL collection of information illustrating the facets of some concept (such as representations of the measurement operations for some measurable concept) is to catalog the item as a part of an ADEPT collection. This involves using the ADEPT-DLESE-NASA (ADN) [1] metadata standard for learning objects with an extension that involves a field for the concept name(s) and facet(s). We are currently exploring an alternate approach that replaces the use of metadata document mark-up in terms of the concept KB and direct links to the concept KB. The long-term goal of this latter approach is to develop, or adopt, automated mark-up systems.
- **Creating services that provide access to the KBs and collections:** We are developing three sets of services over the concept KB and associated collections, including:
  - Services for creating, modifying, and deleting entries in both the concept KB and the collection of illustrative items.
  - Services for searching the concept KB and collection of illustrative items. For the latter collections, such services are just the search services provided by ADEPT over the ADN metadata.
  - Services providing both textual and visual views of the concept KB and collections. We may create various views of the concepts in the KB and their interrelationships. The basic principle concerning their creation is that *the concept KB is the sole source of any representation of concepts and interrelationships.* For example, we are developing software for creating thesauri as textual views of the KB. The software permits the creation of various "subset" thesauri by taking as input only a subset of concepts and interrelationships between concepts. If the subsets of interrelationships include only the *hierarchical relationships* between concepts, for example, then the views correspond to the heart of a traditional thesaurus. Various graphical views of the KB may be constructed, as shown in Figure 2.
    Similar comments apply to visual views of the KB, which include simple concept maps in which nodes represent concept (terms) and links represent relationships in the concept model. Hence, for example, one could view any specified subset of relationships over any subset of concepts from the KB. One may also create a visual view of a single model of a concept and its properties.

One may extend both classes of views to the collections. For example, one may create graphical summary views of the items in the collection that illustrate specified facets of a given subset of concepts.
  - Services for creating "personalized" collections representing, for example, trajectories through the space of concepts represented in the concept KB, together with associated illustrative materials. Such trajectories may be viewed as the "conceptual backbone" of a course designed by some instructor. Such a service allows easy re-use of the KB and collection in creating both course materials for instructors and papers for students.
- **Developing a client that provides access to the KB, collections and services from the perspective of the general user:** This client includes functionality supporting: (1) the creation and modification of the KB and collection; (2) access to the KB and collection [10]; and (3) the creation of views of the KB and collection, and the use of the KB and collection for various learning purposes.

5.2 KRS supporting physical geography

The principle author will teach an introductory course in physical geography, based on a KRS specifically tailored for physical geography. This course will be taught to approximately 200 UCSB undergraduate students in the Fall of 2002. These course support materials will be closely coupled with a specific textbook [11]. This approach to teaching in this domain of knowledge is informed by a set of four experimental lectures that were employed in simulating the use of a trajectory of concepts as a backbone for the topics covered and the use of a KB of structured concepts, a collection, and services.

All scientific concepts and interrelationships referred to in the book have been extracted and classified by: (1) chapter; (2) the classification described in section 4.1 above; and (3) their relative importance in the teaching of the course. This set of concepts has been augmented with additional concepts determined by the instructor.

All of the highest priority concepts are currently being modeled with the use of an XML schema that accords with the SSMC presented in section 4.2. Concurrently the collection of associated materials illustrative of the various facets of each concept model is being created as a regular ADEPT (multimedia) collection, accessible by the operational search middleware services of ADEPT [17]. The metadata created for each item follows the ADN metadata content standard for instructional materials [1] augmented with an additional field for the relevant concept(s) and their facets.

Services for creating textual and visual views of the KB and collection and for creating a trajectory through the KB and collection are under development. The course being developed may be viewed as a set of subtrajectories relating to each of the major topics covered in the course and typically correspond to the chapter organization of the book.

[10] Search of the collection is supported by the ADEPT middleware.

[11] *Elements of Geosystems* by R.W.Chistopherson and published by Prentice Hall, who have generously provided ADEPT with electronic versions of the text.

5.3 Use scenarios

The KRS provides students with access to knowledge that may be used for acquiring scientific understanding of the domain of physical geography. It provides instructors with reusable, re-purposable, and modifiable resources that enable the construction of teaching trajectories through the space of concepts. The services associated with the KRS facilitate: (1) the modification of revised versions of the KB and collection; (2) the creation of various views of subspaces of the concept space (including views of the relevant collections materials); (3) the ability to access and use the information in the KB and collection in various contexts; and (4) the creation of traces of trajectories through the concept space and associated parts of the collections.

To illustrate possible uses of the KB and collection in creating and using learning resources, we consider a specific scenario in which an instructor is creating a subtrajectory of concepts and illustrative materials for use in an instructional module whose goal is to develop (deep) student understanding of the concept of the *hydraulic geometry* (HG) of a *river*. One subgoal of the module, for example, may be to help students to develop understanding sufficient to provide scientifically acceptable answers to the question: "Why do the *cross-sections* of *rivers* take the forms observed in the real-world and what determines the variations in the *cross-sectional form* and the *discharge* of the *river*?

A KB for supporting the *HG* geometry problem contains structured models of all the important concepts used by scientists in representing and analyzing the phenomena of *HG* including, for example, such concepts as *river discharge, river cross-section, channel width, average channel depth* and *average stream velocity*. It also includes structured models of associated abstract and methodological concepts. The associated collection contains illustrative information about the various facets of each ot these concepts, including information that may be used to create and test such hypothetical relationships (i.e., *hypotheses*) between concepts as: *the width of a river is related to its discharge in some systematic manner*.

The services associated with the KB and collection include services for creating a trajectory through concept space that may, for example, form the backbone of a session constructed by the instructor in relation to the following high-level lecture plan.

- The instructor motivates the session with collection items containing (typically) implicit representations of the phenomena encompassed by the concept HG. These may include, for example, *aerial photographs* of *rivers* taken over some significant portion of their *length* and representing (implicitly) aspects of the concept of *downstream HG*. This concept denotes the changes in the *discharge, width, average depth, velocity, etc.* of a *river* as *discharge* changes in the *downstream direction*. They may also include *videos* taken at *cross-sections* of various *rivers* that represent (implicitly) the concept of *at-a-station HG*, or the changes in the *discharge, width, average depth, velocity* etc. of a *river* over *time* in response to a *rainfall event*.

- The instructor begins a discussion of those concepts that might be useful in constructing empirical and theoretical representations of *hydraulic geometry (HG)*. All such concepts may be found by constructing a KB-based view of all the concepts that relate to the topic of *HG*, including the concepts of *width* (w), *average depth* (d), *average velocity* (v) and *total discharge* (Q) at a *cross-section* of a *river channel*. The students might be led to develop specific models of concepts not previously specified in any detail. Using items from the collection, for example, the instructor may lead the students to a representation of a *river* as a sequence of *cross-sections*, each in turn characterized by concepts such as *w, d, v, Q*. The instructor may also encourage development of explicit representations for a *cross-section* in terms of the *tuple* of *width* and *average depth*.

- The instructor may then proceed to interact with the students in establishing hypothetical relationships between concepts (i.e., *hypotheses*) that may be considered to be not only important characterizations of *HG* but, if *verified*, as providing (partial or full) objective representations of the concept of *HG*. The instructor, may for example, provide the students with *datasets* from the collection that can be used for constructing and testing *hypotheses* about *equational relationships* that represent: (1) Q in terms of w,d, and v ($Q = wdv$); (2) changes in w, d,and v in terms of Q as one goes downstream (e.g., $w = aQ^b$, etc.), or *down-stream HG*); (3) changes in w, d,and v in terms of Q as Q changes over time at a given cross-section (e.g., $w = cQ^d$, etc.), or *at-a-station HG*. For such purposes, the instructor may discuss *methodological concepts* relating, for example, to *hypothesis testing*, *data analysis*, or *statistical analysis*.

Similar approaches may be used in constructing laboratory modules for self-guided learning or creating personalized collections of materials for papers.

6. CONCLUSIONS

We believe that the development and use of DL-based learning environments that are based explicitly on the use of KBs of SSCMs will lead to significant benefits to students and instructors. We are fully aware that the justification of our belief requires the use of objective evaluation procedures in assessing its efficacy as a learning support system (see, for example, [3]).

7. 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.

8. REFERENCES

[1] ADEPT/DLESE. Adn joint metadata content model. *http://www.dlese.org/Metadata/*, 2001.

[2] CAS. *CAS Chemical Abstracts Service Homepage*. www.cas.org, 2001.

[3] C.L.Borgman et al. Evaluating digital libraries for teaching and learning in undergraduate education: A case study of the alexandria digital earth prototype (adept). *Library Trends*, 49(2):228–250, 2000.

[4] S. Dagobert. Taxonomy of knowledge organization sources/systems (draft). *http://nkos.slis.kent.edu/KOS_taxonomy.htm*, 2000.

[5] ISO/IEC. *ISO/IEC 13250 Topic Maps*. www.y12.doe.gov/sgml/sc34/document/0129.pdf, 2000.

[6] J.C.Binwal and Lalhmachhuana. Knowledge represenation: Concept, techniques, and the analytico-synthetic paradigm. *Knowledge Organization*, 28(1):5–16, 2001.

[7] J.T.Bruer. *Schools for Thought: A Science of Learning in the Classroom*. MIT Press, Cambridge, MA, 1993.

[8] R. Mayer. *Teaching for Meaningful Learning*. Merrill Prentice Hall, Upper Saddle River, NJ, 2001.

[9] NIST. *MatML: eXtensible Markup Language for Materials Property Data*. http://www.ceramics.nist.gov/matml/matml.htm, 2001.

[10] NRC. *National Science Education Standards*. National Academy Press, Washington, DC, 1996.

[11] NSDL. Core integration systems projects. *http://www.smete.org/nsdl/projects/CIS.html*, 2001.

[12] NSF. *National Science, Mathematics, Engineering, and Technology Education Digital Library (NSDL) Program Solicitation*. http://www.nsf.gov/pubs/2001/nsf0155/nsf0155.htm, 2001.

[13] OGIS. Feature and geometry models. *Geography Markup Language (GML) 2.0*, pages http://opengis.net/gml/01–029/GML2.html#sec1.2, 2001.

[14] P.Gardenfors. *Conceptual Spaces: the Geometry of Thought*. MIT Press, Cambridge, MA, 2000.

[15] R.E.Mayer. *The Promise of Educational Psychology: Learning in the Content Areas*. Merrill Prentice Hall, Upper Saddle River, NJ, 1991.

[16] T.R.Smith et al. Computational modeling systems. *Journal of Information Systems*, 20(2):127–153, 1995.

[17] T.R.Smith, G.Janee, J.Frew, and A.Coleman. Towards the alexandria digital earth prototype system. In *Proceedings of the Joint Conference on Digital Libraries*. ACM/IEEE, June 2001.

[18] D. Weisgerber. Chemical abstracts service chemical registry system: History, scope, and impacts. *Journal of the American Society for Information Science*, 148(4):349–360, 1997.