Ontology Services for Curriculum Development in NSDL

Amarnath Gupta Bertram Ludäscher Reagan W. Moore
San Diego Supercomputer Center, University of California San Diego
9500 Gilman Drive, La Jolla, CA 92093 USA
phone: +1 858 534 5000 {gupta,ludaesch,moore}@sdsc.edu

ABSTRACT
We describe our effort to develop an ontology service on top of an educational digital library. The ontology is developed by relating library holdings to the educational concepts they refer to. The ontology system supports basic services like ontology-based search and complex services such as comparison of multiple curricula.

Categories and Subject Descriptors
H.3.1, H 3.7 [**Information Storage and Retrieval**]Digital Libraries —*system issues*, Content Analysis and Indexing – *abstracting methods*

General Terms: Algorithms, Languages

Keywords: Ontology, NSDL, concept navigation, information integration, science education, knowledge representation

1. INTRODUCTION
The National Science Foundation has sponsored a number of research projects to develop a National Science, Mathematics, Engineering, and Technology Education Digital Library (NSDL). NSDL will consist of a wide variety of instructional material (on-line textbooks, web pages, slides, videos, etc.) so that an educator can use the library to perform tasks like:

1. select audience-appropriate items that cover the topics of instruction
2. evaluate the goodness of some instruction-set (e.g., a textbook) with respect to the topics of instruction
3. develop a complete curriculum for students at a specific level of maturity

As part of the NSDL project, SDSC is developing an ontology and a set of ontology services on top of NSDL to facilitate searching, navigation and comparison of educational material.

2. AN ONTOLOGY FOR NSDL
An ontology is a "formal explicit specification of a shared conceptualization" [4], where conceptualization refers to "an abstract model of how people think of things in the world, usually restricted to a particular subject area" [5]. Given the education-oriented nature of NSDL, we are developing an ontology based on a curricular "roadmap" developed by the American Association for the Advancement of Science (AAAS) [1]. This roadmap covers over 20 areas for kindergarten-to-high-school education in science. We argue that such ontologies will serve as the backbone that glues together diverse and distributed collections in the digital library and can be used to perform curriculum development tasks outlined above.

2.1 Ontology Construction
Following the AAAS model, the SDSC science curriculum ontology divides science education into a number of *setting*s. A setting is divided into a number of *clusters*, which is further broken down into a number of *sections*. For example, the "Physical Setting" (PS) has a cluster called "Structure of Matter" (SoM) that has a section called "Atoms and Molecules" (AM). Each section consists of a set of *teaching units* that outline a complex concept to be imparted to the students as a single body of knowledge. Thus, a unit can be uniquely identified using its fully qualified name (e.g., PS.SoM.AM.4D/1). Two units may be related using the distinguished predicate **prereq**(unit,unit), meaning that the first unit is a *prerequisite* for the second unit. Predicate **prereq** is nonreflexive, antisymmetric and transitive, thereby inducing a partial order on units. The section PS.SoM.AM consists of 21 units (see [2]) connected to units both within and outside the section. In addition, a set of units can have ***named subsets*** with additional properties. For example, each subset may have an attribute "audience" to specify appropriate grade levels. A unit itself is a complex concept, comprising a graph of simpler concepts and their interrelationships.

Example. The teaching unit PS.SoM.AM.4D/1 states: "*Atoms are made of a positively charged nucleus surrounded by negatively charged electrons*". In our NSDL ontology, this is expressed as:

c1 ≡ atom**: concept** c2 ≡ nucleus**: concept**

c3 ≡ electron **: concept** c4 ≡ charge**: concept**

c4**.values = enum**(positive, negative)

c5 ≡ **prop**(has_component(atom, nucleus)) **: concept**

c6 ≡ **prop**(has_component(atom, {electron}) **: concept**

c7 ≡ **prop**(has(nucleus, charge(positive))) **: concept**

c8 ≡ **prop**(has(electron, charge(negative)) **: concept**

c9 ≡ **prop**(surrounds({electron}), nucleus)) **: concept**

u1 ≡ {c1, c2, c3, c4, c5, c6, c7, c8, c9} **: unit**

Here c1…c9 are concept identifiers. Encoded through the distinguished predicate **prop**, c5…c9 are properties that behave as concepts. Concept c5 contains and thus depends on c1. Within the ontology, if **prereq**(u1, u2) holds, then u1 and u2 have at least one concept in common.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
JCDL'02, July 13-17, 2002, Portland, Oregon, USA.
Copyright 2002 ACM 1-58113-513-0/02/0007…$5.00.

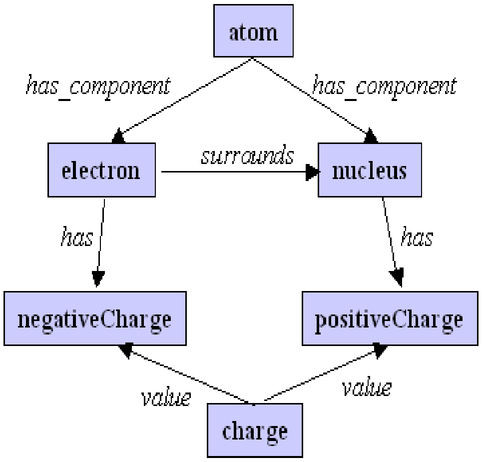

Figure 0. The Ontology Graph for a fragment of the Example Teaching Unit. The ontology can now be modeled as a graph-structured database.

2.2 Registering a Collection with the Ontology

A collection in the digital library consists of single items such as an annotated picture, collection of items such as in a web page, or a more complete discourse such as linearly organized set of chapters in a textbook or a hierarchically structured set of pages in a web site. Although each item and collection may come with its own descriptive metadata, the metadata are usually not semantically linked to each other or to a common vocabulary. This makes navigating through the collections and semantic searching of the material difficult.

In our model, each registered collection and a digital object within a collection is mapped back to an ontology. The system admits a set of mapping predicates such as *example_of, discovery_of, description_of, explanation_of, anecdote_about*, etc. When a library holding $H is placed in the collection, the author fills in a set of predicates, e.g., *example_of*($H, atom), *description_of*($H, hydrogen_atom) where the first term is a reference to the holding and the second term denotes a concept that the digital object "is about". Mapping predicates can be composed. For example, a digital object may represent a mapping like:

anecdote_about($H, *discovery_of*(electron))

It may also provide additional information that would enrich the ontology, such as *isa*(hydrogen_atom, atom). When the entity in a collection is registered, it automatically initiates a search for the keywords in the ontology. If the keyword does not match any concept word or relationship in the ontology, it is reported to the author. Otherwise, all units matching the keywords are fetched. This allows the author to evaluate whether the keywords used to register the material indeed refer to an appropriate teaching unit, and whether the mapping predicate has been appropriately chosen. If not, the author may choose to revise the keywords or the mapping. Once the registration is complete, the ontology manager adds the new element to its index of library holdings relevant for the set of concepts.

3. ONTOLOGY SERVICES

With the ontology constructed and objects registered to the ontology, the SDSC system offers a set of basic and advanced ontology services.

3.1 Basic Services: Navigation and Search

The system allows a user to browse through the graph of concepts and relations within a unit as well as across units. For any combination of concepts, the user may navigate through related concepts, prerequisite concepts and retrieve all library holdings indexed by one or more concepts. The user could also perform ad hoc searches on the holdings, by searching against the mapping predicates. For example, the user may look for all holdings $H, such that

description_of($H, nucleus) **and**

explanation_of($H, *has*(nucleus,charge(positive))).

Here, the predicates defining the mapping are also used as the constructs of an *ontological view definition language* over the set of holdings, such that queries can be formulated over them.

3.2 Advanced Services: Curriculum Analysis

The graph-structure of the units and the dependency graph across units enables the ontology system to support advanced capabilities like partial graph matching. Consider that an educator wants to evaluate a textbook, represented as a linear graph of chapters, where each unit is a graph of concepts. The system employs graph-matching techniques to determine whether the textbook contains missing concepts by first searching for the units touched by each chapter, and then by computing the difference between the concept set within the unit and those covered by the chapter. Similarly, the system can also determine if the order of concept coverage in the textbook preserves the same dependency structure as the curricular ontology.

4. CONCLUSION

We have sketched the ontological framework being developed at SDSC as part of the NSDL project. The first version, implemented using a deductive database based on F-logic, is currently in place. We are revisiting the architecture to be more scalable and usable by a larger body of scientific community. We also plan to evaluate the utility of the system by creating a test bed for schoolteachers and education planners.

5. ACKNOWLEDGEMENTS

We thank David Fulker for pointing us to the AAAS atlas, and the NSF NSDL program for partial support of this research.

6. REFERENCES

1. American Association for the Advancement of Science Web site http://www.project2061.org
2. http://www.project2061.org/tools/atlas/sample/4_7_AM.pdf
3. Decker, S. et al., Ontobroker: Ontology Based Access to Distributed and Semi-Structured Information. *Semantic Issues in Multimedia Systems. DS-8.* Kluwer, 1999.
4. Gruber, T., A translation approach to portable ontologies. *Knowledge Acquisition* 5,2 (1993), 199-299.
5. Gruninger M., and Lee, J., Ontology applications and design. *Commun. ACM 45*, 2 (2002), 39-41.
6. Noy, N.F., Fergerson, R.W., and Musen, M.A. The knowledge model of Protege-2000: Combining interoperability and flexibility. Proc. EKAW 2000.