Fine-grained Representation of Educational Content based on Ontologies

Howard Beck
Agricultural and Biological Engineering Department
College of Agriculture and Natural Resources
University of Florida
Gainesville, FL. 32611 - USA
hwb@ufl.edu

1. Introduction

The process of creating effective educational materials that fully utilize electronic media can be greatly improved. The goal of using instructional technology is, of course, to provide students with better quality educational experiences. Among other things, this means that educational materials should be provided in a way that matches the individual student's learning style, in a variety of selectable formats. At the same time, the cost to produce such materials must be significantly reduced if such an approach is to be widely used. For the most part, instructors are still struggling to prepare their course materials electronically. Conventional tools do not help this situation, and actually make the task more difficult by storing course materials in formats that are not flexible. New tools are needed to assist instructors in creating electronic course materials in a way which minimizes the development time to the instructor, while providing the best possible organization of course content and producing materials of high quality.

In order to achieve these objectives, an architecture for course content management is proposed based on ontologies, description logics, and object databases. This architecture enables educational content to be represented at a very fine-grained resolution, not just at the level of documents or images, but at also at the level of individual concepts occurring within these course-grained media. Ontologies are electronic dictionaries of concepts, and can be used to model the concepts and relationships among concepts in a course. Description logics offer a formal basis for knowledge representation that supports both the ontologies and a supporting database model. The object database is the underlying physical platform for storing content. Course content represented using these technologies is much easier to manipulate, reuse, and present in different styles compared with conventional Web technologies. Most important is a new generation of development tools that instructors can use to create effective course content more easily.

2. Approach

2.1 Ontologies, Description Logics, Object Databases

Ontologies are used to create a comprehensive catalog of the concepts within a domain, and to show how these concepts are interrelated. There is a wide variety of languages used for building ontologies. These range from simple bipartite graphs to advanced description logics. Leading ontology languages, include RDF, Topic Maps, DAML, OIL, and

PGLDB'2003, pp. 105-110, 2003.

Ontolingua, Description logics are the most formal of these languages, and provide a mathematical basis for automatic manipulation of concepts. Figure 1 shows a language that is currently being used to build concepts in the University of Florida Multimedia Database.

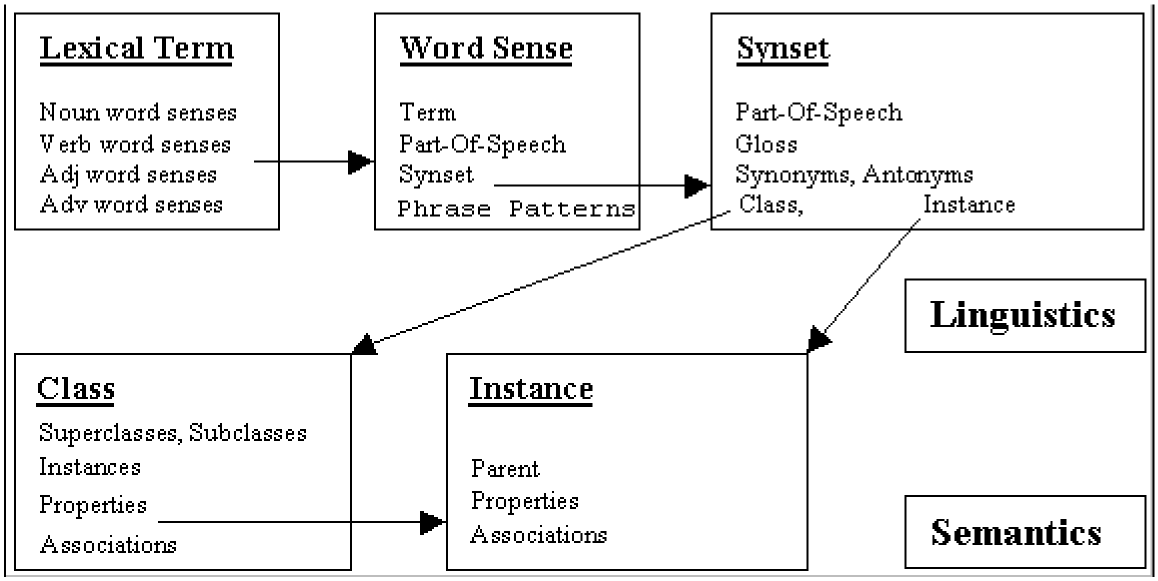

Figure 1. A description-logic language used to build concepts in the UF Multimedia Database.

A commercial object database management system (ObjectStore) is used as the underlying storage device for concepts built using this language. The database is a convenient repository for the complex object. The resulting architecture thus provides an ontology, a formal language for building concept descriptions, and a database management system combined into a single environment. Database objects can be exchanged between distributed servers using CORBA, RMI, XML/Web Services, and other techniques.

2.2 Data Visualization

The tools for creating and managing educational content are based on the assumption that people naturally react to graphic visualization of data much better than character-based presentations. Since the underlying object data structures really are graphs (nodes and links) they lend themselves well to a graphic presentation. Graphic tools rely heavily on aesthetic presentation. They attempt to create an "object world" that is specific to the domain in which a particular individual is interested, whether that individual might be a soil science professor, a county agent working in fruit crops, a student studying genetics, or a programmer developing software.

Figure 2 shows a graph-based browser, designer, and editor for directly manipulating objects in the database. As with all the tools describe here, this browser is accessible through any Web browser, giving on-line access to view and edit the database though visualization tools.

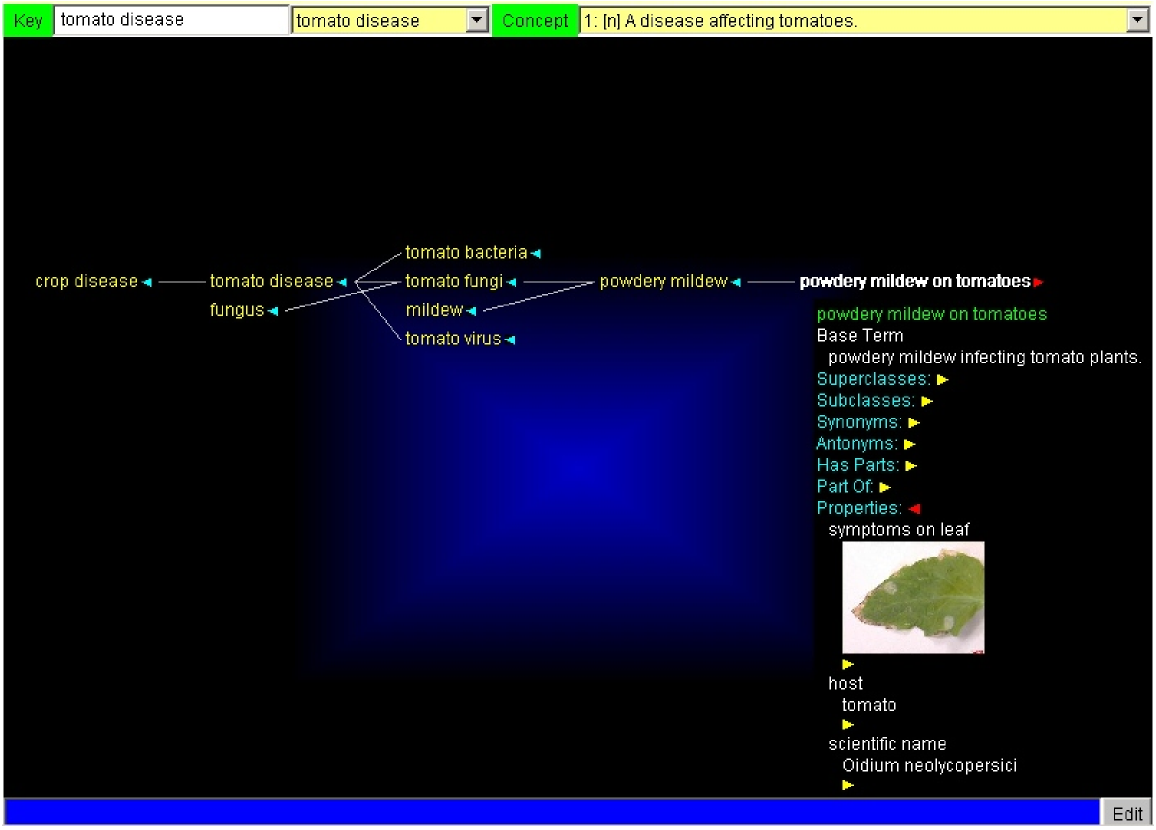

Figure 2. Web-Taxonomy Browser (http://orb.ifas.ufl.edu).

2.2.1 Course Development

Figure 3 shows a graphic tool currently being used by instructors to develop course content. Instructors map the concepts in a course by building diagrams showing concepts and their interrelationships (high-level access to the ontology). Pop-up editors enable media (text, graphs, images, video) to be inserted as part of the course content. All the objects in the course are stored in the back-plane object database management, hence are searchable and reusable and can be combined into other course diagrams. Once the course content is created, it can be presented in a variety of styles (See section 2.4).

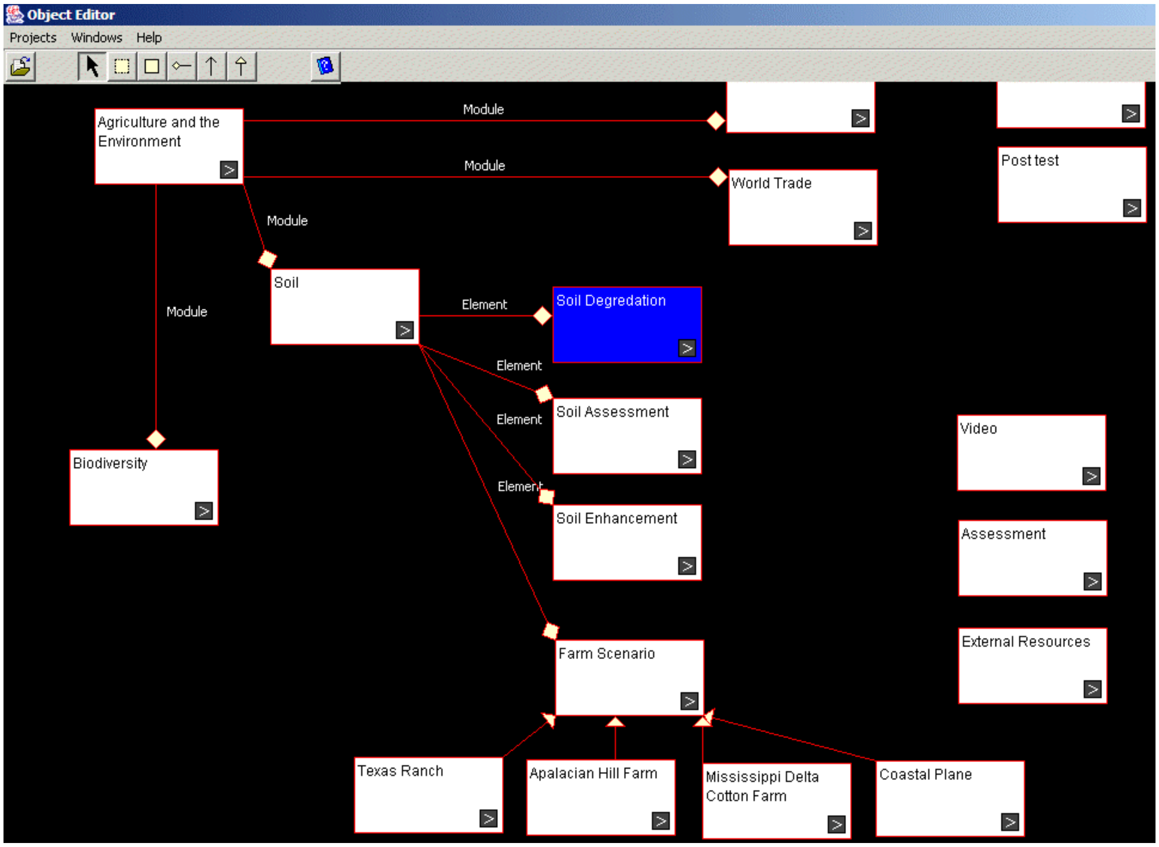

Figure 3. Course Development Tool.

2.2.2 Mathematics and Simulation

A mathematics and simulation environment (to be developed in 2003) will provide graphic tools for design and construction of computer simulations. A symbolic representation for mathematical equations will be incorporated into the database along with tools for creating, editing, exploring, and solving equations. This will be applied to construct irrigation models, fertilizer management tools, crop phenology models, and models of a variety of dynamic systems in agriculture and natural resources.

2.2.3 UML

UML Tools (Unified Modeling Language) are being integrated within this environment to allow for design, development, documentation, and deployment of software systems. The UML Tools include a class diagram which contains all class definitions, along with documentation and source code. Deployment and component diagrams are also available. Use case and sequence diagrams are currently under development.

2.2.4 Workflow

Many applications require that information be routed to and from various people and processes. For example, course materials must route through authors, reviewers, publishers, and others. Data must flow between various organizations. Workflow diagrams can be used to build and design these processes. The nodes in these diagrams can also be enhanced using ontologies.

2.2.5 Web Development Tools

Traditional tools (Microsoft Word, FrontPage, PowerPoint, Adobe Acrobat, Flash) do not attempt to represent content in any organized fashion, and suffer from a failure to cleanly separate content from presentation. New graphic tools will allow experts to focus on content and structuring their knowledge within the context of reusable objects. A wide variety of presentations can then be automatically generated from the same content. Figure 4 shows a content development tool used to create a Web-based presentation. . Instead of slides or HTML pages, this tool arranges information according to content. A presentation layer (Section 2.4) then automatically generates a Web site from this content.

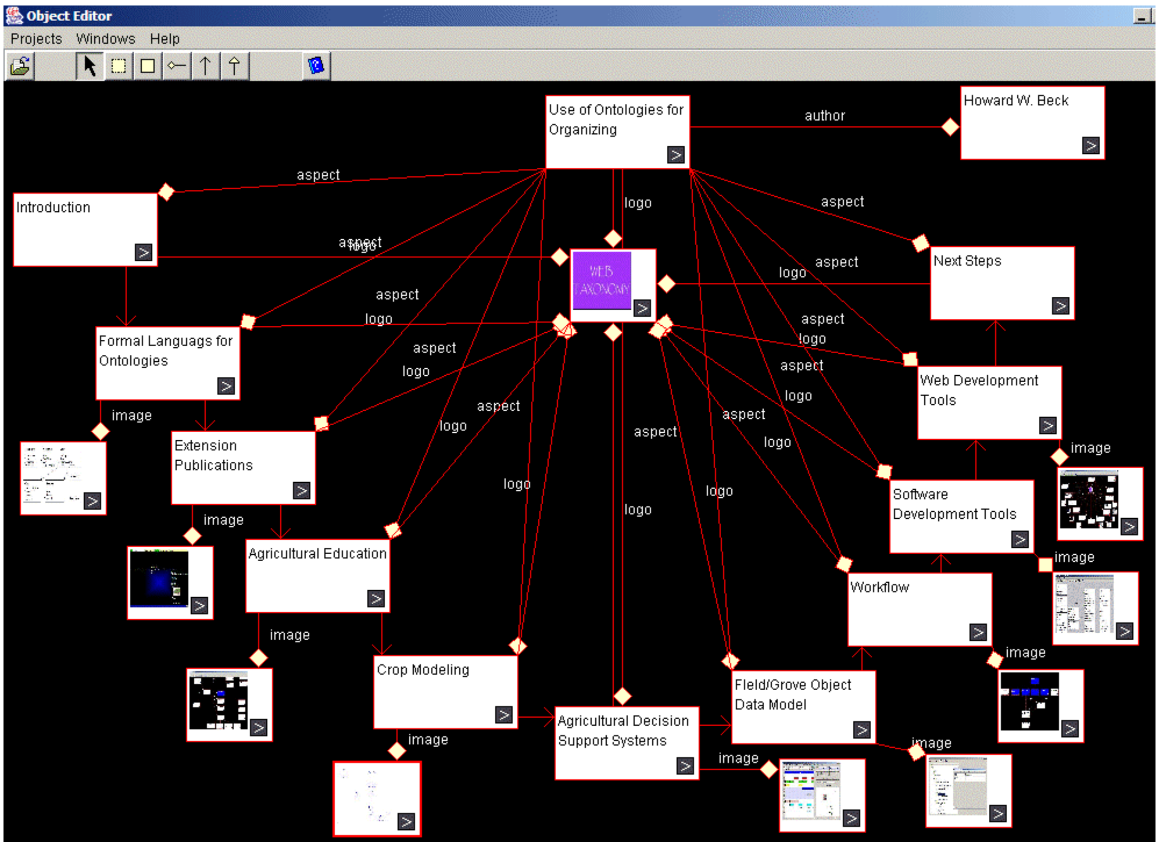

Figure 4. Web-based Development Tool. (This presentation can be viewed on-line at http://orb.at.ufl.edu/ObjectEditor -> APAN 2003 Presentation).

2.3 Linguistics

The ontology provides a dictionary, and when combined with phrase patterns the system provides a platform for natural language applications. An experimental system for information extraction is being built that attempts to create database objects from concepts extracted from the text of figure captions in order to automatically catalog images. The potential for a multilingual ontology that can be used to create multilingual applications also exists within this environment. Currently a multilingual ontology of agriculture is being developed in collaboration with an international project sponsored by the United Nations Food and Agricultural Organization (http://www.fao.org/agris/aos).

2.4 Presentation Layer

Once course content is created and stored in the database, it can be presented in a variety of styles. Using a technique similar to XML style sheets, database objects representing content are matched with objects specifying presentation style. This is used to automatically create Web pages, custom presentations, and even print media from the database content.

3. Next Steps

Currently the system described here is being used to build three prototype courses, on Agriculture and the Environment (Southeast US Agricultural Extension project), Neurology (UF Medical Center), and on bioprocessing (NASA Educational Outreach). This will evaluate the tools for course development, and explore presentation, reusability, and other issues.

The greatest need for research in the ontology/description logic/object database is in data manipulation. Query processing using description logics needs to be expanded and applied to the multimedia database. Automatic tools for ontology construction, including automatic classification, and natural language techniques in information extraction, must be developed to assist in database content generation.