Towards an Interlingua for Information Objects

Daniel D. Suthers
Learning Research and Development Center
University of Pittsburgh
Pittsburgh, PA 15260
+1-412-624-7036
suthers+@pitt.edu

ABSTRACT

To improve the cost/benefit ratio of the educational technology of the future, representations are needed by which information objects can be shared by diverse applications communicating over the Internet, yet without limiting future development in advanced technology. To address this, need we are developing representations based on SGML and KIF that are downwards-compatible with HTML, so that students who do not have advanced display software can still access the media via more common displayers. Client-server technology can then provide such students with access to more advanced functionality.

KEYWORDS: Communication and the Internet, SGML.

INTRODUCTION

During learning that involves the use of computer and network support, a variety of information objects must pass between students, teachers, and supporting software. In order to make intelligent training systems incrementally extensible, shareable, and affordable, these information resources must be reusable, and systems must be capable of using client-server configurations in which information objects can pass freely from one network resource (server or student station) to another. This requires that all information objects involved be specified in standard forms and terms. Yet, we do not want use of standard and reusable media to limit further advances in educational technology. To address these needs, the author and colleagues are designing shareable and reusable representations for educational media that can support advanced functionality while being downwards-compatible with the HTML-based browsers that are rapidly becoming ubiquitous.

SHARED INFORMATION OBJECTS

The representations are based on Standard Generalized Markup Language (SGML)[1] and Knowledge Interchange Format (KIF)[2, 3]. SGML focuses on document structure and KIF focuses on semantics; this is why both are required.

Syntactic Foundation. HTML has raised the "lowest common denominator" of shareable documents to one that includes basic typographical structure. However, HTML does not provide a common format for other kinds of structure, nor does HTML address the problem presenting a given body of information in several substantially different ways. SGML is the established WWW standard for writing Document Type Definitions (DTDs) that indicate how to "mark up" the structure of the information contained in documents, independent of how that information is displayed or used in particular applications. SGML-based parsers can filter out markup not usable by a particular application, thereby providing multiple viewpoints on information, or applications can be designed to simply ignore tags they are unable to process.

Semantic Foundation. KIF is an ontological interlingua, currently under consideration as an ANSI standard. Shared ontologies enable knowledge-based systems to communicate with each other in spite of differences in their representational languages. Participants of ARPA's Knowledge Sharing Effort are collecting and making available an inheritance hierarchy of ontologies ranging from fundamental (e.g., sets and numbers) through intermediate (e.g., components with constraints), to domain specific (e.g., satellite design). We are building on these ontologies, defining extensions in KIF as needed.

Combining SGML and KIF. We need to embed semantic annotations in our documents to enable current and future efforts in intelligent educational technology to use them, but we must do so without impairing accessibility to SGML-based tools. We are experimenting with annotations in SGML DTD headers to reference KIF-based ontologies available via a server. The annotations identify any associated ontologies and specify the mappings between (a) SGML tags and KIF classes; and (b) SGML attributes and KIF relations.

Table 1 illustrates how we identify the ontologies associated with a document type and how we specify the mapping between SGML entities and ontology entities. (The examples are highly simplified due to space constraints.) The DTD is for a miniature version of "ArgML," which describes the logical and rhetorical structure of an argument. We assume for the sake of illustration that we have defined two ontologies, one for science concepts such as empirical reports and hypotheses, and the other for Toulmin argument structures. The ONTOLOGY declarations give the name and WWW location of these ontologies. The INTERPRETATION declarations

```
<!DOCTYPE MiniArgML.DTD>
<!ONTOLOGY NAME=science
   HREF="http://www.pitt.edu/ suthers/ontologies/science.kif">
<!ONTOLOGY NAME=toulmin
   HREF="http://www.pitt.edu/ suthers/ontologies/toulmin.kif">
<! SGML definitions of document syntax. >
<!ENTITY % statement "report | hypothesis | principle">
<!ENTITY % link          "support | rationale">
<!ENTITY % link-attlist "id          ID        #IMPLIED
                         source     IDREF #REQUIRED
                         target      IDREF #REQUIRED">
<!ELEMENT text                      - - #PCDATA>
<!ELEMENT (%statement;|%link;) - - #PCDATA>
<!ELEMENT argument                - - (text|%statement;|%link;)*>
<!ATTLIST (%statement;) id ID #IMPLIED>
<!ATTLIST (%link;)           %link-attlist;>
<! Semantic interpretations of document elements. >
<!INTERPRETATION report       empirical-observation>
<!INTERPRETATION hypothesis hypothesis>
<!INTERPRETATION principle    inference-pattern>
<!INTERPRETATION support      toulmin-support
                         source      grounds
                         target       claim>
<!INTERPRETATION rationale    toulmin-support
                         source      warrant
                         target       toulmin-support>
```

Table 1: Document Type Definition for a simplified version of ArgML, with ontological declarations.

```
<!DOCTYPE MiniArgML.DTD system "mini-argml.dtd">
<MiniArgML.DTD>
<HEAD><TITLE> Why I think HIV causes AIDS </TITLE></HEAD>
<BODY> <ARGUMENT>
<TEXT><H1> Why I think HIV causes AIDS: </H1></TEXT>
<REPORT id=I-1>
   95% of infants born with HIV develop AIDS by 6 years of age.
</REPORT>
<SUPPORT id=I-3 source=I-1 target=I-2>
   <TEXT> Therefore, </TEXT> </SUPPORT>
<RATIONALE id=I-5 source=I-4 target=I-3>
   <TEXT> because </TEXT> </RATIONALE>
<PRINCIPLE id=I-4> correlation suggests causation, </PRINCIPLE>
<HYPOTHESIS id=I-2> HIV causes AIDS. </HYPOTHESIS>
</ARGUMENT> </BODY>
</MiniArgML.DTD>
```

Table 2: Simplified MiniArgML document

give the semantic interpretation of each of the SGML tags and attributes in terms of concepts from the ontologies.

Example. To illustrate our approach, consider the example "mini-ArgML" document shown in Table 2. Figure 1 shows how this document is displayed in a current version of Mosaic (similar results are obtained in other popular WWW browsers). Figure 2 shows how this document could be displayed in our argument grapher [4] with minimal modifications (currently underway) to read the new format.

We are developing a suite of tools that communicate via these representations. A client-server approach will enable any student in the world to make partial but substantial use of our toolset without necessarily having a local copy of these tools, or even an environment powerful enough to run them. Our particular efforts are targeted towards coached apprenticeship in scientific argumentation [4]. However, we ultimately hope to extend the shareable representations and tools to other learning situations and tasks.

Figure 1: Mosaic display of an ArgML document.

ACKNOWLEDGMENTS

This research was conducted while supported by grant MDR-9155715 from the NSF Applications of Advanced Technology program. The author also thanks Alan Lesgold for valuable input and support.

REFERENCES

1. C. Goldfarb. *The SGML Handbook*. Oxford: Oxford University Press, 1990.

2. T. Gruber. A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5(2):199–220, 1993.

3. R. Fikes (Ed.) M. Genesereth. *Knowledge Interchange Format, Version 3.0 Reference Manual, Technical Report Logic-92-1*. Computer Science Department, Stanford University, 1992.

4. M. Paolucci, D. Suthers, and A. Weiner. Belvedere: Stimulating students' critical discussion. In *CHI95 (to appear)*, Denver, CO, May 1995.

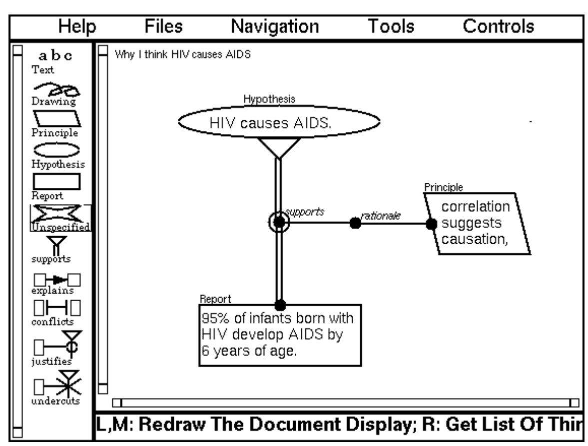

Figure 2: Belvedere display of an ArgML document.