Reagan W. Moore (San Diego Supercomputer Center)

Digital Library and Data Grid Technology Group

Semantic web technology provides the ability to characterize relationships between named entities and the ability to infer metadata based on assertions associated with each type of relationship. The integration of semantic web technology with digital libraries and data grids is particularly well suited to the application of these capabilities to name spaces that are being managed by digital libraries and data grids. The name spaces include logical resource naming, distinguished user naming, logical file naming, metadata naming, and constraint naming. Relationships can then be asserted within a name space and between name spaces (such as access controls). Examples of the use of ontologies include the ability to impose multiple organizational structures on a name space, search for properties associated with names, support mediation on user–defined metadata within a name space, and support mediation across independent name spaces.

Digital libraries and data grids manage state information about data and workflow. Ontologies can be created across the name spaces for types of operations, the name spaces for types of state information, the name spaces for types of services, the name spaces for properties of operations, and even the name spaces for events. By characterizing the relationships between these name spaces, it should be possible to build inference engines for mapping properties between grid services, and develop ontology tools that manage grid service composition.

A particular example is the use of ontologies to name classifications and categorizations created by data mining and knowledge management systems. Data mining infers relationships on data that can be managed in ontologies. Knowledge algorithms can be organized in hierarchies similar to the Guide for Available Mathematical Software. Reasoning across the knowledge algorithms can then be used to identify appropriateness of an algorithm for a given purpose.

A final example is the concept of persistent objects. The preservation community, digital library community, and data grid community have the challenge of understanding how to interact with old encoding formats. A Global Grid Forum effort, led by the

Data Format Description Language Working Group, is developing the ability to characterize the encoding format of a data file as a set of structures. Allowed operations on the structures can be associated with the data file through the use of an ontology. Applications that manipulate data structures can also be characterized by the set of operations that they perform. By mapping the application operations to the allowed operations associated with the data files, a system can be developed that allows new applications to manipulate legacy data, and that allows legacy applications to manipulate current data. The integration of ontology management tools with data structure characterization technology promises to solve the problem of accessing obsolete encoding formats.