Soft-Classing to Create Evolving Ontologies for Distributed Resources

Kenneth J. Laskey, PhD
kenneth.j.laskey@saic.com

Qiang Lin, PhD
qiang.lin@saic.com

Science Applications International Corporation (SAIC)
1710 SAIC Drive
McLean, Virginia 22102 USA

ABSTRACT

The value of a distributed resource is often enhanced by its remaining distributed and under the control of the party responsible for its creation and maintenance. Metadata expressed as XML provides a mechanism to publish the existence of this resource and to support multiple access of the resource from its source location. In this way, uncertainty is minimized about whether the resource is current or from where the resource originates. A challenge then is to characterize the resource in a flexible way which allows the characterization ontology to capture new information in a time frame consistent with the evolution of knowledge in the domain itself. Soft-classing has been developed as a means to support such a responsive system.

Keywords

Soft-classing, ontology, metadata

INTRODUCTION

In order to make effective use of distributed resources, it is necessary to describe what the resource is, where the resource can be found, and how the resource can be accessed. The resource itself need not be Web accessible, only the characterizing metadata and the ability to access the schema through which the resources are classified. In a distributed environment where the information domain contains content from many sources, the desirable attributes of an organizing ontology include

- easy distribution across the community of interest
- flexibility in how concepts are represented
- ease of evolving to accommodate an expanding knowledge base
- a means to evolve the ontology without invalidating data already catalogued.

While the Web supports distribution, soft-classing is introduced as a means to create such a flexible ontology structure, UCLP as the mechanism to capture the metadata defining the structure, and an Oracle implementation as the storage medium to hold both the structure and instances of data which the structure encompasses.

DESCRIPTION

The idea behind soft-classing is to provide a means through which an ontology can be modified by its community of use while requiring a minimum amount of knowledge about the underlying data storage structure. In the current discussion, an ontology is defined to consist of

- a series of hierarchical nodes which represent a type-of or composed-of breakdown of a domain item
- a set of properties associated with each class node which define the metadata which can be searched by users
- instances of the class nodes which contain metadata values describing the instances.

Rather than building a data model to directly implement a domain's ontology, soft-classing uses the XML tags defined through the Universal Commerce Language and Protocol (UCLP) to provide the structure through which the domain can expand its self description. The tag set provides for an information typing that classifies properties such as identifiers, feature sets, and parametric values as shown in the following UCLP syntax examples (attribute values are in *italics* and attributes in brackets [ ] are optional) from the approximately 20 tags defined in Reference 1:

<UC_ID name = *{name}* value = *{value}*
[privacy = *{privacy}*] />

<UC_PAR name = *{name}* value = {value}
[units = *{units}*] [tolerance = *{tolerance}*]
[privacy = *{privacy}*] />

These properties are contained in an XML "UC registration" block which identifies the domain and hierarchy class of interest

<UC domain = *{domain}* version = *{version}*
class = *{class}* [status = *{status}*]
[privacy = *{privacy}*] >

... property elements...

</UC>

An implementation of soft-classing using Oracle 8i is currently in progress. The implementation for the Multiple Information Source Tracking Infrastructure (MISTI) allows the user within a domain to create and implement a hierarchical data model of that domain, building such data model on top of the soft-classing data model and enabling approved personnel of the domain to modify the domain data model without need of a traditional database administrator and without need to modify the soft-classing data model itself. The database schema is designed flexible enough that it can incrementally store and maintain all user-defined domains, class hierarchies and class instances. It is also designed to accommodate revision to UCLP or replacement with other XML markup languages if the end users see the need to build the new or enrich the existing soft-classing scheme.

An example of an ontology captured with the Oracle implementation of soft-classing is shown in Figure 1. For each class node in the ontology, node properties define the ontology structure (such as whether the node is part of a type-of or composed-of branch, parent-child relationships) and class properties define attributes whose data values describe class instances. Class instances are only allowed at composed-of nodes, at the bottom of type-of branches, and at leaf nodes of the hierarchy. Thus, class instances could represent a number of guided missiles, explosive warheads, or antenna, but a proximity fuze would not have instances because more specificity is

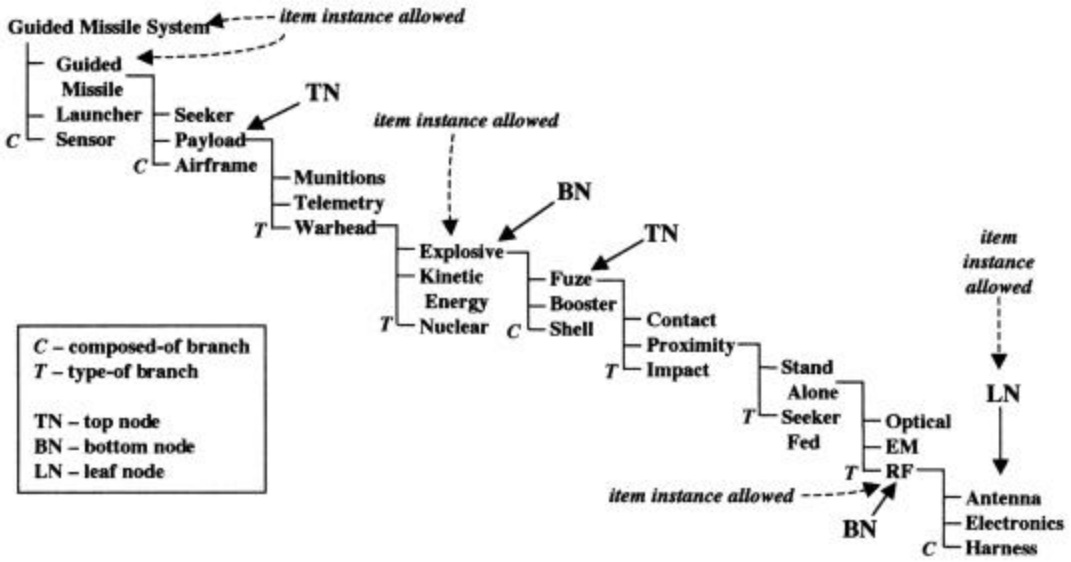

Figure 1 - Sample ontology showing a mixture of type-of and composed-of nodes

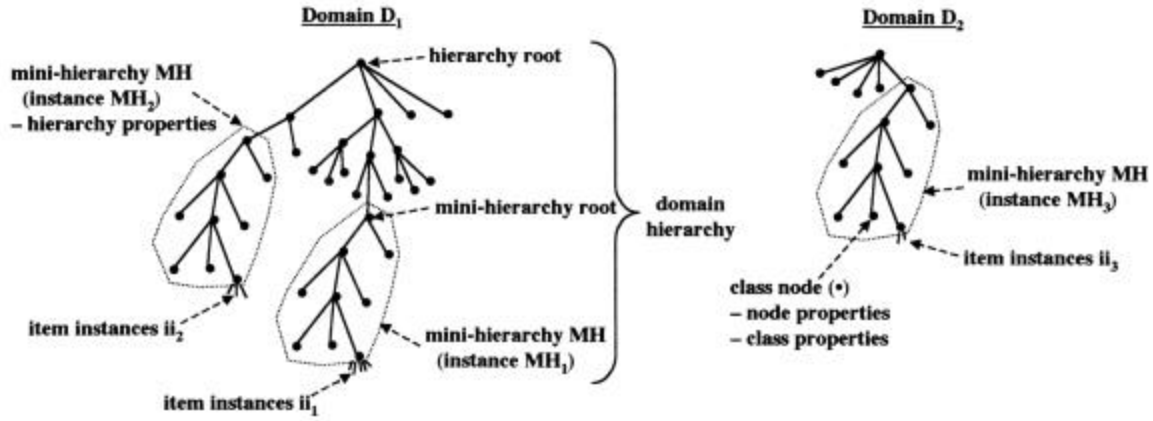

Figure 2 - Example of instantiation of mini-hierarchies

possible in the subsequent type-of breakdown. Inheritance of properties is possible through the type-of structure while the values of such properties as weight could be aggregated through the composed-of structure.

Soft-classing also supports the definition and instantiation of mini-hierarchy, i.e. branches of the ontology which can be reused within the full ontology or across ontologies whose interests overlap.

Instance data is aggregated to the Soft-Classing metadata repository through use of a crawler or other technology. The aggregated data can be stored and indexed in the Oracle8i for future searches. In addition, the searched results can directly link to the original Web sites for further data gathering if necessary. A major advantage of using Oralce8i to store the ontology and the instance metadata is that it can store data in different formats, such as integer, integer range, double, double range, string, date and arrays. Therefore, it can support most business required precise and parametric or probablistic searches that are not possible with the conventional Web text-based searches.

SUMMARY

SAIC has developed the soft-classing methodology over the past several years and is in the process of implementing in Oracle8i to benefit from greater scalability and increased flexibility with respect to evolving the XML tagging itself. UCLP 3.0 was acknowledged as a Submittal by the W3C and SAIC has continued development to increase capability and support greater use of agent technology. Other supporting tools are also in development as part of the MISTI suite.

REFERENCES

1. Universal Commerce Language and Protocol, version 3.0. Submission to W3C http://www.w3.org/Submission/1999/02/

2. "Multi-Industry Supply-Chain Transaction Infrastructure;" R. Chipman, K. J. Laskey; CALS Expo International & 21st Century Commerce 1998, San Diego CA, October 1998.