Robert E. Kent  
Washington State University, WA, USA

Organizing Conceptual Knowledge Online: Metadata Interoperability and Faceted Classification

Abstract: Conceptual Knowledge Markup Language (CKML), an application of XML, is a new standard being promoted for the specification of online conceptual knowledge (Kent and Shrivastava, 1998). CKML follows the philosophy of Conceptual Knowledge Processing (Wille, 1982), a principled approach to knowledge representation and data analysis, which advocates the development of methodologies and techniques to support people in their rational thinking, judgement and actions. CKML was developed and is being used in the WAVE networked information discovery and retrieval system (Kent and Neuss, 1994) as a standard for the specification of conceptual knowledge.

1. Future of the Web

The World Wide Web Consortium (W3C), a standards body that produces specifications and reference software for the Internet, wants to invent a new Web using eXtensible Markup Language (XML) as syntactic base and Resource Description Framework (RDF) (Lassila et al, February 1998) for the semantics of information. XML is a metalanguage used to define other markup languages, which are called its applications. RDF, a generic markup language, is such an XML application. In order to develop an appropriate metadata architecture, in addition to RDF the W3C plans to create:

1. A language for writing RDF schemata (RDFS).
2. A language for expressing information filters (simple Boolean functions of) RDF.

A working draft on RDF Schemas has recently been initiated (Brickley et al, April 1998). The W3C recommendation Platform for Internet Content Selection (PICS) is a standard for metadata specification and information filtering. The second version of PICS will be designed as a particular application of RDF.

However, one important fact should be mentioned: These ideas have already been accomplished in CKML and its subset called Ontology Markup Language (OML). These XML applications were designed during development of the WAVE system discussed below. For the following reasons, the standards, principles and techniques that have been developed in the WAVE initiative, provide a much better solution than RDF and PICS for the representation and management of knowledge on the Internet.

The ontologies of OML provide an enumerative classification mechanism that, until publication of the working draft on RDF Schemas, had been entirely lacking in RDF. The recent development of schemas has brought RDF almost up to the level of OML. What is now lacking, as of May 1998, are the n-ary relations, functions and inferencing assertions contained in OML.

The n-ary relation, function and assertional inference capabilities in OML provide the researcher with a richer means for expressing the entity relationships and constraints in their domain of inquiry.

The synthetic-faceted classification mechanism of conceptual scaling and the concept space is entirely lacking in RDF and PICS. Conceptual scaling is specified in the difference CKML minus OML, and the faceted concept space is constructed by evaluating these scales against the object metadata in OML.
2. The WAVE System

We are developing a prototype distributed knowledge management system for Web resource, which uses principles and techniques from conceptual knowledge processing. These techniques incorporate ideas from both traditional and contemporary approaches which deal with issues of knowledge organization. This includes, the enumerative and synthetic-faceted classification schemes of library science (e.g. DDC, LC, Colon), the metadata gathering and knowledge broker administration architecture recently designed by networked information discovery and retrieval systems (e.g. Harvest), and the standards currently being recommended by the World Wide Web Consortium (PICS, RDF).

An information management software system for the World Wide Web called WAVE (Web Analysis and Visualization Environment) is currently under development (Kent and Neuss, 1994). WAVE is a third generation World Wide Web tool used for conceptual navigation and discovery over a universe of networked information resources. The general goal of the WAVE project is conceptual organization of a community's information space on the World Wide Web. The WAVE system fuses the current NIDR system technology with a mechanism for "dynamic distributed classification." The project seeks to address the following research question: "What is the appropriate architecture for a digital library?" The research goal of the project is to demonstrate in the distributed context of the World Wide Web that the WAVE system, using both the technique of automatic classification and the notion of conceptual space, provides the kernel architecture for a digital library.

Figure 1 is a diagram of the architecture of the WAVE system. This consists of three major components (the digital object store, the metadata object store, and the conceptual space), and three processes (metadata abstraction, conceptual scaling, conceptual browsing) which connect the components.

![Diagram of WAVE System Architecture](image)

**Figure 1: Architecture of the WAVE System**

The digital object store (the world) represents the information space for a community on the World Wide Web as stored in various Web document collections or online databases. The metadata object store (CKML) represents information abstracted from the digital object store. The process of metadata abstraction includes extraction from raw HTML (HyperText Markup Language) documents or translation from annotated XML (eXtensible Markup Language) files, both done by Web robots. HTML is the current lingua franca of the World Wide Web, and XML is a data format for structured document interchange on the World Wide Web. More precisely, XML is a metalanguage used to define markup languages, which are called XML applications. The WAVE system represents its metadata in a markup language (an XML application) called Ontology Markup Language (OML); OML owes much to pioneering efforts of the SHOE initiative at the University of Maryland at College Park. OML is a semantic data model - an extended form of the entity-relationship model of database theory. OML represents information in terms of abstract objects and relations between those objects. The type structure of this information is specified in OML by an ontology, which consists of a generalization-specialization hierarchy or taxonomy of categories (object types) and relational schemata (relation types) between those categories.
The conceptual space is a form of conceptual knowledge representation of the original information. It organizes the ontological information of the metadata object store in terms of the formal concepts of formal concept analysis. The conceptual space represents information conceptually scaled from the metadata object store. The process of conceptual scaling applies conceptual scales to the metadata captured in the ontological categories and relations of OML, thereby transforming it into a framework suitable for the concept lattices of formal concept analysis (Wille, 1982). An extension of OML called Conceptual Knowledge Markup Language (CKML) specifies in the WAVE system the conceptual scales used in the scaling process. The conceptual space component is essentially a concept lattice with a naming facility for book-marking favorite formal concepts. Book-marked concepts are called conceptual views.

3. Facet Classification and Conceptual Scaling

The application of facets in the theory of library classification was first tested and developed by Ranganathan in his Colon classification system. Faceted analysis and classification provides a flexible means to classify complex, multi-concept subjects (Rowley, 1987). Complex subjects are divided into their component, single-concept subjects. Single-concept subjects are called isolates. Faceted analysis examines the literature of an area of knowledge and identifies its isolates. A facet is the sum total of isolates formed by the division of a subject by one characteristic of division. Some examples of facets in musical literature are: composer, instrument, form, etc. Isolates within facets are known as foci.

Comparing these ideas to ontologically structured metadata, facets are identified with conceptual scales or linguistic variables and are often associated with a composite description function, and isolate/foci are identified with scale attributes or linguistic values. A composite description function in the ontology may consist of one function with primitive image values, or a binary relation connecting two categories of objects composable with such a function, or something more complex. In conceptual knowledge processing a conceptual scale computes each facet. A conceptual scale is an active filter or lens through which information is interpreted. Faceted analysis is conceptual scaling. It involves four steps.

1. Gather ontologically structured metadata.
2. Identify conceptual scales of interest and specify attributes within them (abstract conceptual scale).
3. Specify the structure of conceptual scales (concrete conceptual scale).
4. Apply the conceptual scales to the metadata, producing a composable vector of facets (realized conceptual scales) which constitutes the conceptual space.

Conceptual scales come in three types: abstract, concrete and realized. These types are arrayed along an intentional-extensional dimension. Abstract conceptual scales introduce terms (attribute names), and abstractly specify attribute definitions via term-to-term relationships, as structured by a concept lattice. This lattice is specified by a set of term implications and term disjointnesses. Concrete conceptual scales attach meaning and provide definition to the terms in abstract scales by binding a (single-variable) query to each term. These queries are required to respect the abstract term relationships (implications and disjointnesses). Concrete conceptual scales are conceptual filters or lenses. As portrayed in the diagram on the right, when applied to a collection of objects (ontologically structured metadata), concrete conceptual scales produce facets, which are known in conceptual knowledge processing as realized scales. Each facet is a concept space representing a single dimension of the metadata. A faceted conceptual space is a vector of facets, and conversely, facets are the components that underlie a faceted conceptual space.
The process of conceptual scaling is of several kinds (Ganter and Wille, 1989) (Neuss and Kent, 1995): nominal, ordinal, hierarchical, universal, etc. The most common kinds are "nominal" and "ordinal". Nominal scaling refers to names that are members of an unordered set, whereas ordinal scaling refers to sub-ranges of a totally ordered set. Conceptual scaling kind is concentrated in the arguments of the various relations that are associated with the scale. The table on the right describes the various kinds of conceptual scale according to mathematical structure and purpose or use. Mathematical types of scales represent intuitive ideas of design.

<table>
<thead>
<tr>
<th>relation</th>
<th>function</th>
</tr>
</thead>
<tbody>
<tr>
<td>isa:</td>
<td>source:</td>
</tr>
<tr>
<td>Category</td>
<td>BinaryRelation → Type</td>
</tr>
<tr>
<td>Category</td>
<td></td>
</tr>
<tr>
<td>instance:</td>
<td>target:</td>
</tr>
<tr>
<td>Object → Category</td>
<td>BinaryRelation → Type</td>
</tr>
<tr>
<td>argument:</td>
<td>fnSource:</td>
</tr>
<tr>
<td>Relation → Type</td>
<td>Function → Type</td>
</tr>
<tr>
<td>name:</td>
<td>fnTarget:</td>
</tr>
<tr>
<td>Entity → String</td>
<td>Function → Type</td>
</tr>
<tr>
<td>description:</td>
<td>argument:</td>
</tr>
<tr>
<td>Entity → String</td>
<td>Relation → Type</td>
</tr>
</tbody>
</table>

Table 1: Signatures for Relations and Functions
<Ontology name = "Base" version = "0.20">
<Description>
This is the base ontology for the Ontology Markup Language.
</Description>
<Category name = "Entity"/>
<Category name = "Type"/>
<Category name = "Category"/>
<description>
This is the generic concept of a class of objects. When an ontology defines a new category, the object representing that category must be declared as an instance of this object.
</description>
</Category>
<Category name = "Relation"/>
<Category name = "BinaryRelation"/>
<Category name = "Fn2Rel"/>
<Category name = "Function"/>
<Category name = "Object"/>
<Category name = "Relationship"/>
<Category name = "Collection"/>
<Category name = "Ontology"/>
<Category name = "DataType"/>
(DataType name = "String")
(DataType name = "Date")
(DataType name = "Boolean")
<Value name = "true"/>
<Value name = "false"/>
</DataType>
(DataType name = "Natno")
(DataType name = "Integer")
(DataType name = "Real")
<isa specific = "Type" generic = "Entity"/>
<isa specific = "Category" generic = "Type"/>
<isa specific = "DataType" generic = "Type"/>
<isa specific = "Relation" generic = "Entity"/>
<isa specific = "BinaryRelation" generic = "Relation"/>
<isa specific = "Fn2Rel" generic = "BinaryRelation"/>
<isa specific = "Function" generic = "Entity"/>
<isa specific = "Object" generic = "Entity"/>
<isa specific = "Relationship" generic = "Object"/>
<isa specific = "Collection" generic = "Entity"/>
<isa specific = "Ontology" generic = "Entity"/>
<BinaryRelation name = "isa" source = "Category" target = "Category">
<description>
"This indicates the inclusion relation between categories. It is a reflexive, transitive binary relation between categories."
</description>
</BinaryRelation>
<BinaryRelation name = "instance" source = "Object" target = "Category">
<description>
"This indicates that an object is a member of a category, and thus has all of the properties of that category. It is a binary relation between objects and categories, which is closed on the right: If

Table 2: Serialization of the OML Base Ontology as an Ontology

4. Conceptual Knowledge Markup Language

Conceptual Knowledge Markup Language (CKML) (Kent and Shrivastava, 1998) provides a specification standard for the conceptual representation and analysis of networked resources. Although hypertext links enable organization in the small, resource discovery
systems, using content-based access to documents, enable organization in the large (Bowman, 1996). Content-based access requires a good representation for document content and the natural hierarchies associated with documents and related entities (Kent and Bowman, 1995). Ideally, such document content and natural hierarchies would be transparently composable. Concept lattices allow such transparent composability by representing both content and entity type hierarchies as facets of document information.

Concept lattices are lattices with bound objects and attributes. Any arbitrary lattice is a concept lattice, where the objects and attributes bound to it are defined mathematically - they are the atomic (irreducible) elements with respect to joins and meets. Object-attribute incidence relations, called formal contexts, and concept lattices are equivalent structures. Either provides for the conceptual representation of networked resources. When applying conceptual knowledge processing to networked resources, objects are represented as abstracted ontologically structured metadata and attributes are defined by logical queries. Concept lattices, and their equivalent formal contexts, are constructed by the process of conceptual scaling. Conceptual scaling transforms ontologically structured collections of objects to faceted conceptual space by applying conceptual scales.

5. OML and Metadata Interoperability

The digital library community uses the term interoperability to denote the general
problem of linking different systems. The three levels of interoperability are: technical interoperability, informational interoperability, and social interoperability. We propose to study informational metadata interoperability. We are studying informational metadata interoperability. Our study of metadata interoperability is concerned with the connection between RDF with schemas and the OML. We are currently in the process of defining grammar-based translations between RDF/Schema and OML, and are investigating the role of reification in translation. Comparisons between RDF and OML essentially revolve around the connection between the RDF core schema (Brickley et al, April 1998) and the OML Base ontology diagramed in Figure 2.

Figure 2 illustrates the concepts of type, subtype, and entity in OML. A type is depicted by a rounded rectangle, an entity by a large dot. In Figure 2 arrows represent definition: they are drawn from an entity to the class that it defines. A dot appearing within a class rectangle depicts the membership relation, and the subtype relationship (containment) is depicted by complete enclosure. Sets in Figure 2 with uncolored backgrounds are partitions of their contained colored subsets. The relations and functions in Figure 2 have the signatures listed in Table 1. Two alternate serializations are possible: Table 2 gives a serialization of the base OML ontology as an ontology, whereas Table 3 gives a serialization of the base OML ontology as object collections. In addition, inferencing can be represented in OML. The RDF core schema specification mentions that the RDF subclassOf relation is transitive. Transitivity of the corresponding OML isa binary relation is expressed as an assertion in Table 4.

```
<Assertion kind = 'expression'>
 <Forall name = 'c1,c2,c3' type = 'Category'>
  <implies>
   <and>
    <BinaryRelation name = 'isa' source = 'c1' target = 'c2'/>
    <BinaryRelation name = 'isa' source = 'c2' target = 'c3'/>
   </and>
   <BinaryRelation name = 'isa' source = 'c1' target = 'c3'/>
  </implies>
 </Forall>
```

Table 4: Assertion Expressing Transitivity of ISA

6. Summary and Future Work

In this paper we have given an overview of conceptual knowledge processing, explained the motivation behind the design of the conceptual knowledge markup language, discussed the process of conceptual scaling in an online environment, explained the translation interoperability between CKML and the metadata standard RDF with Schemas, and discussed the online WAVE conceptual browser.

In future work we would like to apply established knowledge representation methodologies, such as the conceptual descriptions of Bill Woods (Woods 1991, 1997) and conceptual graphs of John Sowa (Sowa 1984), to the specification of metadata and profiles. This will involve further development of a markup language for conceptual graphs (CGML), and investigation of its interoperability with OML (and RDF).
References


