MAKO: Multi-Ontology Analytical Knowledge Organization based on Topic Maps

A Riki Y. Morikawa and Larry Kerschberg
E-Center for E-Business, Department of Information and Software Engineering, George Mason University
Fairfax, Virginia, USA

Abstract

This paper addresses how the XML Topic Map (XTM) 1.0 standard can be used to develop an analytical knowledge base comprised of multiple ontologies to support intelligence assessments. Termed the Multi-Ontology Analytical Knowledge Organizational (MAKO) framework, it incorporates a Multidimensional Ontology Model (MOM) that organizes subjects into separate conceptualizations based upon common-sense groupings. Topic, association and occurrence elements are temporally serialized, according to the Temporal Layer Model (TLM), to accommodate, and historically preserve, modifications to the knowledge base as world events change.

Introduction

Today's intelligence analyst has access to an overwhelming amount of data and information. Often times, they are asked to provide quick turn around assessments regarding a particular issue. When this occurs, they must rely heavily upon their own tacit knowledge, knowledge of fellow colleagues, and information contained in databases. However, sifting through disparate databases and online documents is time consuming and labor intensive.

In order for analysts to make assessments that are defensible, they need access to a knowledge base of information that is organized in a meaningful way by fellow analysts and subject matter experts. The knowledge base represents the semantic understanding of a particular subject as seen by subject matter experts. As a shared resource, a knowledge base built upon a semantic standards helps to ensure interoperability between intelligence organizations.

This paper introduces the Multi-Ontology Analytical Knowledge Organization (MAKO) Framework, which is built upon the XML Topic Map version (XTM) 1.0 standard. MAKO organizes information into a semantically connected cross-matrix of ontological domains and temporal concepts. Analysts can then conduct inter-ontology searches, and build assessments using topics, associations and occurrences that can be easily shared with other analysts.

Organization of Paper

Section 2 provides a brief description of the Multidimensional Ontology Model (MOM). The Temporal Layer Model and the concept of temporal serialization are discussed in section 3. Section 4 presents our conclusions and suggestions for future work.

Related Research

Both the Web and artificial intelligence communities have been working in the areas of knowledge management (KM). Many knowledge management systems have incorporated standardized formats or protocols at varying levels of design and implementation; however, interagency sharing or use of multiple integrated ontologies has not been a major focus in many of these efforts [1-5].

Other knowledge architectures exist such as the Knowledge Management Framework, [6]. Within this conceptual architecture, the MAKO framework fits nicely into the Information Management Layer.

Although the XTM standard does not specifically address ontological declaration, the use of topic templates has been offered as a possible solution [7]. The MAKO framework leverages this concept and expands it by describing a matrix of multiple interconnected ontologies.

Multi-Ontology Analytical Knowledge Organization Framework

The MAKO framework, shown in Table 1, is a cross matrix of several major concepts. The Multidimensional Ontology Model decomposes the knowledge base into separate conceptualizations based upon common sense groupings. The Temporal Layer Model segregates topics according to their time intervals, while the Interpretive Layer represents an analyst's short-term assessments.

Table 1. MAKO Framework

Temporal Layer Model		Multidimensional Ontology Model: CATEGORICAL	STAKEHOLDER	GEOGRAPHICAL
	INTERPRETIVE LAYER	ASSESSMENT SCENARIO		
	OCCURRENT LAYER	DYNAMIC SUBJECTS	DYNAMIC ENTITY	DYNAMIC LOCATIONS
	CONTINUANT LAYER	STABLE SUBJECTS	STABLE ENTITY	STABLE LOCATIONS

Multidimensional Ontology Model

An ontology is defined as a "specification for a conceptualization". The conceptualization is comprised of a set of concepts, or classes, that relate to one another in some logical fashion. In essence, an ontology describes concepts that are in a domain of discourse. We use the term *Semantic Layer* to refer to an Ontology that has been populated with class instances. The Semantic Layer, combined with rules and constraints, constitute the beginnings of a knowledge base [8, 9].

In terms of the Multidimensional Ontology Model, a set of *root* ontologies are integrated through standard association interfaces. There are several qualities that a *root* ontology possesses in our model. The first quality is *independence*, which means that the ontology can represent a complete domain of knowledge without depending upon other domains (e.g. temporal, geographical domains). The second is *reusability*, where the root ontology can be reused by other knowledge bases with minimal, if any, modification. While reusability is required for all reference ontologies (e.g. temporal), it is not a necessary condition for other ontologies that may be representing highly specialized knowledge domains. Third, root ontologies must have *standardized interfaces* that enable them to be connected to other root ontologies. Finally, root ontologies can be specialized (or decomposed) into multiple sub-ontologies. The following definitions describe the ontological relationships within our model.

Definition 1. The Multidimensional Ontology Model is defined by a set of interrelated root ontologies, *rootOnt*. A *rootOnt* is defined by a set of interconnected *topic*, *association*, and *role* classes. Each *rootOnt* can be further divided into multiple sub-ontologies, *subOnt*.

Let TC, AC, and RC be sets of all topic, association, and role classes contained within the knowledge base, and $TC_i \subset disjointP(TC)$, $AC_i \subset P(AC)$, and $RC_i \subset P(RC)$, where i represents the i^{th} root ontology, disjointP(TC) represents the power set of TC such that no topic class is contained in more than one set within the powerset, and P represents the power set.

$rootOnt_i$ defined by $\{TC_i, AC_i, RC_i\}$

$rootOnt_i = \{subOnt_{(i,1)}, subOnt_{(i,2)}, subOnt_{(i,3)}, \ldots, subOnt_{(i,n)}\}$

Definition 2. *Complementary* ontologies are defined as two or more root ontologies that are interrelated to one another, thereby serving to enhance the overall knowledge represented. Complementary ontologies are connected through a subset of ontology defining association and role classes.

Let $rootOnt_i$ and $rootOnt_{(i+1)}$ be complementary ontologies that are connected through the subsets AC' and RC' where $AC' \subset AC_i$, $RC' \subset RC_i$, for the i^{th} root ontology

$[rootOnt_i \cap rootOnt_{(i+1)}] = \{AC', RC'\}$, where $AC' = AC_i = AC_{(i+1)}$

Definition 1 states that a root ontology is comprised of a set of XTM class elements. The interface between root ontologies, per definition 2, represents a subset of association and role classes that are common to two or more interrelated ontologies. Therefore, by standardizing AC' and RC', you essentially standardize the interfaces to the root ontologies. For example, standardized association classes to a temporal ontology might include "(Open or Closed) Time Interval", "Occurs (Before, After, or During) Time Point". These standardized classes represent the standard interfaces used to connect to the temporal reference ontology. In order to reduce ambiguity and misinterpretation when integrating ontologies, each AC' and RC' is defined by either a namespace or Published Subject Indicator (PSI).

Rath [7] describes a technique by which ontology topics can be represented within a topic map template[1]. The current ISO 13250 and XTM standards do not specify a mechanism for declaring a topic map ontology. As such, an ontological structure is not defined until it has been referenced by a class-instance within the topic map. The topic map template, which is uniquely identified by a PSI, provides the mechanism by which topic classes can be declared ahead of time. We adopt this concept to declare multiple complementary ontologies that are connected together through association and role class elements.

By treating all class elements initially as topics, we are also able to explicitly declare class properties such as data types [10] and constraints[2]. In our model, we declare properties by using the topic's occurrence child element, which refers to PSIs for explicit property definitions.

An example knowledge base was developed to test certain aspects of the MAKO concept. In our knowledge base, titled "World Oil Market Analysis", four integrated root ontologies were specified - Categorical, Stakeholder, Geographical, and Temporal. Although we chose these ontologies, it is emphasized that any set could have been specified.

---

[1] *Topic map template* is a "semi-official" term used by the ISO working group to refer to the ontology portion of a topic map.

[2] The proposed XTM standard does not formally address data typing or constraint identification.

As a metadata framework, multiple ontological and semantic layers can be defined with no adverse impact to the underlying resources. Herein lies the strength of applying the XTM standard as the basic building block.

Sub-Ontology Organization within the Root Ontology

Sub-ontologies form a hierarchical tree structure where each branch represents a child sub-ontology. Taxonomic paths are explored by tracing paths from the root to each leaf. The leaf node represents a topic class or instance that is contained within a sub-ontology.

The taxonomic structure in this model differs from traditional hierarchical taxonomies in two basic ways (see Figure 1):

1. Taxonomies are created in a "top-down" and "bottom-up" fashion. This means that taxonomic paths are first created in the traditional manner from root to leaf. This is followed by a "bottom-up" pass where leaf nodes are allowed to connect to other branches as appropriate. Doing this helps to prevent "dead-end" searches where few alternatives are available once a user has reached the leaf node.
2. Because leaf nodes are allowed to attach to any branch, strict inheritance through superclass-subclass associations is not always possible. As a consequence, we adopted two additional types of relationships for use in our taxonomic hierarchy. The first is a *mereological* association, "part_of", with the role types "part" and "whole". The second is an *aggregation* association, "member_of", with the role types "member" and "community". With the aggregation association, any topic can be members of any aggregated set. Membership is determined solely by the user.

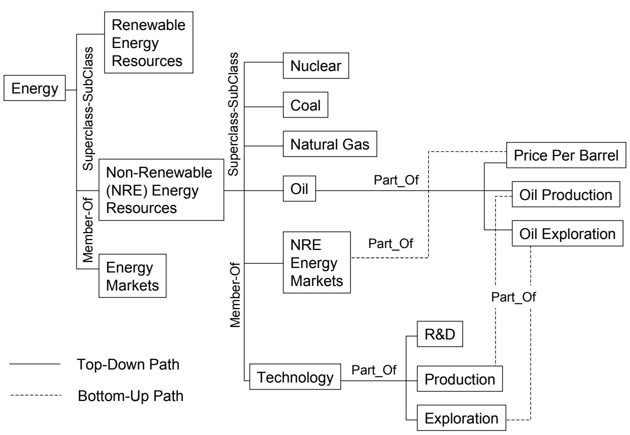

Figure 1. Taxonomic relationships in the "World Oil Market Analysis" example are constructed using a top-down and bottom-up classification approach.

The Categorical Ontology and Semantic Layer. The Categorical Ontology covers the main concepts that comprise the knowledge base. Topic instances within this layer reify subjects and play roles as members within an association. Roles can be assigned in such a way as to imply directionality between member topics.

The Geographic Ontology and Semantic Layer. The Geographic Ontology describes physical features and national boundaries. It is used to provide information regarding a topic of interest. For example, if discussing oil exploration in the Caspian Sea, this layer will indicate that Russia and Iran are two coastal nations (see Figure 2). Following semantic links provides some insights about these neighbors and OPEC.

The Stakeholder Ontology and Semantic Layer. The Stakeholder Ontology is comprised of government, corporation, and organizational sub-ontologies. It provides the analyst with an overall understanding of the economic or political factors surrounding stakeholder motivations. For example, understanding Russia's economy and oil industry gives insight on how Russia might react to an OPEC request to reduce oil production (see Figure 2).

The Temporal Layer Model. A key feature of the MAKO framework is the Temporal Layer Model (TLM), which provides a temporal ordering relation for all knowledge base components. The TLM segregates time into continuant and occurrent classes. Continuant objects are stable topics and associations that do not change over the valid time interval of the knowledge base. Occurrent objects are created, destroyed, or their definitions modified during the valid time interval [11]. For knowledge bases that are populated with numerous occurrent elements, the temporal reference helps to preserve the historical mapping between elements for any time frame of interest.

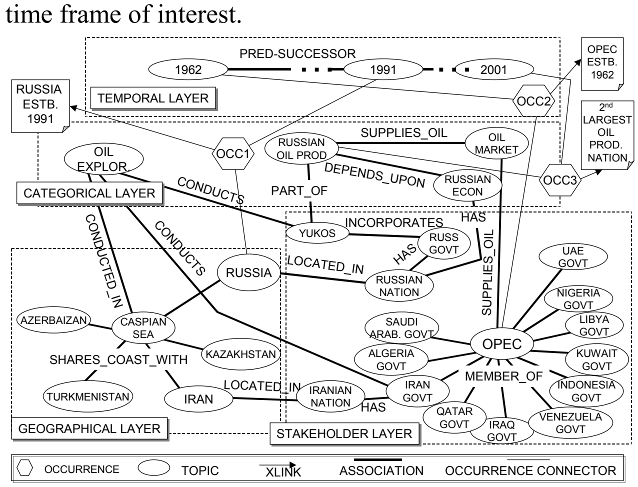

Figure 2. The Geographical Layer can provide semantic connections of interest.

Temporal Serialization

The Base Timeline (BT) consists of the set of continuant and occurrent objects that have been serialized over the valid time line of the knowledge base. The BT is defined by t_{open}, and t_{close}, or $t_{current}$, time points. The closing time point is used when a knowledge base covers a finite time period. The current time point is used for knowledge bases that are continually updated with new information. In the topic map depicted in Figure 3, the valid time interval of the knowledge base is between 1900 and 2002.

The following definitions will help us develop a serialization methodology for our example in Figure 4.

Definition 3. Baseline Time (BT) represents the topic map's overall valid time interval (t_{open}, t_{close} or $t_{current}$) and is defined by the opening valid time t_{open}, and either the closing or current valid time, t_{close} and $t_{current}$ respectively. (e.g. t_{open} = 1900 and t_{close} = 2002). BT is expressed by the following set members:

$$BT = \{x_i(t_{Open}),\ x_i(t_{close} \text{ or } t_{current}),\ y_j(t_k),\ y_j(\Delta t_k),\ A_r(T,\ t_a),\ A_r(T,\ \Delta t_a)\}$$

The first two elements of the set BT, represents the opening, t_{open}, closing, t_{close}, and current , $t_{current}$, end points for a continuant variable, $x_i(t)$. The third and fourth elements represents occurrent topic, y_j, with opening, t_k, and closing, Δt_k, time points for the interval (t_k, Δt_k). The final two elements represent association elements, A_r, with opening, t_a, and closing, Δt_a, time points.

Definition 4. The variable $x_i(t)$, represents a *continuant topic*, where $x_i(t_{open}) = x_i(t_{close}$ or $t_{current})$. Continuant objects, and the subjects they reify, are valid for the duration, although some of their properties and occurrence elements are allowed to change over time. As topic classes, they provide the anchors that stabilize ontologies within the knowledge base. As topic and association instances, they are the fundamental building blocks from which semantic understanding is based.

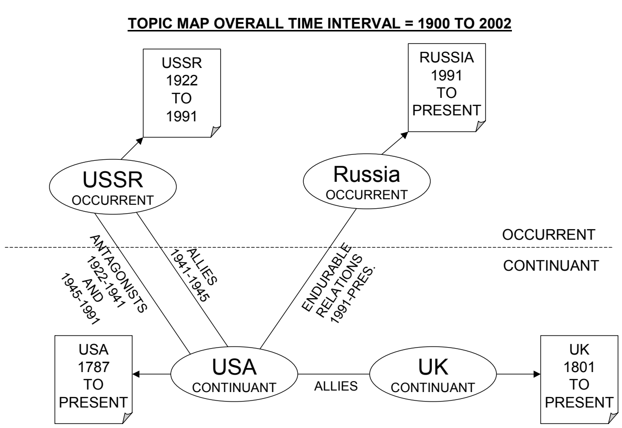

Figure 3. The Base Timeline (BT) consists of topics and associations from the occurrent and continuant layers.

Definition 5. An occurrent object is defined by the variables $y_j(t_k)$ and $y(\Delta t_k)$, which represent the opening and closing time points , and where ($tk+\Delta t_k$) represents the valid time interval for y_j such that ($t_{open} \leq (t_k + \Delta t_k) \leq$ (t_{close} or $t_{current}$)).

By distinguishing continuant topics and associations from occurrent ones, several advantages are realized. First, since the continuant layer consists of elements that do not change during the valid time interval of the knowledge base, once serialized, they will not have to be re-serialized when new occurrent elements are added or modified. Second, the continuant layer provides stable knowledge that represents the fundamental principles by which all assessments can be reliably based. Third, the occurent layer, allows the knowledge base to be flexible (i.e., addition of new information). This enables capture of historical knowledge, which allows analysts to search past "snapshots" of the knowledge base. This also offers some unique possibilities for multivariable trend analysis.

Definition 6. Let A denote the total set of topic map association elements, such that $A_r \in A$, and A_r is the r^{th} association contained within the knowledge base. Associations are completely dependent upon the topics to which they are connected, as such, they are identified together during serialization. Let T denote a set of continuant or occurrent topics that are incident to A_r. Associations are then represented by $A_r(T,t_a)$ and $A_r(T,\Delta t_a)$, which denote opening and closing valid times, respectively.

Definition 7. *Serialization* is defined by the partially ordered pair (BT, $<_t$), where BT is the baseline time as defined in Definition 3, and $<_t$ is the ordering relation over time.

Example. Given the topic map in Figure 3, we proceed to serialize topics and associations. We start by identifying topics, associations and their respective time interval end points.

Let: x_1 = USA, x_2 = UK, y_1 = Russia, y_2 = USSR;
A_1 = "Allies", A_2 = "Antagonists", A_3 = "Endurable Relations";
T_{open} =1900, t_1 = 1922, t_2 = 1941, t_3 = 1945, t_4 = 1991, t_{close} = 2002;

The topics and associations above are then temporally *serialized* according to the TLM, which provides the ordering reference. Figure 5 graphically depicts the final serialization. Below is the linear representation.

$[x_1(t_{open}),x_2(t_{open}),y_1(t_{open}),A_1(x_1,x_2,t_{open})]<[y_1(t_1),y_2(t_1),A_2(x_1,y_2,t_1)]<[A_2(x_1,y_2,t_2),A_1(x_1,y_2,t_2)]<[A_1(x_1,y_2,t_3),A_2(x_1,y_2,t_3)]<[y_1(t_4),y_2(t_4),A_2(x_1,y_2,t_4),A_3(x_1,y_1,t_4)]<[x_1(t_{close}),x_2(t_{close}),y_1(t_{close}),A_1(x_1,x_2,t_{close}),A_3(x_1,y_1,t_{close})]$

From the serialization shown in Figure 4, we see that more than one node can share the same point in time. This is due to the coarse granularity of the time scale. As we refine granularity, the number of concurrent events decreases and the serialization is refined. Although concurrent events do not invalidate serialization, the lowest feasible granularity should be selected to improve temporal inferencing.

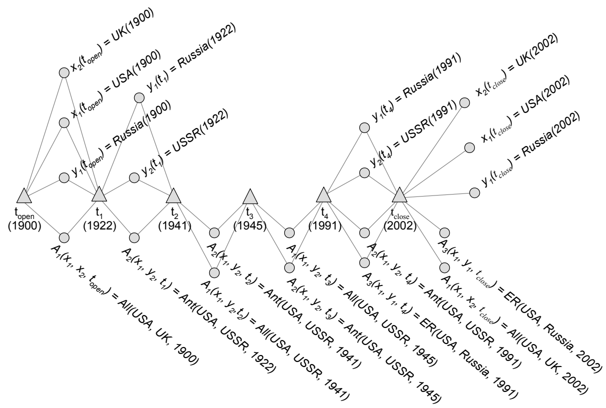

Figure 4. Graphical depiction of temporal serialization

By serializing the entire knowledge base, the correct topic mapping for any time point or interval can be presented. This helps analysts understand an overall sequence of events that, in turn, helps them identify and assess causal relationships.

Conclusions and Future Work

This paper has presented the Multi-Analytical Knowledge Organization (MAKO) framework based on XML Topic Maps (XTM). The Multidimensional Ontology Model (MOM) can be used to organize topics according to complementary domain-specific viewpoints. The Temporal Layer Model (TLM) serves as a reference for the knowledge base by which objects can the serialized. Serialization enables temporal inference, constraint management based upon time, and the capture of historical knowledge in the form of semantic relationships.

By combining these elements into an analytical knowledge base, an analyst can search through rich ontological structures that have been created by subject matter experts. By using XTM as a basic semantic building block for the knowledge base, an organization can capture and share a subject matter expert's knowledge through interconnected topics, associations and roles.

This paper has only touched upon some of the Web standards that are currently being developed. The Web Ontology Language (OWL) [12] in conjunction with either XTM or RDF should be investigated as complementary ways to define ontological structures within a multidimensional knowledge base. Use of analytical namespaces and/or Published Subject Indicators (PSI) should be further incorporated to enable inter-organizational knowledge sharing.

Acknowledgments. This work was sponsored by a NURI from the National Geospatial-Intelligence Agency (NGA). This work was also supported in part by the Advanced Research and Development Activity (ARDA). Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the U. S. Government.

References

[1] A. Agostini, S. Albolino, R. Boselli, G. D. Michelis, F. D. Paoli, and R. Dondi, "Stimulating Knowledge Discovery and Sharing," presented at 2003 International ACM SIGGROUP Conference on Supporting Group Work, 2003.

[2] J. Golebiowska, R. Dieng-Kuntz, O. Corby, and D. Mousseau, "Building and Exploiting Ontologies for an Automobile Project Memory," presented at International Conference on Knowledge Capture, Victoria, British Columbia, Canada, 2001.

[3] D. Fowler, "Harnessing the Power of Knowledge Management," in netWorker, ACM Press, vol. 5, 2001, pp. 18-24.

[4] J. Davies, A. Duke, and Y. Sure, "OntoShare: a knowledge management environment for virtual communities of practice," presented at International Conference on Knowledge Capture, Sanibel Island, FL, USA, 2003.

[5] C. Weng, M. Kahn, and J. Gennari, "Temporal Knowledge Representation for Scheduling Tasks in Clinical Trial Protocols," presented at American Medical Informatics Association Fall Symposium, San Antonio, Texas, 2002.

[6] L. Kerschberg, "Functional Approach to in Internet-Based Applications: Enabling the Semantic Web, E-Business, Web Services and Agent-Based Knowledge Management," in The Functional Approach to Data Management, P. M. D. Gray, L. Kerschberg, P. King, and A. Poulovassilis, Eds. Heidelberg: Springer, 2003, pp. 369-392.

[7] H. Rath, "Topic Map Fundamentals for Knowledge Representation," in XML Topic Maps: Creating and Using Topic Maps for the Web, J. Parker and S. Hunting, Eds. Boston: Addison-Wesley, 2003, pp. Chapter 14.

[8] L. Orbst and H. Liu, "Knowledge Representation, Ontological Engineering, and Topic Maps," in XML Topic Maps: Creating and Using Topic Maps for the Web, J. Parker and S. Hunting, Eds. Boston: Addison-Wesley, 2003, pp. Chapter 7.

[9] N. F. Noy and D. L. McGuinness, "Ontology Development 101: A Guide to Creating Your First Ontology," Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880 March 2001.

[10] M. Altheim, "Datatypes for XML Topic Maps (XTM) using XML Schema Datatypes (http://kmi.open.ac.uk/psi/datatypes.html)," Knowledge Media Institute, The Open University, 2002.

[11] J. F. Sowa, Knowledge Representation: Logical, Philosophical, and Computational Foundations: Brooks/Cole Thomas Learning, 2000.

[12] S. Bechhofer, F. v. Harmelen, J. Hendler, I. Horrocks, D. L. McGuinness, P. F. Patel-Schneider, and L. A. Stein, "OWL Web Ontology Language Reference," W3C, 2003.