IDEA GROUP PUBLISHING

701 E. Chocolate Avenue, Suite 200, Hershey PA 17033, USA
Tel: 717/533-8845; Fax 717/533-8661; URL-http://www.idea-group.com

ITP5148

This paper appears in *Managing Modern Organizations Through Information Technology*, Proceedings of the 2005 Information Resources Management Association International Conference, edited by Mehdi Khosrow-Pour. Copyright 2005, Idea Group Inc.

From Ontologies to the Semantic Web: The Engineering of Knowledge

Antonio M. Lopez Jr.

Dept of Computer Sciences & Computer Eng., Xavier University of Louisiana,

One Drexel Dr., Campus Box 50, New Orleans, LA 70125, USA, tlopez@xula.edu

ABSTRACT

In any context area, knowledge management requires a holistic view of knowledge from the single, individual user to a global community of practice. This paper suggests a methodology that supports the engineering of knowledge with this vision in mind. The approach uses ontologies to structure the information so that it is comprehensible to a single, individual user at the user's level of understanding or "view". Various types of ontologies are incorporated into a Semantic Web environment allowing larger communities of practice to use them as a unified whole. The methodology is demonstrated using a complex problem space — ballistic missile defense.

INTRODUCTION

Like beauty, knowledge exists in the mind / eye of the beholder. One person's knowledge is another person's irrelevant piece of information. Although knowledge and information are closely related, they are different concepts with the latter being easier to define. Smith and Farquhar (2000) defined knowledge as information in action. Lopez (2004) highlighted the subtleties surrounding that definition and offered a somewhat expanded definition: Knowledge is the process of putting information into action to solve a problem. The information used in the process can come from more than one context area (i.e., domain). The solution to the problem creates new information that can become new knowledge when someone or something uses it to solve either the same problem at a different time, or a different problem in the same or different domain. Lopez also argued that "process" presupposes the existence of an entity capable of putting information into action and called such an entity a knowledge-holder.

In today's age, knowledge-holders can be either human or intelligent agents (For information regarding intelligent agents see Plekhanova, 2003). Consequently, a high-level view of knowledge management is the management of knowledge-holders (i.e., humans and intelligent agents) and the continued development of their knowledge. This view of knowledge management is consistent with that offered in some of the business and technical literature. Gupta and Sharma (2004) put forward knowledge management as an emerging, interdisciplinary organizational model dealing with all aspects of knowledge within the context of an organization, encompassing both its technological tools and the sociological behaviors of its members. While groupware technologies can be used, the knowledge management goals are identifying information resources, establishing the relevance of those resources to a given situation, and sharing knowledge within an organization.

Acknowledging that knowledge management is a recent area in business administration that deals with how to leverage knowledge as a key asset, Schreiber et al. (2002) tie it strongly to knowledge engineering. Knowledge engineering is a subfield of artificial intelligence that is concerned with capturing and structuring the ways in which humans put contextual information into action (i.e., knowledge) so that intelligent agents can be knowledge-holders. In the late 1970s, knowledge engineering was focused only on the development of information systems in which knowledge played an underlying role (i.e., knowledge-based systems). Today, knowledge engineering does more. It provides tools and techniques for understanding the structures and procedures that knowledge-holders use. Knowledge engineers can help identify opportunities in organizations for the development, application, and distribution of knowledge resources (Schreiber et al., 2002). Knowledge engineering has turned from trying to extract knowledge from humans to a modeling activity. The focus of knowledge modeling is the conceptual structure of knowledge. The emphasis is on the context of organizational problem solving in a real world, workplace situation. The concepts as well as the relationships between concepts reflect the real world domain and are expressed in a vocabulary that people working in that domain understand.

For years, the Department of Defense (DoD) has made use of its Secret Internet Protocol Router Network, which supports a classified World Wide Web look-a-like for the exchange of classified data and information. Now, DoD and other Federal Agencies are pursuing a holistic view of knowledge management to encompass a variety of domains (Kenyon, 2003), and they see the Semantic Web as the vehicle (Ford, 2004). Military contractors and information-technology creators not usually associated with weapon systems are working to weave weapons, intelligence and communications into a seamless web (Weiner, 2004). This paper uses unclassified segments of a military problem to demonstrate a methodology for engineering knowledge.

As a contextual information area, ballistic missile defense is one of the most scientifically complex and politically controversial problems facing research and development teams working for the DoD (Graham, 2001). The origins of the problem date back 50 or more years. In 1983 President Ronald Reagan established the Strategic Defense Initiative program that created a governmental organization whose charge was to lead efforts in finding solutions to the multifaceted problem. Over time and under different presidential administrations (both Republican and Democrat), the organization has evolved into the Missile Defense Agency (MDA), which guides numerous research teams (i.e., communities of practice) – government, industrial, and academic – in addressing different aspects of the problem. Many well-known and capable scientists (Wright and Postol, 2000; Folger, 2001) have raised objections to the research work that MDA has funded. They point to the numerous shortcomings in technology and physical capabilities. Such criticisms serve to highlight the need to engineer the knowledge that enables communities of practice to challenge assertions and revise knowledge, as knowledge is refined over time.

WHO KNOWS WHAT?

In the context of Operation Iraqi Freedom, Dr. Phillip Meilinger (2003), a defense analyst with Science Applications International Corporation, made statements that truly transcend that context. He asserts that technology has made fighting wars more precise, but that the human decision-making cycle has not kept up with the technology. In his commentary, Meilinger focused on the "sensor-to-shooter" cycle – the time necessary to spot a target, identify it as unfriendly, and destroy it. This cycle is a complex process that relies on sophisticated sensors generating data that must be gathered, collated, and analyzed then passed along to the decision-maker for authorization to destroy the target. The sensor-to-shooter cycle is nowhere more sensitive to human decision-making than when it comes to ballistic missile defense.

If a nation on one side of the world were hostile enough to launch a ballistic missile toward a nation on the opposite side of the world, it would take the missile approximately 3 minutes from launch to clear the earth's atmosphere; this is called the Boost Phase. The missile would then travel another 15 to 20 minutes in outer space; this is called the Mid-course Phase. In what is called the Terminal Phase, the missile would re-enter the atmosphere and take about 1 minute to hit its target. The total time period is shortened considerably if a short-range ballistic missile is launched from an innocent-looking merchant ship off the target nation's coast. In either case, a human decision-maker might have 5 to 10 minutes to authorize a defense (if one is available). The outcome hinges on what the decision-maker (human or intelligent agent) knows — time, space (distance, geography, atmospheric, etc.), and actionable resources.

In any hostile missile launch scenario, a multitude of sensors might be providing decision-makers at various levels with data about the missile. In the Boost Phase, the influx of sensor data would have to be organized (i.e., fused) becoming information. But knowledge is required to put that information into action giving the decision-maker at that level additional information such as what kind of missile is inbound, where is it going, and how much time before it hits. During the Mid-course Phase, the belligerent who launched the missile would most likely use counter-measures, such as decoy warheads, to increase the sensor data thus attempting to overpower the sensor fusion facilities and leave the decision-makers with more questions than answers. Again, knowledge is needed at this level, but it is different from the knowledge needed at the previous level. In the Terminal Phase, the decision-making, if there is any to be done, will be localized and depend entirely on knowledge of defense resources.

The knowledge required for each level of decision-making in the scenario described above can be left to human expertise and chance during a crisis. On the other hand, the knowledge can be engineered, tested, and re-calibrated in a non-crisis timeframe and during the crisis intelligent agents can assist human decision-makers in taking action, both understanding fully that chance still plays a role on the final outcome. Although it is beyond the scope of this paper and its underlying research to contribute directly to the enormous ballistic missile defense problem domain, the domain itself supports exploration into the proposed knowledge engineering methodology required for a holistic view of knowledge management. The methodology involves (1) the structuring of information using various ontologies, and (2) the sharing of knowledge between communities of practice using the Semantic Web.

STRUCTURING INFORMATION

Ontologies have been an area of research since the early 1990s. A short, popular definition of ontology dating back to 1993 is a *specification of a conceptualization* (Gruber 1993). A lengthier definition is that an ontology is a logical theory, which gives an explicit, partial account of a conceptualization; it is an intentional semantic structure that encodes the implicit rules constraining the structure of a piece of reality (Guarino & Giaretta, 1995). An ontology models the vocabulary that is common to knowledge-holders in a particular domain. It explicitly describes the different concepts and the relationships that exist between concepts thus giving structure for the knowledge. Every knowledge model has an ontological commitment; this is to say, that every knowledge model has a partial account of the intended conceptualization of a logical theory (Noy and Hafner, 1997). Building an ontology can also clarify human thinking by forcing the disclosure of tacit human knowledge thus turning it into explicit knowledge that other humans and intelligent agents can use (Lopez et al., 2004).

Table 1. Ontology Classifications

Problem Solving Classification

Independent

- ❑ Generic ontologies
- ❑ Representational ontologies
- ❑ Domain ontologies
- ❑ Application ontologies

Dependent

- ❑ Task ontologies
- ❑ Method ontologies

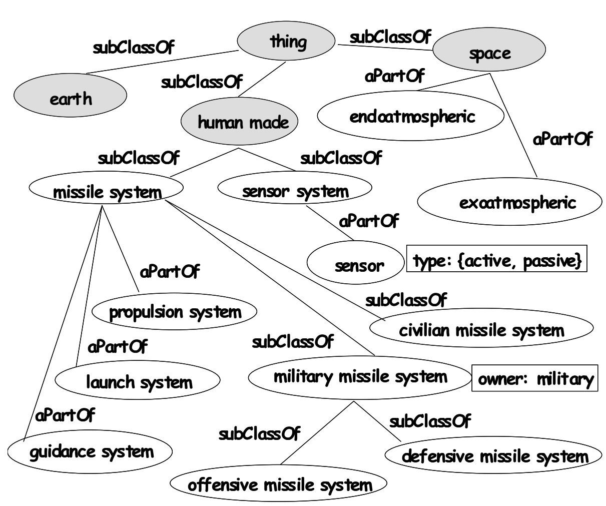
Relational Classification

- ❑ Top-level ontologies
- ❑ Domain ontologies
- ❑ Application ontologies
- ❑ Task ontologies

There are several types of ontologies. Unfortunately, the terminology used is not always consistent; this is probably because ontology development is a young field of endeavor much like knowledge management. Table 1 offers two established sets of ontology classifications. The first (Studer et al., 1998) categorizes ontologies by whether they are problem solving independent or dependent, while the second (Maedche, 2002) categorizes ontologies by how they relate to each other. Both classifications show levels of understanding or "views." Generic ontologies provide super theories valid across several domains (e.g., gravity, conservation of energy). This is somewhat similar to the Top-level ontologies that describe very general concepts (e.g., time, nature). In both classifications the Domain ontologies, the Application ontologies, and the Task ontologies are subtly hierarchical. Domain ontologies structure information for a particular domain (e.g., military, medical). Application ontologies contain the knowledge model for a portion of a Domain ontology (e.g., respectively: ballistic missile defense, colon cancer research). Task ontologies provide terms specific for problem solving in a particular Application ontology (e.g., respectively: X-band radar, colonoscopy). Method ontologies provide the terms specific to a problem solving approach for a given task within a Task ontology. For example, if the Task ontology is about determining the real warhead versus decoy warheads during the Mid-course Phase then Bayesian network terms and procedures might be part of the Method ontology. Representational ontologies are more "housekeeping" ontologies providing representation for entities without stating what should be represented, for example frame-based systems have named concepts, filled or unfilled slots, and filled slots have actual or default values for instances. This paper demonstrates the structuring of information using ontologies described in the Relational Classification scheme.

Regardless of the type of ontology being developed, graphical techniques are usually employed to help clarify the structure of the information that is desired. The graphical techniques used in developing ontologies are similar to those used to describe semantic networks — concepts are nodes

Figure 1. An Ontological Fragment



and relationships are labeled lines between nodes. However, in semantic networks attributes are treated in the same way as relationships; this is not the case with the graphical techniques used to describe ontologies. Figure 1 shows an ontological fragment that offers some of the various levels of concepts encountered in an understanding of the ballistic missile defense domain. The root for any ontology is typically called "thing" or "object". In Figure 1, the concepts of *earth*, *space*, and *human made* are Top-level ontology elements. Concepts such as *missile system*, *sensor system*, or *endoatmospheric* can be developed into Domain ontologies directly or left as "stubs" for later importation from a published ontology. The Domain ontology of *missile system* could contain those common concepts and relationships that go into building any kind of missile. Similarly, the *sensor system* Domain ontology could contain the typical concepts and relationships that would be used when constructing sensors. The *endoatmospheric* Domain ontology could contain the concepts and relationships of atmospheric realities between the surface of the earth and 60 miles up towards outer space.

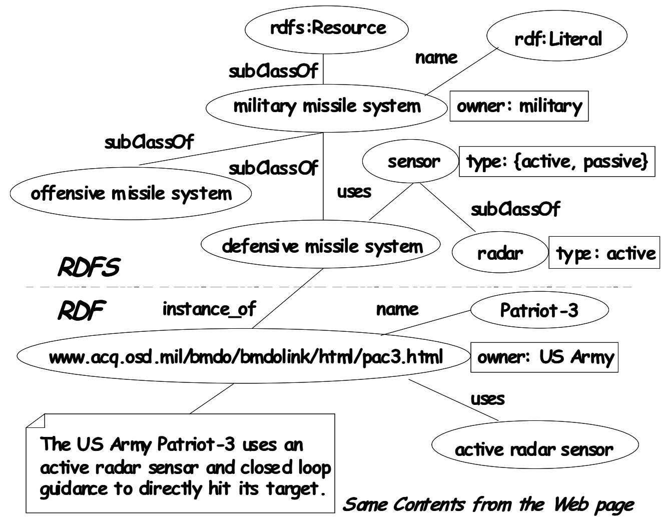
Figure 1 shows the concepts *military missile system* and *civilian missile system*, and each of these might be further developed into distinct Application ontology. The reasoning supported by each would be quite different, yet each might make the same use of other Domain ontologies. For example, a critical time for any missile system takes place in the endoatmosphere during the Boost Phase. Hostile military ballistic missiles are most vulnerable to attack and destruction during the Boost Phase. On the other hand for a civilian missile carrying a weather satellite with various sensor systems, an endoatmospheric event such as a lightening strike will more than likely cause the sensor systems to be checked for damage before the satellite is put into geosynchronous orbit. As Domain ontologies can have numerous Application ontologies, so too Application ontologies can have various Task ontologies. Within an Application ontology, there is also higher-level ontology re-use by the various Task ontologies. For example, in the *military missile system* Application ontology one of the Task ontologies might be tracking a missile during its Boost Phase, so this Task ontology might use information structured in the *sensor system* ontology as well as the *endoatmospheric* ontology. Finally, ontologies support inheritance of attributes and their values. Figure 1 shows the concept of a *military missile system* with attribute *owner* whose default value is *military*. As instances of *military missile systems* (Spencer, 2000) populate the ontology the *owner* value will be overridden with specific information.

SHARED PRACTICE

Today, the World Wide Web (WWW) is the main technological infrastructure for online information exchange between people. Communities of practice are making great use of this infrastructure, but it can be improved significantly. Berners-Lee et al. (2001) wrote, "The Semantic Web will bring structure to the meaningful content of Web pages, creating an environment where software agents roaming from page to page can readily carry out sophisticated tasks for their users." So the fundamental idea behind the Semantic Web is to embed Web pages with structured information that intelligent agents can use to perform sophisticated tasks for their human user. The envisioned end-state for the Semantics Web is words, images, and audio wrapped in organizing concepts and relationships (i.e. ontologies). Ontologies are a key technology for the Semantic Web, enabling a shared and common understanding of a domain, an application, or a task that can be communicated between people and intelligent agents (Davis et al., 2003).

The World Wide Web Consortium (W3C) has taken the lead in specifying how information on a Web page can be given well-defined semantics. The challenge is to provide a means of expressing both information and rules for reasoning about the information on the Web page. Toward this end, two important technologies have been developed – XML (extensible markup language) and RDF (Resource Description Framework). XML allows creators of Web pages to produce and use their own markup tags. If other users of the Web know the meaning of the XML tags, then they too can write scripts that make use of those tags.

Figure 2. An Ontological Fragment Tied to a Web Page



To give meaning to the XML tags, a community of practice can create a Domain ontology in which the XML tags are concepts and the relationships between concepts reveal the common and shared meaning of the Web page. Unfortunately, XML does not provide standard data structures and terminologies to describe problem-solving methods, thus the need for RDF (Gomez-Perez and Corcho, 2002). The RDF data model consists of resources, properties, and statements written using XML tags. Universal Resource Identifiers (URIs) can be used in identifying resources, properties, and statements. The RDF Schema (RDFS) provides a means of defining relationships between resources and properties. RDFS provides the basics for defining knowledge models.

Figure 2 shows how an ontology can provide the conceptual underpinning required for the proper structuring of information and its integration into the knowledge work of a community of practice. The RDFS is equivalent to a portion of the ontological fragment in Figure 1. In general then, existing ontologies can be translated into Semantic Web markup languages (McGuiness et al., 2002).

SUMMARY

Knowledge and the entities that hold the knowledge are at the very heart of knowledge management. Knowledge-holders today can be human beings or intelligent agents, each constrained to certain levels of knowledge. In the context area of ballistic missile defense, this paper demonstrates a methodology for the engineering of knowledge that all knowledge-holders can use, given their level of knowledge. The methodology involves (1) the structuring of information using various ontologies, and (2) the sharing of knowledge between communities of practice using the Semantic Web. Ontologies have been established for knowledge sharing and are used as a means for conceptually structuring information in communities of practice. The various types of ontologies presented in this paper support views of the world ranging from an individual user focused on a particular task to a community of practice providing opinions and suggestions to the solution of a very complex, multi-domain problem. The paper demonstrates how an ontology can be used to structure knowledge for the Semantic Web.

ACKNOWLEDGMENT

Production of this material was funded by the Missile Defense Agency under Cooperative Agreement HQ0006-02-2-0002. The U. S. Government is granted a paid-up, nonexclusive, irrevocable, worldwide license to reproduce, prepare derivative works, distribute copies to the

public, and perform and display publicly by or on behalf of the U. S. Government.

REFERENCES

Berners-Lee, T., Hendler, J., and Lassila, O. (2001) The Semantic Web. *Scientific American*, 284, 5, 34-43.

Davis, J., Fensel, D., van Harmelen, F. (Editors) (2003) *Toward the Semantic Web: Ontology-driven Knowledge Management*, West Sussex, England: John Wiley and Sons Ltd.

Folger, T. (2001) *Shield of Dreams*. Discover, 23, 11, 58-67.

Ford, P. (2004) Uncle Sam's Semantic Web. http://www.xml.com/pub/a/2004/09/15/egov.html.

Gomez-Perez, A. and Corcho, O. (2002) Ontology Languages for the Semantic Web. *IEEE Intelligent Systems*, 17, 1, 54-60.

Graham, B. (2001) *Hit to Kill*. Public Affairs: New York, NY.

Gruber, T., (1993) A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition*, 5, 2, 199-220.

Guarino, N., and Giaretta, P. (1995) Ontologies and Knowledge Bases: Toward a Terminological Clarification. In *Towards Very Large Knowledge Bases: Knowledge Building and Knowledge Sharing*, N. Mars (Editor), 25-32, ISO Press: Amsterdam, The Netherlands.

Gupta, S. and Sharma, S. (2004) *Creating Knowledge Based Organizations*. Idea Group Publishing: Hershey, PA.

Kenyon, H. (2003) Computer Languages Seeks Deeper Meaning. *Signal*, 57, 10, 59-61.

Lopez, A. (2004) Knowledge Engineering Enabling Knowledge Management. *Proceedings of the 2004 Information Resources Management Association International Conference*, 17-19, New Orleans, LA.

Lopez, A., Comello, J., and Cleckner, W. (2004) Machines, the Military, and Strategic Thought. *Military Review*, 84, 5, 71-77.

Maedche, A. (2002) *Ontology Learning for the Semantic Web.* Kluwer Academic Publishers: Norwell, MA.

McGuinness, D., Fikes, R., Hendler, J. and Stein, L. (2002) DAML + OIL: An Ontology Language for the Semantic Web. *IEEE Intelligent Systems*, 17, 5, 72-80.

Meilinger, P. (2003) U. S. Technology Can Outpace Decision-making. *USA Today*, March 24, 19A.

Noy, N., and Hafner, C. (1997) The State of the Art in Ontology Design: A Survey and Comparative Review. *AI Magazine,* 18, 3, 53-74.

Plekhanova, V. (2003) *Intelligent Agent Software Engineering*. Idea Group Publishing: Hershey, PA.

Schreiber, G., Akkermans, H., Anjewierden, A., deHoog, R., Shadbolt, N., de Velde, W., and Wielinga, B. (2002) *Knowledge Engineering and Management: The CommonKADS Methodology*. MIT Press: Cambridge, MA.

Smith, R., and Farquhar, A. (2000) The Road Ahead for Knowledge Management: An AI Perspective. *AI Magazine*, 21, 4, 17-40.

Spencer, J. (2000) *The Ballistic Missile Threat Handbook*. The Heritage Foundation: Washington, DC.

Studer, R., Benjamins, V., and Fensel, D. (1998) Knowledge engineering: principles and methods. *Data and Knowledge Engineering*, 25, 1/2, 161-197.

Weiner, T. (2004) Military Aiming to Change the Face of War. *The Times-Picayune*, November 21, A-30.

Wright, D. and Postol, T. (2000) Missile Defense System Won't Work. *Boston Globe*, May 11, also at http://www.commondreams.org/views/051100-101.htm.