A Semantic Web Approach for Predictive Searching in the Intelligence Imagery Domain

MARY K. PULVERMACHER AND MICHAEL S. COKUS

The MITRE Corporation
1155 Academy Park Loop, Colorado Springs, CO 80910
Phone: (719) 572-8241 Fax: (719) 572-8345
pulver@mitre.org msc@mitre.org

Abstract. This paper describes an approach for using semantic web technologies to help predict where to find information that exists in a set of imagery repositories accessible on a secure web. The approach uses information relevant to the repository to predict the likely location for a given image. This inference is accomplished using knowledge of the mission and responsibilities of the organizations that own the repositories. Although our application is in the military intelligence imagery domain, we believe this approach can be generalized to other domains.

Keywords: *Semantic Web Use Case, Ontologies, Semantic Web Inference Schemes, Imagery, Imagery Product Library, IPL, NIMA*

1 Introduction

This paper describes an approach for using semantic web technologies to help predict where to find information that exists in a set of imagery repositories accessible on a secure web. The approach uses information about the repositories, and the missions and responsibilities of the organizations that own them, to predict the likely location for a given image. Thus, information external to a repository is used to infer its expected contents. Although we chose the intelligence imagery domain, we believe this approach can be generalized to other domains.

General Military Intelligence (GMI) analysts need access to exploited imagery on simple interfaces like a web browser. An exploited (i.e., not "raw") image is one that has been analyzed, annotated, and cataloged. Figure 1 shows the difference between an image in its unexploited and exploited state. A set of repositories exist, called Image Product Libraries (IPLs), that contain exploited imagery and are accessible to analysts via a secure web. However, the search process to find a desired image among the over 400 IPLs that exist on the secure web is manually intensive. Further, the heuristics regarding which of the IPLs is likely to contain the desired image exist only in the minds of each analyst.

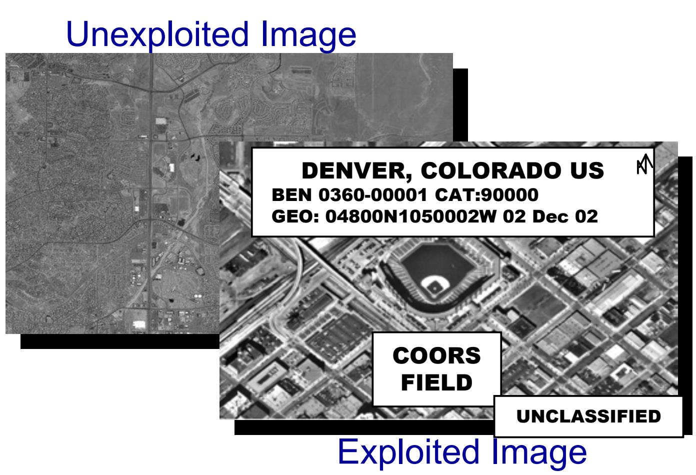

Figure 1. Sample of Unexploited versus Exploited Image

The goal of this research effort is to investigate semantic web and software agent technologies to understand how they may be applied to military domains in the future. Although our goal is not necessarily to solve the IPL searching problem described above, we do see the possibility of contributing a potential solution as an added benefit.

2 Problem

GMI analysts often require rapid access to exploited imagery via a simple web interface. IPLs store finished, annotated image products that may be dynamically converted to JPEG format. The good news is that these IPLs are accessible to analysts via a secure web using a standard interface and images are cataloged using a standard set of metadata. So, analysts can submit the same query, in the same fashion, to any accessible IPL.

One challenge is that the IPL architecture is very complex. The IPL architecture is organized along United States military organizational boundaries. Each Combatant Commander, Military Service and Agency has a set of IPLs, which they may configure in any way that makes sense for their organization. Therefore, the IPL configurations are not consistent, are often dynamic, and knowledge of the specific configurations is not readily available.

Another challenge is that there is no standard way to find the location of an IPL. Often an organization will provide a link from their home page to their IPL. However, the location of this link on a web page is not standardized. It may be on the home page or one may have to navigate to other pages to find the IPL link.

A further challenge is that there is no easy way to determine which IPLs are likely to have the desired image. There is no master index of the contents of all the IPLs. Therefore, today analysts use their own heuristics to identify which IPLs to search.

These challenges make searching for a needed image in over 400 IPLs a daunting task. The current process is very manually intensive. Today analysts log on to their secure networks and use their own heuristics to identify the IPL most likely to contain their desired image. The analyst then finds this IPL and queries it. If the required product is not found, they must guess at the next most likely source IPL and continue this process until a satisfactory image is located or they get discouraged and quit. After a fruitless hour or two, many analysts give up and submit a request to obtain a new image.

3 Approach

3.1 Overview

We simplify this manually intensive process by developing an IPL Search Agent to perform “smart searches” on behalf of the user. The analyst would specify imagery needs in the same manner as before. However, the analyst would no longer need to determine which IPL to search, locate the web address (i.e., URL) for a desired IPL, launch the user interface software for that IPL, submit a query for the desired image, and then repeat this process for additional IPLs if the desired image is not located. The IPL Search Agent would do this for the analyst.

Our IPL Search Agent exploits knowledge contained in ontologies and rules to determine which IPL to search given a set of desired imagery characteristics. Rules are used to capture analyst heuristics on which IPLs are likely to contain their desired image. Details on our approach, including descriptions of the contents of the ontologies and rules follow.

Use of an IPL Search Agent should decrease the amount of time it takes an analyst to perform their search, increase the likelihood of finding the desired image, and decrease the amount of knowledge needed by the analyst. Saved time translates to more effective operations and saves money. This approach also avoids the cost (time and resources) of creating (capturing, analyzing, annotating, and cataloging) a new image when a suitable image already exists but may be difficult to locate. Another benefit of this approach is that it accommodates the dynamic nature of the IPL architecture. Any changes to the IPL architecture would mean changes to the ontology instances (or perhaps to the rules) but not to the agent software.

3.2 Ontology

As mentioned previously, our ontologies do not model the contents of the IPLs. Rather, we capture information about the organizations that own IPLs and the about IPLs themselves in ontologies. The organization ontology captures information about the organization including its mission, location in the organizational hierarchy, geographic responsibilities, and functional responsibilities. This information allows us to infer the expected contents of an IPL based upon the organizational mission and responsibilities. (See section 3.3 for an explanation of this inferencing approach.) The organization ontology also points to the instances of an IPL that are owned by that organization. The IPL ontology provides details on an IPL including its host network, location (i.e., URL), and owning organization.

We modeled our ontologies using Protégé-2000[Protégé], an open source ontology and knowledge base editor. We also plan to experiment with the use of other ontology vocabularies, specifically DAML+OIL[DAML] and OWL[OWL]. Figure 2 is a Protégé screen print of our ontologies created with the Protégé OntoViz plug-in.

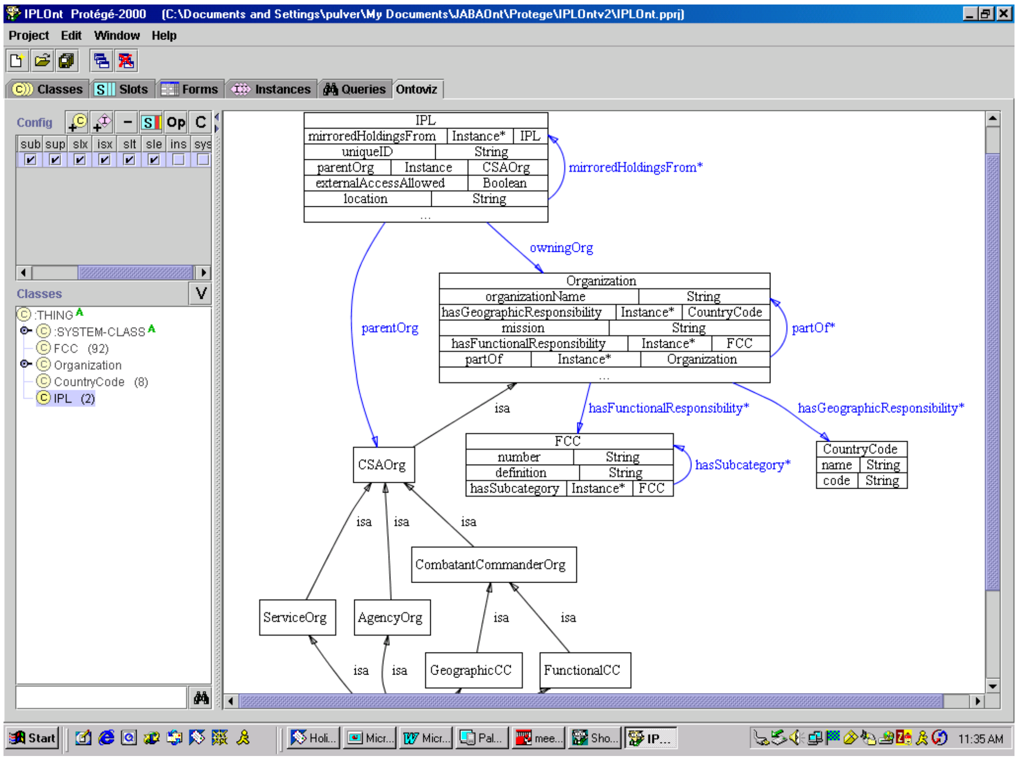

Figure 2. Protégé-2000 Screen Print of IPL Related Ontologies

3.3 Inferencing

Our approach uses inferencing to predict which IPLs could have a desired image by using information contained in the ontologies and a set of rules. These rules capture analyst "rules of thumb" on which IPLs to search and in what order. We created our rules using knowledge captured through interviews with actual IPL users. From our sample of IPL users, we found it relatively easy to discern a pattern for analyst heuristics.

Consider two of the rules we use to illustrate our inferencing approach. Each Combatant Commander, Service and Agency owns one or more IPLs. Each of these organizations has a set of responsibilities. Geographic Combatant Commanders are responsible for a region of the world that includes a set of countries. Functional Combatant Commanders are responsible for worldwide issues. For example, United States Transportation Command (USTRANSCOM) is responsible for transportation issues that affect all other Combatant Commanders. Similarly, Services and Agencies have specific responsibilities and missions. The organization ontology captures these geographic and functional responsibilities and points to IPLs owned by each organization. Now, let's assume an intelligence analyst wants an exploited image of a particular facility. One analyst heuristic is to determine within what country this facility is located and search the IPL that belongs to the Geographic Combatant Commander whose geographic responsibility includes this country. If the analyst does not find the desired image in that IPL, another heuristic is to then go to the National Imagery and Mapping Agency (NIMA) IPL since NIMA is the producer for national issues, has a worldwide mission, and supports all Combatant Commanders. These two sample rules illustrate how we use information about IPL owning organizations to predict the likely location for a desired image.

Some of the knowledge of how IPLs are used is implicit in the rules. For example, the rules assume that a particular IPL will contain exploited imagery relevant to its owning organization's mission. Another implicit assumption in our current rules is, when searching IPLs in other organizations, to search only the IPLs at the top organizational level. This is because in our construct, the top organizational level includes all the major intelligence imagery producers and for the most part, imagery produced below this level is funneled up to the headquarters level.

We implemented our rules using Jess™ [Jess], the Java Expert System Shell, a rule engine and scripting environment that is based on CLIPS. Because we modeled our ontology using Protégé, we used the Protégé JessTab[JessTab] plug-in to create a Jess knowledge base. The JessTab provides functions that made it easy to generate a knowledge base by automatically generating Jess templates for each class in our Protégé ontologies and asserting our ontology instances as Jess facts. The Jess reasoning engine operates on the Jess knowledge base to match our Jess rule patterns against the facts in the knowledge base. Execution of our Jess rules results in the creation of a prioritized list of IPLs that are predicted to contain imagery that meets the constraints provided by the user. The IPL Search Agent uses this list to submit queries to these IPLs to determine if they contain imagery that meets the given constraints.

3.4 IPL Search Agent

An overview of the IPL Search Agent is shown in Figure 3. A GMI analyst interacts with the User Agent to specify the constraints on their desired image. The User Agent converts these constraints into an Extensible Markup Language (XML) file of imagery needs and passes them to the Build Search Agent. The Build Search Agent uses these imagery needs to create the standard query and to interact with the Jess knowledge base. The Build Search Agent starts the reasoning engine and fires the Jess rules. The Jess rules match on facts created from instances in the IPL ontologies and create a prioritized list of IPL instances. The Build Search Agent then converts these instances into an XML formatted list and sends both the query and the IPL list to the Manage Search Agent. The Manage Search Agent handles submitting queries to the IPLs and interacting with the User Agent to obtain user feedback on the conduct of the search.

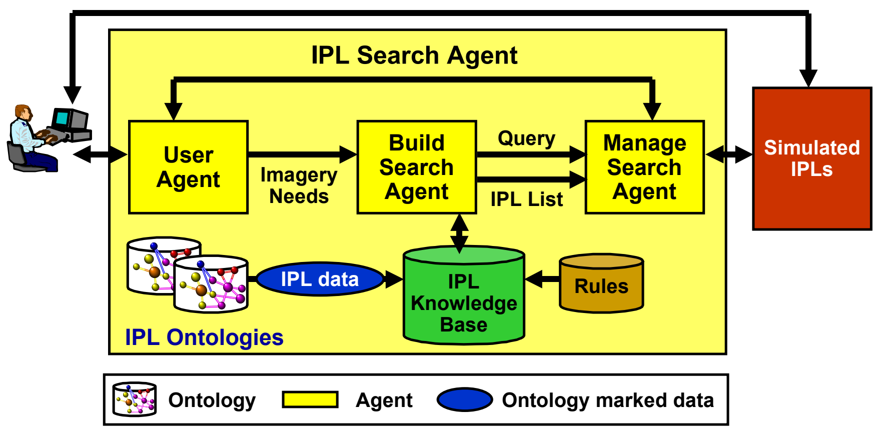

Figure 3. IPL Search Agent

We are using a minimal set of software tools to implement the IPL Search Agent. Our initial implementation uses Java to create the agent software although we plan to experiment with the use of a software agent framework (see section 3.5). A user need only have a web browser to use the IPL Search Agent. As previously described, we used Protégé to model the ontologies and to capture the instance data and used Jess as our reasoning engine. For the simulated IPLs we are using the actual software used to host and interface to IPLs.

3.5 Agent Framework

At the time of writing, we have begun development of the anticipated software agents. Initially, we are prototyping the agent software using conventional programming techniques and inter-process communication (e.g., Java and sockets). The next phase of development will involve transitioning the initial prototype software to a FIPA[FIPA] compliant agent platform. Our plan is to use the Java Agent DEvelopment (JADE) Framework[JADE] to implement the agent version of our predictive search prototype. Our initial experiences with JADE indicate that it is a stable and easy environment with which to work. It appears to be an excellent tool for exploring the synergy between software agent and semantic web technologies.

Our goal is to distribute the agents within the JADE environment, and to have each agent operate as independently as possible. This should facilitate the ability of new agents to use one or more of these agents in pursuing their objectives. For instance, at some time in the future, the User Agent could be replaced by another agent in the network that acts as an interface to some application that does not interact directly with the user. Experimenting with both types of user agents (human and machine interfaces) co-existing in the same environment could be subject of follow-on work. One could envision another example where a user request causes multiple agents to be invoked. In this case an agent could notify a user that not only was desired imagery found, but other related material (document, web page, message traffic, etc.) was also discovered.

4 Potential Role in Air Force Enterprise Vision

Another aspect of our research involves the integration of semantic web technologies into a United States Air Force (USAF) information vision. The current USAF Information Vision is "Enabled mission capabilities through seamlessly integrated access to the right information anytime, anywhere."[USAF] An antecedent for this information vision, known as the Joint Battlespace Infosphere (JBI), is described in the December 1998 USAF Scientific Advisory Board (SAB) report [SAB]. The JBI concept was proposed in response to the increasing volume and importance of digital information in military activities. The fundamental requirement proposed in the JBI vision is for an "information-centric" enterprise "To provide the right information at the right time, in the right way."

This information-centric approach emphasizes information objects, and brokering and transformation services to support a paradigm of loosely coupled interactions. The services provide capabilities to bring together the information necessary to support information consumers and their missions. As an analogy, the JBI functions as an "information supermarket" or "infomart" as depicted in Figure 4 [Miller]. Although distributed, the infomart provides a virtual central location for one-stop information shopping for information consumers and a one-stop information outlet for information producers. The "goods" and services made available for exchange are all information-based.

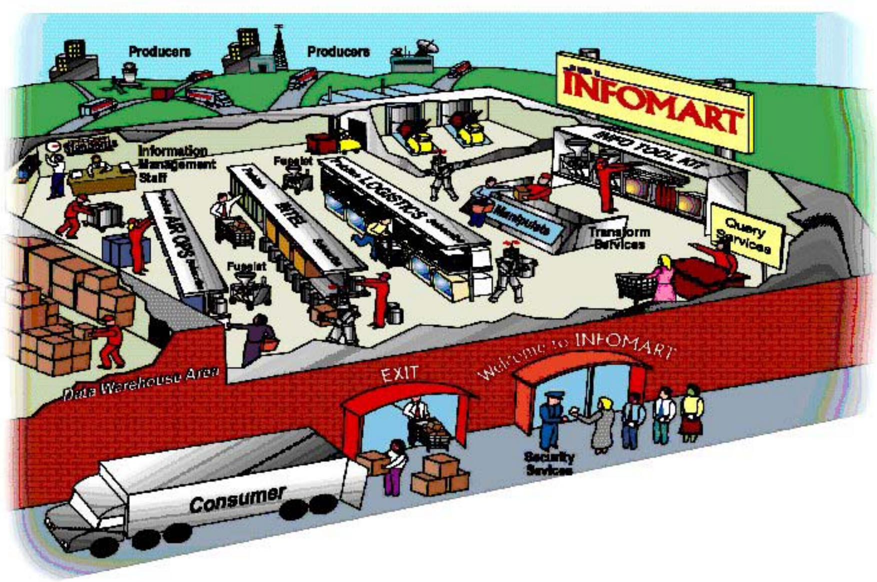

Figure 4. The Infomart

A common method for information producers and consumers to share information is fundamental to the infomart concept. Information producers place their information products into the infomart by "publishing" while information consumers receive information products available through the infomart by "subscribing". In addition, information consumers typically require the capability to find and extract only the information they need. Consumers search for and retrieve precisely the needed information through the infomart's query services to reformat and translate retrieved information in a user-specified fashion.

MITRE has prototyped JBI concepts using web services technologies and is currently investigating the selection and integration of policy-based management services[Cherinka]. We plan to explore the role that semantic web technologies might play in the evolving JBI environment. In addition, we will consider the original vision of the JBI, as expressed in the SAB report[SAB], to consider how a semantic web approach might be applied on the basis of first principles, rather than as an extension to the current realization of the JBI.

5 Conclusions

Our experience to date with semantic web technologies (i.e., use of ontologies and a reasoning engine) lead us to believe they are a promising approach for predictive searching where information about the content of a repository can be inferred from external knowledge. We believe our research effort has demonstrated that an inferencing approach maps well to our selected problem. We found implementation with the combination of Protégé, Jess and Java worked well. It was fairly easy to model the information sources, along with relevant information on their owning organizations, with an ontology. After an initial learning curve, we also found it easy to capture our rules with Jess. It is easy for us to see logical extensions in the use of ontologies, rules, and software agents. However, it is too early in our project for us to extrapolate to meet a goal of this effort – to create informed opinions on when and where these technologies could be applied to our customers' domains (i.e., the United States military).

6 References

[Cherinka] Cherinka, R., Wild, C., Allen, D., Smith, C., Zhang, Y., Panek, R., Semy, S., "Policy-based Information Management for the Joint BattleSpace Infospere", IEEE 4th International Workshop on Policies for Distributed Systems and Networks (accepted), Lake Como, Italy, June 5, 2003

[DAML] DARPA Agent Markup Language + Ontology Inference Layer (DAML+OIL), http://www.daml.org/language/

[FIPA] The Foundation for Intelligent Physical Agents (FIPA), http://www.fipa.org/

[JADE] Java Agent DEvelopment Framework, http://sharon.cselt.it/projects/jade/

[Jess] The Java™ Expert System Shell (Jess), Sandia National Laboratories in Livermore, CA, http://herzberg.ca.sandia.gov/jess/

[JessTab] Protégé-2000 Plug-ins Library, http://protege.stanford.edu/plugins.html

[Miller] Miller, R., M. A., Malloy, E. Masek, and C. Wild, "Towards an Information Management Framework", Information Knowledge Systems Management Report, 2001

[OWL] Web Ontology Language (OWL), World Wide Web Consortium (W3C) Web-Ontology Working Group, http://www.w3c.org/2001/sw/WebOnt/

[Protégé] The Protégé Project, Stanford University School of Medicine, http://protege.stanford.edu/index.html

[SAB] United States Air Force Scientific Advisory Board, Information Management for the Warrior, Technical Report, 1998

[USAF] Air Force Information Strategy, August 2002.