



Semantic Technologies for Knowledge Synthesis

The National Infrastructure Simulation and Analysis Center (NISAC), a program under the Department of Homeland Security's (DHS) Preparedness Directorate, provides advanced modeling and simulation capabilities for the analysis of critical infrastructures, their interdependencies, vulnerabilities, and complexities. These capabilities help improve the robustness of our nation's critical infrastructures by aiding decision makers in the areas of preparedness, consequence and risk analysis, policy analysis, investment and mitigation planning, education and training, and near real-time assistance to crisis response organizations.

Sandia National Laboratories (SNL) and Los Alamos National Laboratory (LANL) are the prime contractors for NISAC, integrating the two laboratories' expertise in infrastructure disruption/vulnerability modeling and simulation under the direction of DHS's Infrastructure Protection/Risk Management Division.

Knowledge Synthesis

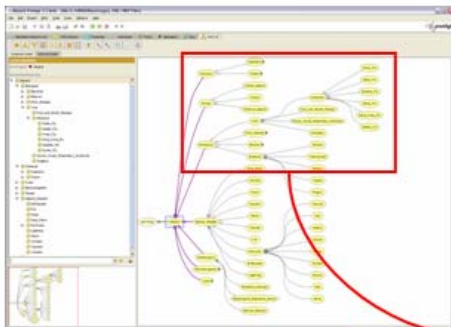
NISAC analysts and modelers require vast amounts of data and information for their critical infrastructure interdependency work. The

knowledge synthesis effort provides an access-controlled, organized, and searchable view on the data for rapid retrieval and update through a suite of Web-based tools and programmatic interfaces. Four semantic technology projects are underway to help improve information access:

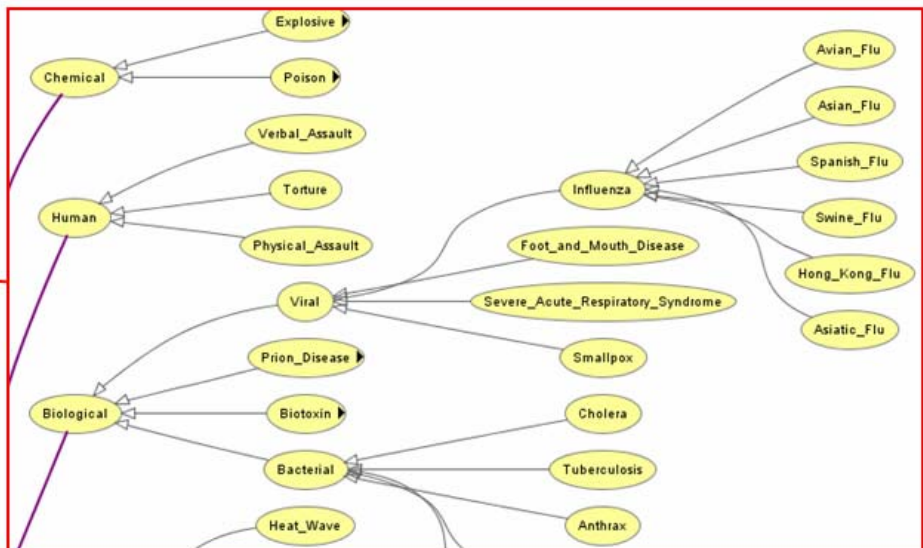
- Ontology Development
- Semantic Navigation
- Keyword Expansion (Synonyms)
- Automatic Semantic Metadata Determination

Ontology Development

An ontology is a map of the key terms in a domain of knowledge and the relationships among the terms. It forms both a vocabulary and a cognitive framework for reasoning about the domain. By structuring domain knowledge and making relationships both within and among subject areas explicit, NISAC analysts are able to categorize and access information more efficiently. Additionally, the ontology is flexible enough to be adjusted as the understanding of the domain evolves. High level structural components of the current ontology for the critical infrastructure protection domain include hazards, critical assets, geographical location, consequences, and organizations.



A prototype of the NISAC ontology of hazards displayed in Protégé



Semantic Navigation

The NISAC Critical Infrastructure Protection Knowledge Management (CIP KM) Portal is a one-stop shop where NISAC analysts, modelers, and administrators can contribute and share information. Semantic navigation allows the Knowledge Management Portal to be searched in terms of concepts and relationships, rather than by keyword only. This approach overcomes the semantic ambiguity and unrecognized synonym issues associated with keyword search. As a result, analysts can navigate the portal at the “speed of click” using categories drawn from the domain of critical infrastructure protection.

Keyword Expansion

A step along the way to full synonym resolution is the expansion of keywords used to search the CIP KM Portal. Each keyword is expanded to include synonyms related to the concept being searched, in priority order. For example, a search on the term “bird flu” brings back documents containing that phrase and also those containing “avian influenza” and “H5N1.” The synonyms are drawn from those commonly used by critical infrastructure protection analysts, as well as a more general set of synonyms taken from a public domain thesaurus.

Automatic Semantic Metadata Determination

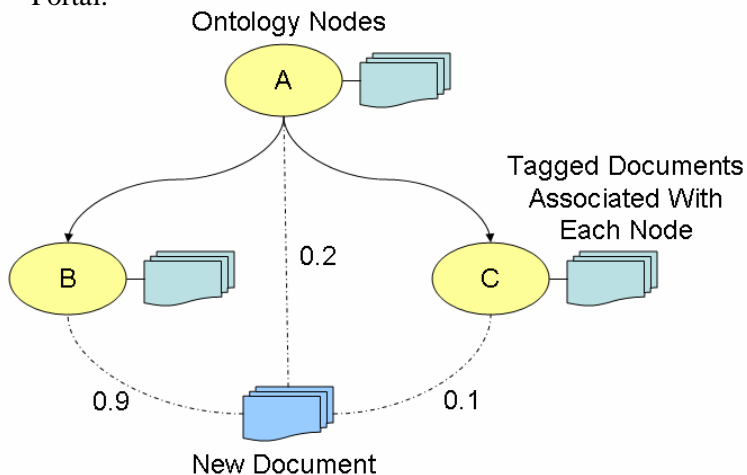
As new documents are added to the CIP KM Portal, they must be tagged under metadata categories defined within the ontology in order to be available for semantic navigation. This is a tedious and error-prone process when carried out manually. A system is being developed to automatically recommend the appropriate metadata categories using a text analysis tool. However, the person submitting the information will review the recommendations

Web-based Semantic Navigation of the CIP KM Portal

and make the final decision. The benefits of automatic semantic metadata determination include:

- Efficient addition of new documents, which minimizes human involvement in determining semantic metadata
- Potentially better quality and consistency of semantic metadata.

An extension of this application is being designed to target information on the Internet, making a tremendous amount of other data resources available to users of the CIP KM Portal.



A new document comes in and shows the highest correlation (0.9) with node B of the ontology. The system recommends that the document be semantically tagged using the semantic metadata of that node.

Contacts:

Merrick Krause
DHS-OIP
(703) 235-5410; e-mail:
merrick.krause@dhs.gov

Theresa Brown
Sandia National Laboratories
(505) 844-5247; email:
tjbrown@sandia.gov

