

# Semantic Technology Applications for Homeland Security

D. Avant, M. Baum, C. Bertram, M. Fisher, A. Sheth and Y. Warke

[Semagix Inc.](#)<sup>1</sup>, Athens, GA

## 1. INTRODUCTION

Semantic Content Organization and Retrieval Engine (SCORE) is among the earliest commercialized Semantic Web technologies. Based on supporting and exploiting domain specific ontologies, it offers advanced capability in heterogeneous content processing analysis, and integration at a higher semantic level-- rather than merely syntactical and structural level approaches based on XML and RDF. These capabilities are now being demonstrated in addressing requirements of very demanding Homeland Security and National Security applications. This paper briefly describes two of them.

## 2. HOMELAND SECURITY REQUIREMENTS & SCORE CAPABILITIES

Semantic Web technology as realized by SCORE [1,2] is used to satisfy a number of requirements found in Homeland Security applications:

- Extract, organize, and normalize valuable, meaningful information from many disparate and heterogeneous content sources, including structured, semi-structured, and unstructured sources; static and dynamic (e.g., database driven) sources, that may be internal or external to the organization (incl. deep web)
- Extract, organize, and normalize contextual knowledge (people's names, places, organizations, etc., and relationships between them) from heterogeneous sources and formats (database tables, xml feeds, PDF files, streaming media, internal documents)
- Co-relate extracted information to discover previously unknown or non-obvious connections between documents and/or entities
- Make efficient use of the extracted knowledge and content by providing tools that enable fast and high quality querying and browsing

Following core capabilities of the SCORE technology address the above requirements:

**Semantic Metadata:** Metadata that describe domain-specific information about content (in the right context) based on a custom (e.g., industry-specific or enterprise specific) metadata model, called WorldModel. For example, if the content is from

the Business domain, the relevant semantic metadata could be company name, ticker symbol, industry, sector, executives, etc., whereas if the content is from the Intelligence domain, the relevant semantic metadata could be terrorist, event, location, organization, etc. Metadata elements that offer more insight 'about the document' fall under the semantic metadata category.

**Automated Extractor Agents:** SCORE's distributed software agents created using toolkits by nonprogrammers monitor desirable information sources, and extract semantic metadata as well as contextual knowledge. Automatic extraction of semantic metadata enable automatic semantic annotation.

**Semantic Normalization:** Associating (as much as possible) the same metadata for content belonging to the same domain or category regardless of source and format, coupled with homogenization of multiple names of a single entity into one uniform (canonical) name.

**Semantic Association:** Leveraging domain-specific real-world semantic relationships between objects and entities to greatly strengthen the relevance of content to the end user. SCORE uses a Knowledgebase to store a "semantic net" of real world entities and named relationships between them.

**Semantic Query and Personalization:** Personalized information retrieval in the appropriate context (to granular levels of detail), transcending the boundaries of syntactic keyword-based matches. SCORE operates on main-memory indices that make even complex querying about two orders of magnitude faster than RDBMS. Java- and HTTP-based APIs enable rapid development of complex semantic applications that provide Blended Semantic Browsing and Querying (BSBQ) features, an intuitive mix of information querying and knowledge navigation.

## 3. PASSENGER THREAT ASSESSMENT

The objective of this semantic application for transportation environments such as airports is to provide various kinds of personnel -- most importantly, security officers such as interrogators -- with useful tools to effectively assess and react to passenger threats while maintaining a high rate of passenger flow through the system. It is also a building block of an application that can monitor the national airspace.

The information that is collected and derived from semantic data processing supports important decisions such as adding a marshal to a flight, granting push-back permissions, searching carry-on luggage, or requiring a brief interrogation of a particular passenger before letting him or her advance to the gate. Interrogation specialists will find BSBQ helpful to quickly gather a wealth of relevant information that can be used for the interviewing process (see Figure 1).<sup>1</sup>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.  
CIKM '02, November 4-9, 2002, McLean, Virginia.  
Copyright 2002 ACM 1-58113-492-4/02/0011...\$5.00.

<sup>1</sup> Formerly Voquette, Inc. and Taalee, Inc. SCORE is a patented technology of Taalee, Inc. <http://www.voquette.com>



Figure 1: Passenger Threat Analysis Console

The data sources for passenger threat analysis applications typically comprise

- Watch lists: Office of Foreign Asset Control (published as PDF), FBI Most Wanted List (Web site), Denied Persons List (database table)
- Passenger information: flight history, reservation information, and ad-hoc information entered by the check-in or security personnel
- Flight details: origin, destination, aircraft type
- Massive data repository that provides secure access to public record and background information on 285 million individuals
- Biometrics information such as fingerprints or photographic images
- Various news sources: CNN, BBC, NewsML feeds, internal memos, etc.

SCORE interfaces with all of the above sources, extracts relevant bits of information in near real-time, and then organizes and normalizes them based upon the WorldModel. Using its

APIs, it co-relates the extracted information from knowledge and content sources to determine possible threats. Adjustable threat scoring algorithms include

- **Link analysis:** The Knowledgebase can show whether a passenger is directly or indirectly (through related organizations, people, or places) linked to a watch list
- **News analysis:** Analysis of the information stored as semantic metadata can reveal whether the passenger's name appears in the same news story or internal report as known terrorists
- **Personal information analysis:** Analysis of other information such as biometrics, form of payment, type of ticket, background information as provided by certain data stores, etc.
- **Cumulative threat analysis:** Individual passenger scores can be combined to calculate cumulative flight scores that take seating arrangements into consideration; similarly, threat scores can be calculated for the entire airport

## 4. ANTI-TERRORISM

In another environment, SCORE is deployed to effectively interface with other applications to form a powerful Knowledge Discovery Toolkit that enables intelligence analysts to detect possible or planned terrorist activities (see Figure 2).

**Knowledge-Tagger** [3]: Enables your organization to offer enhanced metatagging abilities to enrich content with real value - information connected to explicitly-mentioned and associated concepts, categorized according to relevancy and customized for the unique language, structure and needs.



Figure 2: Anti-Terrorism Knowledge Discovery Analysis

While BSBQ alone is already a helpful tool for analysts, an important role of the SCORE technology is to interface with and retrieve the best value from various software packages that all expect data in different formats. One package, for instance, can take a person's name as it was extracted from a document and perform a similarity analysis using numerous databases to detect possible fraudulent activity associated with that name. Another package takes the unstructured form of a news story to discover common concepts among many documents. Yet another application specializes in visualizing data such as the links between named entities as found by SCORE, working off a fixed database schema. SCORE can also work with "notebook" and reporting tools that allow for the organized storage of information deemed interesting by the analyst, and then process them to uncover non-obvious relationships among them.

The integration of all of the above mentioned software packages, coupled with advanced SCORE capabilities such as the ability to support the identification of semantic associations, provides the analyst with a powerful toolset. Their capabilities are maximized by SCORE's extraction of a large number of heterogeneous content and knowledge sources, and the successive normalization and enhancement of the extracted metadata. SCORE technology discussed here is available both as licensed enterprise software as well as QuickStart Web Services solutions on an ASP basis. Core products include:

**Enterprise Semantic Platform:** The Enterprise Semantic Platform (ESP) transforms content from multiple sources into useful information by providing a single view of complex inter-relationships by intelligently correlating content with contextual real-world knowledge, thus making the information more relevant and actionable for enterprise users.

## 5. REFERENCES

- [1] A. Sheth, C. Bertram, D. Avant, B. Hammond, K. Kochut, and Y. Warke, "[Semantic Content Management for Enterprises and the Web](#)," IEEE Internet Computing, July/August 2002, pp. 80-87.
- [2] A. Sheth, D. Avant, and C. Bertram, "System and Method for Creating Semantic Web and Its Applications in Browsing, Searching, Profiling, Personalization and Advertisement," U.S. Patent #6,311,194, 30 Oct. 2001.
- [3] B. Hammond, A. Sheth, K. Kochut, "[Semantic Enhancement Engine: A Modular Document Enhancement Platform for Semantic Applications over Heterogeneous Content](#)," in Real World Semantic Web Applications V. Kashyap and L. Shklar, Eds., IOS Press, 2002.