

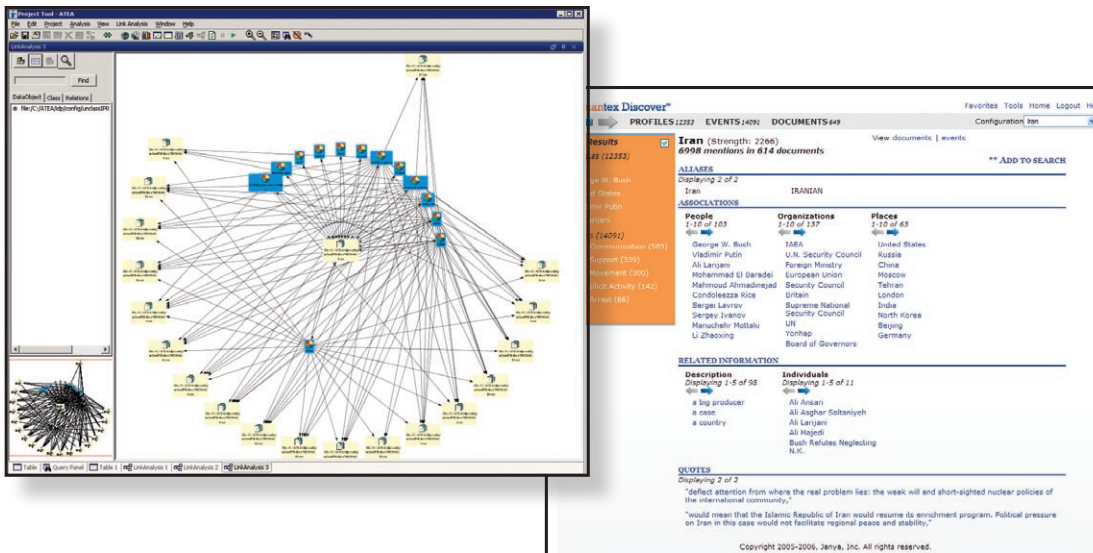
# Solution Brief

## Unstructured Text Extraction and Visualization

Today's complex intelligence analysis problems cross domain boundaries and involve a wide variety of unstructured and semi-structured text sources. To address this need, analysts need a solution that can handle this complexity of data source, fusing the information and providing the analyst with the results they need, in both text and an array of visualization options.

### The Semantex-WebTAS Solution

Janya and Intelligent Software Solutions (ISS) have brought together their leading platforms for text extraction, data access, and visualization to create a combined solution that leverages the strengths of both products, while providing a seamless integration that makes the results of the Semantex processing available to WebTAS. Users can submit source data to Janya's Semantex™ engine for extraction and Semantex will automatically process the results and populate a database with the extracted information. Janya has worked with ISS to define a standard schema for the data so that the WebTAS client can import the data quickly and provide users with a wide array of visualization options. In the end, the analyst gets the data they need the way they want to see it, including access to the original source, and can quickly focus in on the relevant information.



## About Semantex™

Semantex™ is a complete information extraction system that reads large volumes of electronic text information in order to detect entities, attributes, relationships and events. Semantex™ represents a hybrid model for information extraction, combining machine-learning and grammatical approaches, and leveraging the strengths of each to achieve a high degree of accuracy. Semantex™ can process a wide variety of unstructured and semi-structured text source, including open source news, HUMINT, USMTF and email.

Semantex™ is modular, scalable and portable to various domains of interest and is used to power both commercial and intelligence applications. It can exploit multiple processors to achieve the throughput necessary for high-volume environments typically encountered in data-intensive government applications. The XML output reflects the rich features assigned to elements of the text based on syntactic and semantic knowledge.

## TECHNICAL SPECS

Semantex™

### Product Details

- Automatic extraction of:
  - Entities – Over 50 predefined types
  - Relationships – Over 50 predefined types
  - Events – Over 40 types predefined
  - Attributes
- Date/Time Normalization
- Entity Profiles
- Scalable, Fault-Tolerant Architecture
- Extensive Customization Capabilities

### System Requirements

- Dual Processor CPU, 2.0 GHz or better
- 1 GB RAM minimum, 2 GB recommended
- Linux RedHat EL3 and EL4, Microsoft Windows XP Professional, SP2
- MySQL Database Support

## About WebTAS

WebTAS is a modular software toolset supporting access to and the fusion of large and disparate data sets. WebTAS provides visualization, project organization and management, pattern analysis and activity prediction, and various presentation aids. These tools have been successfully used by many Federal and regional government organizations and are ideally suited to solutions ranging from a single user to a full service oriented architecture, enterprise solution. WebTAS provides an integrated toolbox to users needing to make sense of large and diverse sets of data.

The unique architecture provides a virtual file cabinet that can exploit a diverse set of information sources and formats. In addition to the local data services provided with the software, WebTAS can access external databases and real-time feeds and provide links to multimedia such as imagery, video, and audio. This allows integrated access to all of this information from within a single application.

WebTAS

### Product Details

- Flexible data access methodology to many disparate databases, streaming and proprietary data sources
- Multiple sources of data into common, fused pictures
- Sophisticated, multi-modal visualization components for viewing data
- Ad-Hoc query of disparate data
- Collaborative analysis with integrated textual and multi-modal collaboration
- Trend and pattern detection
- Dynamic web content generation based on these data sources
- Easily configured domain customization

### System Requirements

- 1.0 GHz or better CPU
- 1 GB RAM minimum, 2GB recommended
- Microsoft Windows 2000 or newer
- Multiple relational databases supported

## CONTACT:

### Janya

info@janyainc.com  
www.janyainc.com  
202-742-6590

### ISS

info@issinc.com  
www.issinc.com  
719.457.0690