

## SOLUTION BRIEF

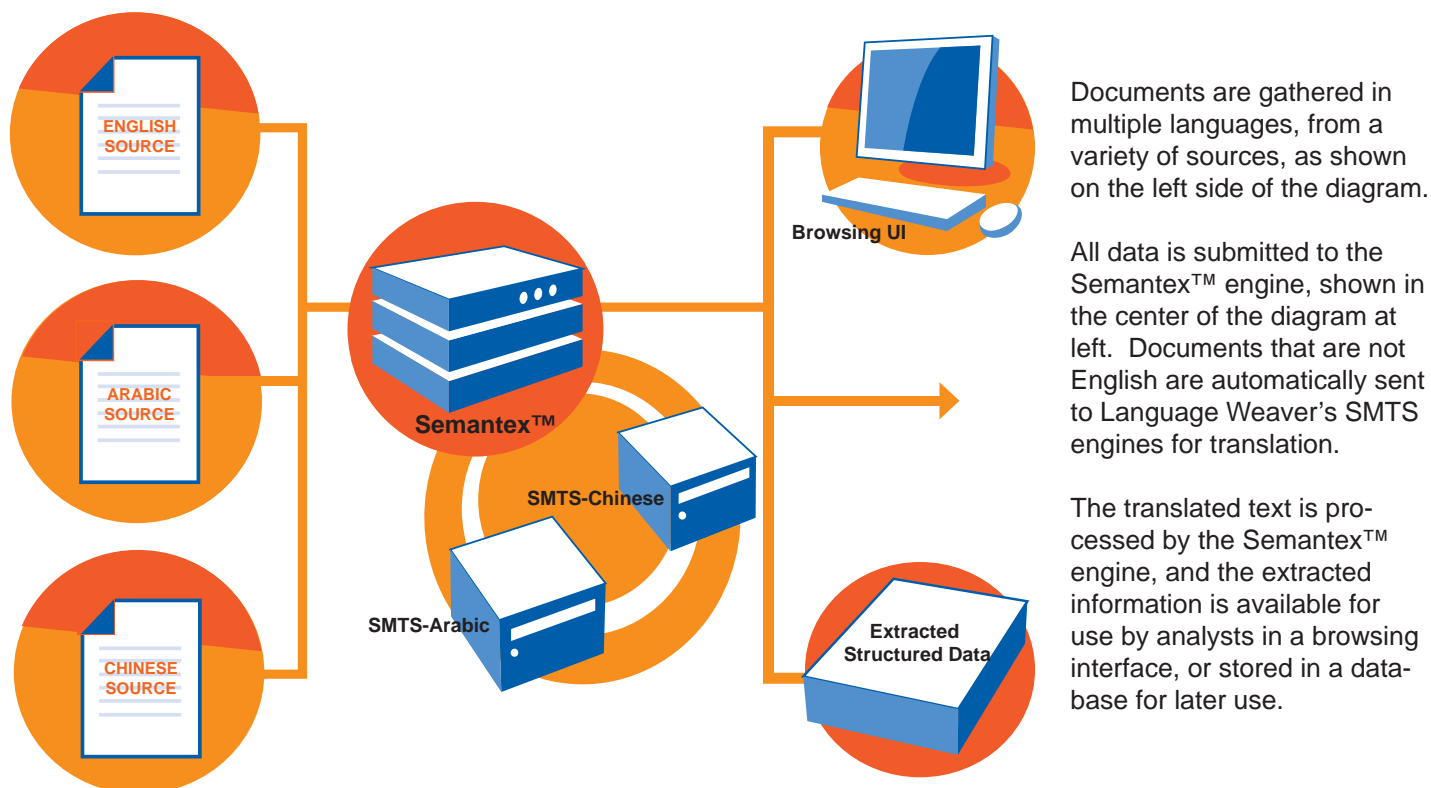
### MULTILINGUAL TEXT EXTRACTION SOLUTION

Today's complex intelligence analysis problems cross boundaries of language and subject domain, and involve a growing variety of unstructured and semi-structured text sources. The amount of data requiring analysis continues to grow, and analysts - especially those that are multilingual - are in greater demand. To address this complex need, analysts need a text extraction solution that can handle multiple data source types and languages, but also intelligently consolidates information in a format that is easy to digest and provides actionable results.

#### THE SEMANTEX-SMTS SOLUTION

Janya and Language Weaver have brought together their leading platforms for text extraction and language translation to create a combined solution that leverages the strengths of both products, while providing a seamless integration that eliminates the need for running the applications separately and manually managing the results. Users can submit source data to Janya's Semantex™ engine for extraction, and if that source is in a non-English language, Semantex will automatically submit the source to the Language Weaver's statistical machine translation software (SMTS) server for translation, and process the returned results. In the end, the analyst gets the data they need, including access to the original source in both English and the source language, if applicable.

With this integrated solution, analysts have access to more relevant information that crosses multiple languages and sources so they can analyze information more effectively and make better-informed recommendations and decisions.



**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™ 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.

**ABOUT SMTS**

SMTS is statistically based, state-of-the-art language translation software. Translations are more accurate and fluent than other machine translation solutions, and translation quality continues to improve as more data is added to the system, as algorithms improve, and as processing power increases.

Language Weaver's translation software is scalable and has a well-documented, straight-forward API (Web Services or C++) which makes integration fast and easy.

The screenshot shows the Semantex Discover interface. At the top, there are navigation tabs for 'PROFILES', 'EVENTS', and 'DOCUMENTS'. Below this, a table lists extracted entities categorized into 'People', 'Organizations', 'Places', and 'Other'. The 'People' list includes names like George W. Bush, Tony Blair, and Gordon Brown. The 'Organizations' list includes Democrats, Al Qaeda, and Labor Party. The 'Places' list includes United States, Afghanistan, and Iraq. The 'Other' list includes Ethiopia and Fort Campbell Kentucky. Below the table, there is a text snippet from a document, and a 'Show Arabic source' button is highlighted with a red circle. To the right, a window displays the 'Show English source' of the same text, showing the original English text with some words highlighted in red.

Articles that were once in another language are translated and available within the Semantex interface. Users can also view the original source document.

**Semantex™**

**Product Details**

- Automatic extraction of:
  - Entities – Over 50 predefined types
  - Relationships – Over 50 predefined types
  - Attributes
  - Events – Over 40 types predefined
- 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
- MySQL Database Support

**SMTS**

**Product Details**

- Easy customization to specific subject domains
- User dictionary
- Web Services or C++ API
- Side-by-side display for web page translations
- Unicode internal
- Continuous improvement in speed and quality
- Rapid development of new languages
- Easy integration with multiple applications

**System Requirements**

- OS: Windows 2003 Server, Windows 2000 Advanced Server (SP3+)
- Processor: 2.8 GHz Pentium 4
- RAM: 4 GB

**Languages Supported**

- Arabic
- Chinese
- Persian
- Russian
- Hindi
- Somali
- French
- Italian
- German
- Spanish
- Portuguese
- Swedish
- Dutch
- Polish
- Czech
- Romanian
- More languages coming soon!

**CONTACT**

For further information on Semantex, SMTS or the integrated solution, please contact Janya or Language Weaver

**Janya, Inc** | Janya Sales | 202-684-7053 | sales@janyainc.com | www.janyainc.com  
**Language Weaver, Inc** | LW Sales Team | 310-437-7300 | sales@languageweaver.com | www.languageweaver.com