

Semantex™ is a highly scalable and customizable multilingual semantic analysis platform supporting text analysis driven applications in deployments ranging from low data volumes to big data environments.

## Information Extraction

- **Named and Nominal Entities, Coreference**
  - ◆ Named entities such as people, locations, and organizations.
  - ◆ Numerical expressions such as date, year, percentage, etc.
  - ◆ Nominal entities such as *the phone*, etc. which get slotted into the proper category.
- **Relationships and Attributes**
  - ◆ Entity relationship extraction, e.g., relationship associated with *person* include employer, associates, spouse, location, etc.
  - ◆ Descriptive information such as age, job title, affiliation, etc.
  - ◆ Contact information such as address, e-mail, URL, phone number, etc.
  - ◆ Generic descriptor phrases used to describe entities, e.g., *the largest software manufacturer*, in reference to Microsoft.
- **Events**
  - ◆ Captures *who* did *what* (to *whom*) *where* and *when* automatically.
  - ◆ Support for user-defined event categorization.
- **Time and location Normalization**
  - ◆ Time normalization and support for TIMEX2 standard.
  - ◆ Location normalization and geo-tagging (latitude, longitude) support for over 10 million geographical entities.
- **Multilingual Support**
  - ◆ Native support for Unicode UTF-8 text.
  - ◆ Within-language entity, event, and sentiment tagging support for multilingual configurations in Simplified Chinese, Urdu, Arabic, Russian, Pashto.
- **Ontology and Taxonomy Support**
  - ◆ OWL, RDFS ontology, taxonomy support for ontology-driven extraction.

## Architecture

- **Hybrid model and Customization**
  - ◆ Statistical/machine learning modules and toolkit for custom categorization.
  - ◆ Grammar/rule modules.
  - ◆ Lexicons/gazetteers.
  - ◆ Semantex Workbench™, intuitive graphical development environment for user-enabled customization.
- **Optimized Configurations**
  - ◆ Support for multiple simultaneous extraction configurations, and optimized levels *express*, *basic*, and *full*.
- **Janya Document Filters**
  - ◆ Support for over 100 word processing, spreadsheet, email, and other file formats including Word, Excel, PDF, and Outlook. Comprehensive list of supported formats available on request.
- **Integration**
  - ◆ Web services SOAP interface.
- **Scalability**
  - ◆ Multi-threaded processing support.
  - ◆ Open source and commercial load balancer support, including HAProxy and Windows 2008 load balancing.
- **Output Formats**
  - ◆ Configurable inline and stand-off XML.
  - ◆ RDF and HTML output.
  - ◆ Database output with custom schema support.
  - ◆ RDBMS support via SQLAPI for MySQL, Oracle, SQL Server, PostgreSQL, etc.
- **Hardware Requirements**
  - ◆ Dual processor CPU, 2.0 GHz or better with 3GB RAM minimum per Semantex™ processing thread.
  - ◆ Windows XP, 7, Redhat EL/CentOS 5.5.



[www.janya.com](http://www.janya.com)



[sales@janya.com](mailto:sales@janya.com)

## Mine

Enterprise and  
Web Content



Social Media  
and Customer  
Conversations



News  
Subscriptions



Multilingual  
Content



Scientific and  
Technical  
Content



## Discover

Categorization



Semantic Analysis

Entities  
Relationships  
Events and activities  
Sentiment  
Facts



Native Multilingual  
Analysis



Customized Analysis

Machine learning  
Grammar rules  
Lexicon lists



## Act

Interactive  
Analytics and  
Intelligence



Multilingual  
Faceted Search



Customized  
Reports



Alerts and  
Warnings



1408 Sweet Home Road,  
Suite 1  
Amherst, NY 14228  
Phone: 716-565-0401  
Fax: 716-565-0308



1050 Connecticut Avenue NW,  
10th Floor  
Washington, DC 20036  
Phone: 202-684-7053  
Fax: 202-772-2370

# Janya