



# Janya, Inc. KDD Proposer's Day Briefing

## Who we are:

- Language technology company with offices in Amherst, NY and Washington DC
- 30 employees with expertise in machine learning, computational linguistics, several employees with clearances
- Facilities to store and process classified data

## What we do:

- Develop and market text mining software and solutions that extract critical information from unstructured and semi-structured data.

## Products and Services:

- Semantex™ - Scalable, customizable, multilingual text extraction platform

## Principal Investigators:

- Rohini Srihari, Phd. – Text Mining, Social Media Mining
- Tom Cornell, Phd. – Ontologies, Semantics
- John Chen, Phd. – Machine Learning

## Current Team Members Capabilities:

- Multimedia data, scalable repositories, behavior modeling, predictive modeling, link analysis and visualization

Janya is involved in a number of research or deployment efforts related to Knowledge Discovery and Dissemination:

## Social/Behavioral Modeling and Mining

- Processing noisy data inherent in email, chat, blog posts, etc. allowing for information discovery and modeling of future actions

## Multi-Lingual Extraction and Context Aware Translation

- Extracting entities, events, relationships within foreign language scripts and supporting machine translation tools

## Information Integration:

- Creating a holistic picture of people, organizations, and events across document collections

## Active Learning for Automated Extraction

- Supporting analyst in the loop

## Handling Ambiguous Locative Expressions

- Solving for "3 miles east of Baghdad"; Paris Texas vs. Paris France

## Corporate Capabilities:

- Dedicated team of computational linguists and experts in machine learning whose core team has worked together over 8 years
- Experience with full spectrum of software from academic research to product development and support, both in govt. and commercial sectors
- Personnel cleared to TS/SCI levels with classified storage capability on-site in Amherst, NY
- Woman-owned small business

## Semantex Platform Superlatives:

- Rich Entity and Event Profiles – relationships, attributes, and connections about and between entities and events
- Hybrid Processing – lexical, grammatical, and statistical approaches for best of breed extraction
- Domain Customization – customizable by subject matter experts
- Ease of Integration – CORBA and SOAP interfaces; natively ingests and outputs XML
- Scalability – Semantex processes can be spread across multiple CPU's or multiple serve for enterprise-level support
- Multi-lingual Support – new languages require linguistic resources not new code

## Research Areas of Interest:

- Automated transformation of data and extraction requirements between ontologies and application towards social media
- Application of extraction efforts towards multi-lingual social media data sets with emphasis on sentiment analysis
- Cross-document entity and event disambiguation
- Semi-automated and Automated efforts to include “analyst in the loop” capabilities for knowledge management systems.

## Partners/Teams of Interest:

- Systems integrators
- Universities
- National labs
- Complementary Technology Vendors.



## Contact Information

John Weaver  
Vice President, Business Development  
Janya, Inc.

Phone: 202-684-7053

Email: [jweaver@janyainc.com](mailto:jweaver@janyainc.com)

Web: [www.janyainc.com](http://www.janyainc.com)