



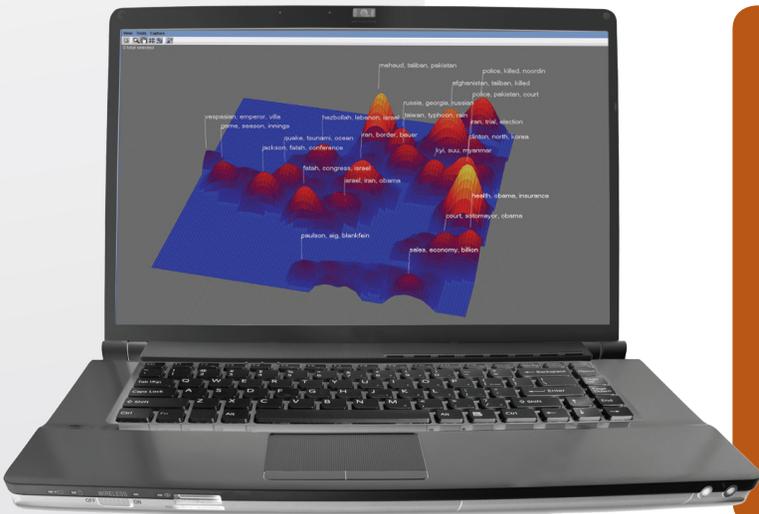
**Pacific Northwest**  
NATIONAL  
LABORATORY

*Proudly Operated*  
by **Battelle** Since 1965

Information and Visual Analytics at PNNL

# Pushing the Boundaries of Information and Visual Analysis

The U.S. Department of Energy's Pacific Northwest National Laboratory (PNNL) has developed information and visual analytics software technologies that infuse breakthrough speed and opportunities for deeper discovery to the analysis and understanding of vast amounts of data. While many of these tools have already demonstrated strong commercial value, PNNL welcomes further industry feedback and interest in commercial opportunities specifically for its text analytics capabilities.



**PNNL's information and visual analytics tools complement a wide range of applications, including:**

- » Competitive and business intelligence
- » Customer analytics
- » Market research
- » Library and information sciences
- » Fraud-risk detection, analysis, and monitoring
- » Legal and intellectual property research
- » Scientific study and disease detection
- » Intelligence analysis
- » Law enforcement
- » Threat assessment
- » Social network analysis

## DIVING DEEP INTO TEXT ANALYTICS

**Huge amounts of structured and unstructured text information?** No problem. The PNNL software suite includes a variety of capabilities for the visualization and analysis of large amounts of structured and unstructured text. These tools facilitate rapid processing of a massive number of documents or other text sources to enable extraction of key data and subsequent organization, analysis, and visualization of these data.

In fact, PNNL's tools provide value across the entire data analysis and visualization process:





## COVERING THE GAMUT OF DATA ANALYSIS AND VISUALIZATION CAPABILITIES

### Access/Collect/Cleanse ►

- » Extract data from unstructured sources
- » Collect data from social and traditional media to uncover relationships—both written and human entity—not detectable in documents alone
- » Cleanse information to provide a normalized format for analysis
- » Process high-volume, real-time data streams
- » Automatically capture data provenance
- » Collect data across multiple languages (e.g., English, Spanish, Arabic, Chinese)

### Parse ►

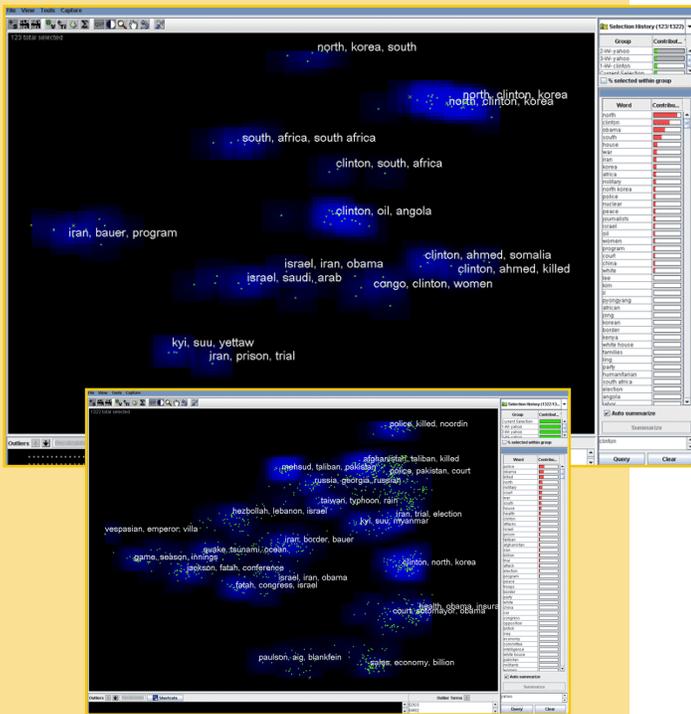
- » Automatically organize data by key topics without requiring users to develop keywords
- » Create keywords as phrases, multiple words, entities, and actions or events
- » Process data using both natural language and computational methods to provide greater insight into concepts and semantics—not just keywords
- » Organize documents thematically based on keywords and concepts to facilitate simplified interaction
- » Distinguish attitudes and sentiment through natural language processing

### Map/Vector/Cluster ►

- » Extract central themes from a pool of information
- » Develop a variety of relationships (e.g., network, geospatial, human entity, temporal)
- » Assess keyword prevalence
- » Search using concepts and semantic methods

### Project/Display/Visualize ►

- » Reveal key themes and their relationships via 2-D galaxy maps
  - Extract relevant documents through visual queries
  - Remap key themes dynamically as data is explored, based on user guidance and provided concepts
  - Zoom in on a particular concept of interest
- » View and visually query the relative importance of topics via 3-D elevation maps
- » Detect spikes or unexpected increases in keyword usage or themes via temporal analysis
- » Extract semantic information and construct link analysis diagrams to uncover relationships between entities (e.g., people, organizations, locations, publications, vehicles, etc)
- » Monitor dynamic data (e.g., chat room feeds, Google alerts) in real time to detect conversation and keyword patterns
- » Conduct concurrent visual analysis in multiple formats (e.g., see how key themes in a dataset change over time and impact both geospatial and human entity relationships)
- » Deploy a visualization toolkit to portable handheld devices such as PDAs and cell phones
- » Via handheld devices, collect real-time information from field personnel (e.g., GPS coordinates) and share information analytics across remote teams to improve operational awareness and decision-making capabilities

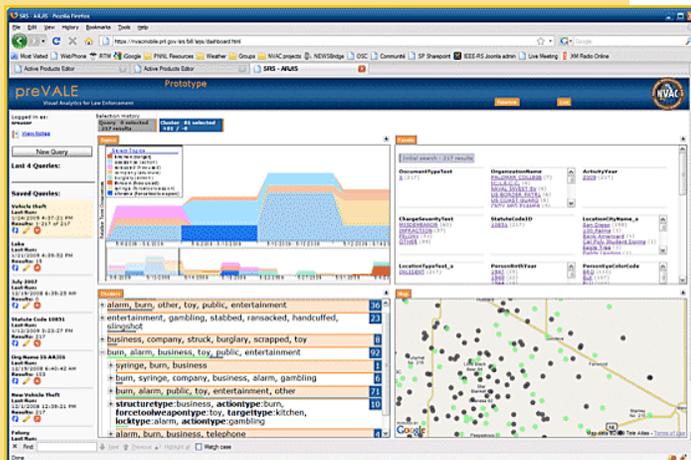
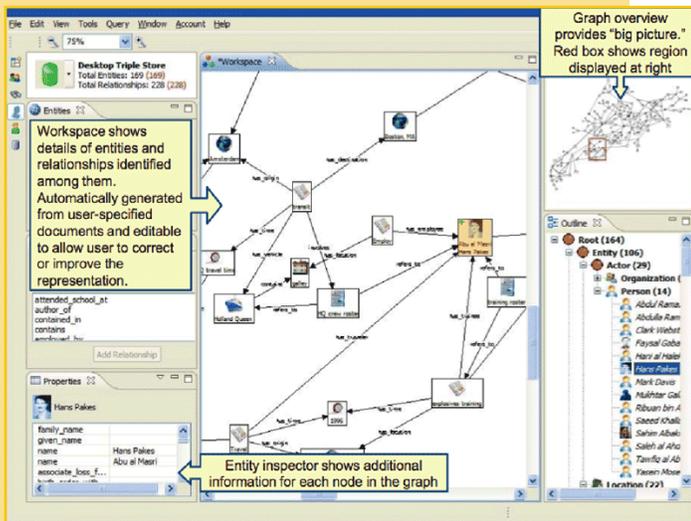


## Rank/Report/Analyze

- » Correlate keywords and dominant themes with temporal analysis to show how content themes change over time
- » Summarize blogs to highlight activity over a period of time, including most prevalent topics, user demographics, key opinion leaders, and human entity relationships
- » Compare keywords by their statistically derived signatures, extracted document count, total document count and via a query response from a search index
- » Create a “smart report” that provides automatic updates when new information is received from internal and external data sources (e.g., Web pages, news feeds)
- » Detect a “story” within a large set of documents, building a “narrative” from detected relationships

## Additional Capabilities

- » Analyze photos and videos to identify social relationships from facial recognition
- » Develop models based on existing sets of documents that can be used to automatically discover new relevant media
- » Leverage visualization capabilities within common software applications (Microsoft® software programs, desktop environments)
- » Apply thematic analysis and visualization to spreadsheet data
- » Create electronic briefing books
- » Consolidate data into a single repository for use by multiple data visualization tools
- » Integrate multiple data visualization tools within a common user interface



PNNL’s information and visualization technologies help analysts in numerous information-driven fields to:

(Top) reveal key themes and relationships via 2-D galaxy maps (the inset is a galaxy view of over 1,300 documents clustered by keyword content and relationships, while the larger image has been drilled down from the larger set to those containing a specific keyword)

(Middle) extract semantic information and construct link analysis diagrams to uncover entity relationships (here, an entity relationship diagram shows links between people, organizations, locations, publications, vehicles, and more)

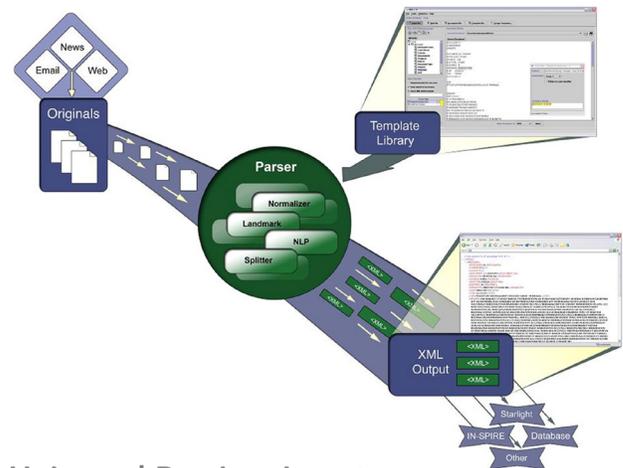
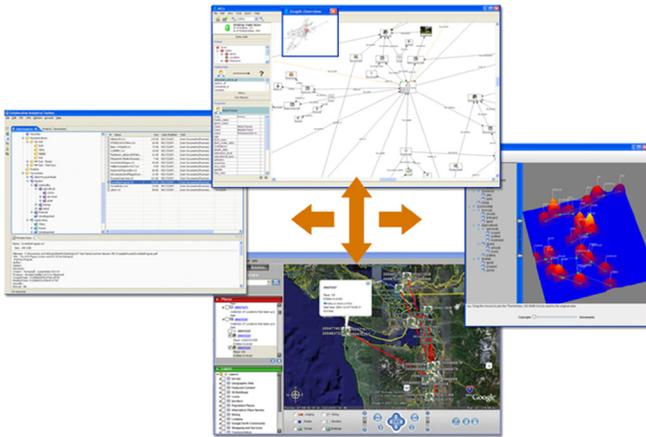
(Bottom) conduct concurrent visual analysis in multiple formats

(Cover) view and visually query the relative importance of topics via 3-D elevation maps (this image provides an elevation view of the same 1,300+ documents shown above on this page).

## Robust Tools with Powerful Commercial Potential

Many of PNNL's tools are available—or may soon become available—for license by commercial partners. Additionally, they represent numerous opportunities for further collaborative research with industry. If you'd like to know more about doing business with us, please visit <http://www.pnl.gov/business/>. To find out more about technologies currently available for license, visit <http://availabletechnologies.pnl.gov/>. Or give us a call. We welcome a conversation with you.

## JUST A FEW OF PNNL'S VISUALIZATION SOFTWARE TECHNOLOGIES INCLUDE

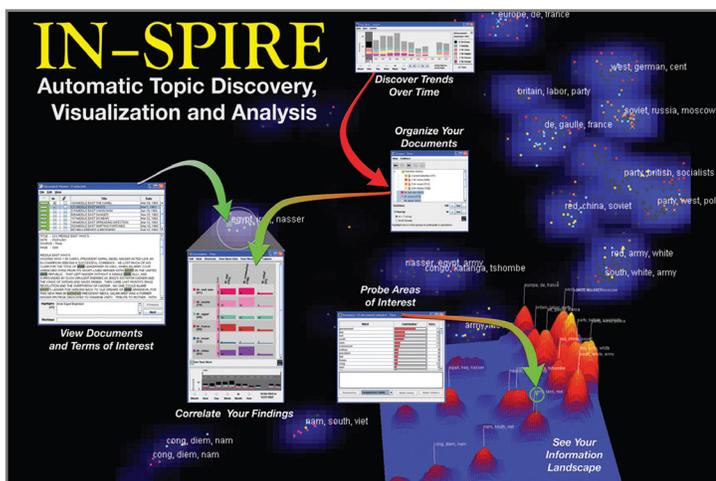


### Fused Analytic Desktop Environment (FADE)

PNNL's FADE suite represents a robust collection of information analysis tools not previously available in a consolidated solution. This integration of five PNNL-developed software technologies allows analysts to uncover relationships among entities in data that likely would have otherwise gone undiscovered—greatly strengthening understanding and decision-making.

### Universal Parsing Agent

One of the challenges associated with importing information from heterogeneous data sources is the diversity of formats presented. The Universal Parsing Agent (UPA) eases this challenge by providing a range of processing templates associated with common formats, greatly expanding the range and volume of document types that can be readily ingested and outputted for analysis using visualization applications. For more information about UPA, visit <http://availabletechnologies.pnl.gov/technology.asp?id=100>.



Dave Thurman  
Program Manager  
Pacific Northwest National Laboratory  
(206) 528-3221  
dave.thurman@pnl.gov

### IN-SPiRE™

Using statistical analysis, IN-SPiRE sorts through a set of documents and determines which are most thematically similar. This enables users to rapidly discover both known and hidden information relationships, and focus further time and effort only on the information most pertinent to their task rather than sifting through seemingly endless volumes of data. The tool has received an R&D 100 Award, and is currently available for license. For more information about IN-SPiRE, visit <http://availabletechnologies.pnl.gov/technology.asp?id=129>.



**Pacific Northwest**  
NATIONAL LABORATORY

Proudly Operated by **Battelle** Since 1965