



Text Analytics with IN-SPIRE™

The ability to uncover relationships or trends hidden in data can lead to new knowledge or insights in a variety of fields. IN-SPIRE™ text analysis software works with a person's natural ability to discover and learn from visually conveyed information. Its powerful underlying analytic algorithms extract key entities and topical information which are then conveyed through a series of engaging, interactive visualizations. Created by researchers at Pacific Northwest National Laboratory, IN-SPIRE can be applied to variety of document types including plain text, PDF, and Microsoft Word.

CHALLENGE

In today's data rich environment, analysts can be buried in information, often in the form of narrative textual reports or web pages. In the past, analysts have typically used keyword queries or manual scanning of titles to find relevant information. While keyword queries can be effective, the analyst must know what to look for prior to

the query. Manual scanning is not only time consuming but is only effective when the body of information being scanned is relatively small. IN-SPIRE provides a solution for those cases where analysts need to review thousands of documents for new information or insights that cannot be defined ahead of time.

SOLUTION

IN-SPIRE quickly and automatically conveys the gist of large sets of unformatted textual data such as technical reports, social media, web data, newswire feeds and message traffic. A series of statistical information extraction and clustering algorithms prepare the data for visualization without the need for advanced key wording or training. Its integrated suite of analysis tools presents the information in a suite of interconnected, interactive visualization components that enable the user to see common themes and hidden relationships within the data collection. IN-SPIRE can process both static



Query and display tools enable users to explore trends and relationships in the data, both quantitatively and qualitatively.

collections of documents, and streaming data such as RSS feeds. Its statistical algorithms also enable analysis of foreign language.

IN-SPIRE provides a number of query and display tools to support deeper analysis and interrogation of the information space. The software supports query capabilities including Boolean word search, wild cards, date range, numeric field, and “fuzzy” search based upon topical content. Visually oriented tools help you explore trends over time, perform categorical group correlation, and discover relationships between concepts. Documents can be categorized according to keywords or field values and ranked according to lists of terms (lexicons) that correspond to subject domains or sentiment content. The tools in IN-SPIRE enable both qualitative and quantitative measurement of language features in the dataset, plus tools for discovery of previously unknown phenomena, such as the identification of changes in the use of keywords and phrases over time.

IMPACT

IN-SPIRE allows analysts to spend more time exploring the information they find most relevant and less time sifting through the masses of irrelevant documents. Its data driven approach can be applied to any corpus of documents, including technical and patent literature, social media, marketing and business documents, web data, accident and safety reports, newswire feeds, message traffic, and more. It has applications in many areas, including information analysis, situational awareness monitoring, strategic planning, and medical research.

ABOUT PNNL

Interdisciplinary teams at Pacific Northwest National Laboratory address many of America’s most pressing issues in energy, the environment and national security through advances in basic and applied science. PNNL employs 4,300 staff, has an annual budget of nearly \$1 billion, and has been managed for the U.S. Department of Energy by Ohio-based Battelle since the laboratory’s inception in 1965.

For more information, contact:

Gus Calapristi

Project Manager
 gus.calapristi@pnnl.gov
 (509) 375-6853

Dave Thurman

Manager, Analytics Programs
 National Security Directorate
 dave.thurman@pnnl.gov
 (206) 528-3221



Proudly Operated by **Battelle** Since 1965