



DARISM

Data Analytics & Research in Social Media

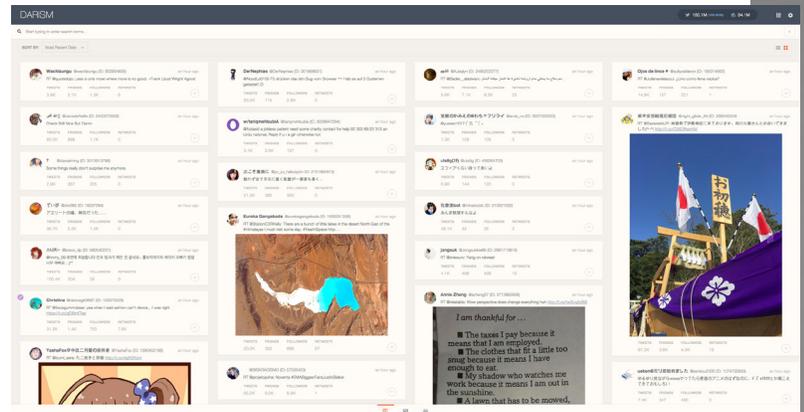
Performing analytics across the entire landscape of social data is often impractical, inefficient, and cost prohibitive. Researchers at the Pacific Northwest National Laboratory (PNNL) have developed the Data Analytics & Research in Social Media (DARISM) platform as a means to perform large-scale analysis of social data for the most popular social media platforms, making the “fire hose” of data much more manageable and informative for decision makers.

CHALLENGE

Social media environments have broken the traditional information flow from media-driven to consumer-driven. Through online environments, anyone can be a producer of information which has resulted in a shift in how people discover, read, and share news, information, and content. This shift allows social networks to provide early indicators of security and health threats, public sentiment, and actor motivation. The challenge arises in trying to leverage that information for appropriate action. Due to the massive volume of streaming data, analysis becomes overwhelming, difficult, and complicated. Moreover, the variety, veracity, and velocity of social data requires sophisticated algorithms and methodologies to achieve useful insight. Beyond scaling issues, human language algorithms that worked for longform documents fail when applied against the new social “shortform.”

SOLUTION

DARISM provides a suite of analytic capabilities to enrich the social data stream. These capabilities include sentiment and psychological analysis, user characterization, link unwrapping and characterization, trend identification, and media harvesting. Its web application provides users tools such as customizable alert mechanisms, social network analysis, and geographical and image/video views.



The DARISM workflow is characterized by the following three simple ideas—discover, triage, and analysis—that enables decision makers to effectively analyze this novel open-source media type. In the discovery stage, a user-friendly web application interface allows the user to search across massive amounts of social media data, and set up future data requirements by defining delivery filters. A triage stage presents a commonly used faceted browsing technique so the user can tightly define their interest and reduce to the critical social media objects that help answer their query. Available facets include common characteristics of social data such as hashtags, authors, and location, as well as computationally defined characteristics such as language detection, and sentiment/psychological analysis. Multiple views provide insight, such as a display that shows the results in data-rich tables, geographically on a world map, or provide insight into the shared media being mentioned in the post.

Finally, the analysis stage teases apart the social data collected to provide insightful commentary on analytical questions through means such as social network analysis, thematic visualizations, and multimedia presentations. The entire experience is wrapped around a walk-up usable design, created through ethnographic study, and task-focused interviews of knowledge workers across the U.S. government enterprise.

