



COLLECT
Documents and metadata on identified technology domains are gathered from **open-source databases**. Collection can be automated through APIs and done via custom import of text files, Excel workbooks, etc.



PREPROCESS
Text and metadata are preprocessed and formatted into a **common document template**. Preprocessing focuses on cleaning up text, normalizing names, fetching geolocation data, etc.



EMBED
Documents are transformed into **fixed-length feature vectors**, giving each document a unique "fingerprint." Similar documents will be geometrically closer together in "feature space."



IDENTIFY
Feature vectors are fed into a graph-based **clustering algorithm** that identifies the fastest-growing trends. The algorithm also finds the most important documents for each trend.



SUMMARIZE AND VISUALIZE
Trends are summarized using a combination of **automated and expert-driven analytics**. Output can include topic summaries, keywords, geographic distributions, author networks, and many other insights.



DSIAC Success Story

Novel Horizon-Scanning Approach

www.dsiac.org

Customer:	Air Force Research Lab (AFRL/CZ)
Challenge:	Establish a repeatable, automated process to facilitate big data analysis and horizon scanning of emerging science and technology, with disruptive innovative potential to address Air Force (AF) warfighter requirements in 2030 and beyond.
Approach:	DSIAC researched tools, techniques, and processes for conducting big data analysis and horizon scanning to identify potentially disruptive innovative technologies capable of addressing targeted AF challenges. Upon completing an assessment of alternatives, DSIAC established a horizon-scanning approach employing modern machine-learning techniques. This "wide-aperture" approach avoids possible institutional biases by using broad search criteria and leveraging clustering algorithms to automate the identification of emerging science and technology.
Value:	This novel horizon-scanning tool provides the AF with a repeatable process for mining and analyzing big data automatically to identify innovative, disruptive science and technology ideas relevant to specific technical interests. The result minimizes the labor intensity of data analysis while streamlining the identification of emerging science and technology solutions for AF warfighter challenges.

The Defense Systems Information Analysis Center (DSIAC) is operated by SURVICE Engineering Company under contract number FA8075-14-D-0001.

Distribution Statement A: Approved for public release; distribution is unlimited.