MFGS INC

Plan. Build. Deliver. Run.

SIMPLIFY DATA ANALYSIS AND CORRELATION FOR IMPROVED NATIONAL SECURITY

Collect, analyze, and visualize your data with OpenText™ Knowledge Discovery and OSINT

As of 2020, there were 4.8 billion internet users creating in excess of 2.5 quintillion bytes of data every day on websites and social media feeds[1]. Much of this information is benign when evaluated on its own. Yet, there remains a significant amount of dangerous data that when combined with the 'harmless' data, generates risk to industry, populations, and governments. Access to public information can help to thwart these threats. However, without the right software solution in place, it's impossible to filter through and connect the virtual sea of available data – no matter how many human resource hours are thrown at the task.

Using a tool created to first ingest your own data and then connect it to open-source intelligence (OSINT) means it's possible to secure and plan an incident response, or perhaps even preempt an incident in the first place. Although this information is available in the public domain, government agencies have certain restrictions around the manner OSINT tools are used in relation to sensitivity and legal protocol, as they leverage publicly available information for use in:

- National security
- Counterterrorism
- Cyber tracking terrorists
- Supplying policymakers information to support decision making
- and more

About OpenText™ Knowledge Discovery

OpenText™ Knowledge Discovery is an ideal solution for monitoring and analyzing publicly available information sources, such as websites, social media feeds, video streams, and television and radio broadcasts, in over 150 different languages. And, Knowledge Discovery accomplishes all of this within the restrictions imposed by legal and ethical requirements.

Knowledge Discovery is an artificial intelligence-based unstructured data analytics platform for video, image, audio, and text data; it removes data silos, monitors real-time data feeds, uncovers covet trends, patterns, and relationships, and saves massive labor resources while streamlining workflows. By virtually tapping any data repository using the extensive Knowledge Discovery Connector Framework, Knowledge Discovery serves as a nexus point of knowledge discovery, enabling unprecedented insights and offering advanced natural language processing (NLP) functionality. In June of 2024, Knowledge Discovery was named a leader by Forrester Wave in the AI-Based Text Analytics platform.

Knowledge Discovery provides unified text, speech, and video analytics with support for over 1,100 data formats. It enables out-of-the-box

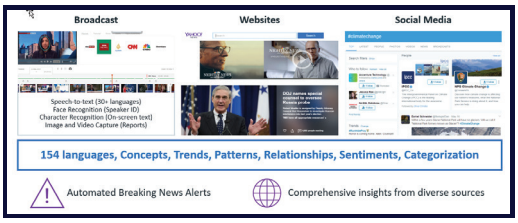


Figure 1 - Access a broad range of sources and data types

Customer Success Story: International Agency - Advanced OSINT analytics to identify risk

Challenge

- Agency closely monitors published media, social media and public data, and more to identify developing situations and trends.
- Mitigate risks from important information being overlooked.
- Multiple languages across the globally dispersed locations of member states.

Solution

- Knowledge Discovery Enterprise Search and Collaboration

Result

- Identify patterns and highlight potential security risks at an early stage, enabling the agency to better coordinate responses.
- Enabled staff to find relevant information quickly and easily, regardless of where or how it is stored.
- Improved collaboration, identify similar issues, avoid duplication.

[1] https://seedscientific.com/how-much-data-is-created-every-day/

access to 150 data repositories behind and beyond your firewall, e.g. Facebook, Twitter, YouTube, and Dropbox. It can index data without relocation and disruption. Built on proven, world-class technology and innovations such as machine learning and deep neural networks, Knowledge Discovery unlocks hidden insights by revealing trends, patterns, and relationships.

With modular offerings of hundreds of advanced analytics functions, as well as an open and scalable architecture for easy third-party integration, Knowledge Discovery lends itself to supporting diverse use cases spanning a broad spectrum of industries, especially OSINT.

Knowledge Discovery provides an automation platform to deliver a regular process flow for ingestion, enrichment, and analysis of collected publicly-available open source information. The result is a corpus of normalized, processed content that is deliverable to decision makers through dashboards, or available for interactive interrogation to identify patterns and information of interest.

Knowledge Discovery detects noteworthy correlations, patterns, trends, and more across a broad set of publicly available data sources, with a high level of confidence. The insights gleaned from using Knowledge Discovery for OSINT benefit an organization massively. For instance, strategic planning for a large public event or state visit that requires detailed and comprehensive background research into people, places, and previous occurrences can be quickly completed with the highest levels of accuracy and efficiency. Further, in an enterprise setting, analysis of specific companies with regard to competitors and stability can be done as a step prior to making a large investment.

Knowledge Discovery gives you the ability to connect seemingly disparate data, locations, persons, events, etc. across multiple sources to unmask critical trends, patterns, and relationships, validating developing intelligence and spotting nuances. It assists in answering critical questions, such as: What are the key topics of interest across the world? How do they differ? What are the public opinions around the globe? How do they differ?

Key Supporting Features for OSINT Provided by Knowledge Discovery

Ingest Near-Real-Time Social Media Feeds

Knowledge Discovery has a large family of connectors able to connect to and ingest data from hundreds of data repositories at an API level. Connecting at an API-level allows for Knowledge Discovery solutions to perform additional actions beyond ingestion if supported by the repository, allowing for the management of data in place. These connectors facilitate near-real-time ingestion of all data on these repositories, including from social media platforms such as Twitter.

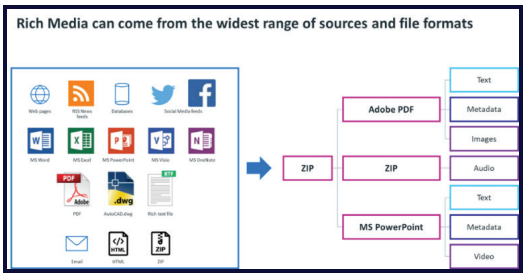


Figure 2 - Knowledge Discovery supports over 150+ data sources and 1100+ formats

Complete Analysis of Ingested Data

Knowledge Discovery utilizes its large library of connectors and KeyView file filter technologies to ensure all textual content and metadata (both file and repository-based) are extracted and indexed into the Knowledge Discovery platform for search. Knowledge Discovery has hundreds of functions designed for analyzing unstructured data and supports a wide variety of retrieval methods including keyword, metadata, faceted/parametric, Boolean, and conceptual.

Customizable Grammar Libraries

Knowledge Discovery can automatically identify and extract terms in documents that lend themselves to key fields, such as the names of weapon systems or people, locations, addresses, and telephone numbers. Knowledge Discovery offers several thousand entity grammars out of the box across 150+ languages. Organization-specific entities are easily trained.

Video, Image, and Streaming Media Capabilities

Optical Character Recognition (OCR) is used to extract text from image files. The use of OCR on

scanned or photographed documents and pictures facilitates their conversion into a computer-readable format for easier storage and analysis. Upon the ingestion of the input of the image, Knowledge Discovery's Visual Server can return the identified text, the confidence score, and detect the region of the image in question. The detection region can be bound to certain areas to decrease noise from the rest of the image, and the region-accuracy can be fine-tuned to the position of each character; for example, video feed of a train station platform can exclude the area with tracks and train travel, and highlight only the station platform activity. Knowledge Discovery visual and audio analysis supports face, object, speech-to-text, speaker detection, and much more, to provide a complete index of human-generated information.

Comprehensive Search of Open Source Data

Knowledge Discovery is designed to index and analyze unstructured data in all its forms, and supports a wide variety of retrieval methods including keyword, metadata, faceted/parametric, Boolean, and conceptual. These searches are all performed in near real time, allowing for users to see results immediately.

Advanced Media Analytics

Media Management and Analytics Platform (MMAP) provides all components necessary to facilitate analysis, management, and visualization of large volumes of rich media alongside traditional text-based sources. It allows video, audio, and image analysis and enrichment, real-time monitoring and alerting, and search and retrieval across a wide array of sources such as audio/video streams, CCTV, broadcast TV, radio, file-based rich-media archives, and social media video, image, and audio content.

Graphical Visualization and Correlation (trends, topics, geospatial, etc.)

Knowledge Discovery makes it easier than ever to correlate and visualize complex data types from multiple sources. So long as geospatial/time metadata exists or can be extrapolated from the content of the document or message, the new data points can immediately be plotted on maps and time horizons. The data can be further segmented by concept, analyzing result sets by what is being said or how those concepts evolve over time – including public reaction to them.

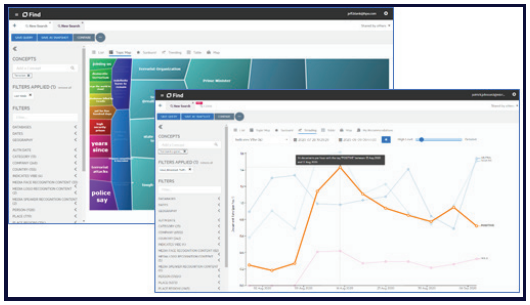


Figure 3 - Topic map of related concepts specific to search results. Sentiment charting of search results using Trending Visualization.

Knowledge Discovery can extract and process any metadata or textual content from documents and repositories. Through the use of entity extraction, information can also be extracted from text, creating additional metadata tags. This information can then be plotted on trend analysis tools built into the out-of-the-box FIND GUI.

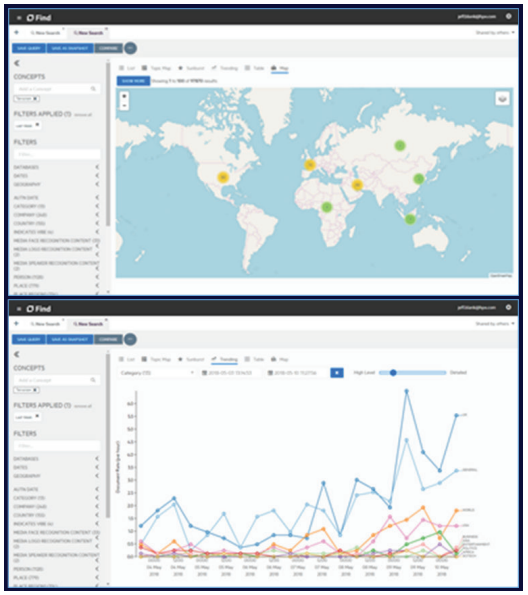


Figure 4 - Map Visualization and Trending Visualization

Alternatively, Knowledge Discovery natively supports categories which can be trained in a number of ways including: conceptual keyword, Boolean, metadata, and via example training documents. These categories can be audited constantly against the indexed data and, if a sudden uptick in events matching the category are found, can be used to alert users.

Sentiment Analysis on Open Source Data

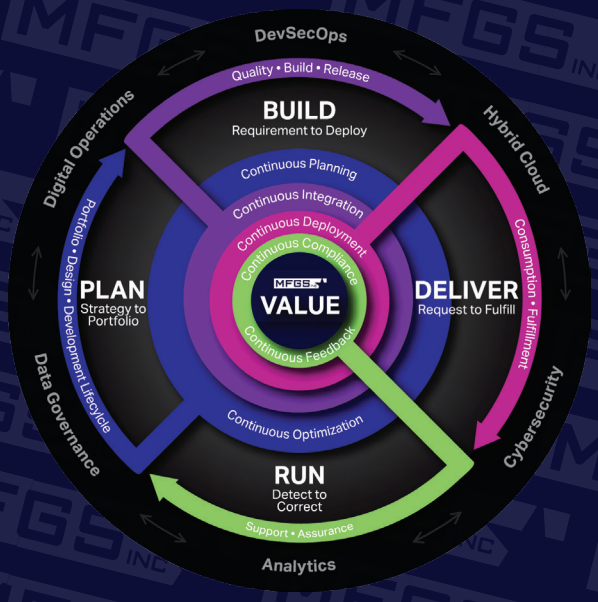
Knowledge Discovery can perform sentiment analysis upon any block of text. Typically performed at index time, the Knowledge Discovery platform will provide a general vibe (positive/neutral/negative) and a score based upon the full body of text presented to it. These tags can then be analyzed by Knowledge Discovery in the same manner as any other piece of metadata.

The Knowledge Discovery platform is language agnostic, able to index documents of any language, identify what language they are, and apply appropriate linguistic rules upon them. This capability is supplemented by a wide range of additional language packs, including those for entity extraction (detecting PII, abiding by US FIPS, EU GDPR, etc.), sentiment analysis, OCR, and speech-to-text.

Summary Takeaways

Similar to technology advancements, the sheer volume of information and the speed at which it gets created increases exponentially every year. While most of that data is innocuous, other data presents a risk to national security, or can lead to further information about impending dangers. To be most effective, it's imperative to simultaneously correlate internal information with open-source intelligence, which Knowledge Discovery has been able to collect and analyse for many years.

Yet, collecting, analyzing, and acting on that data is an impossible task when attempted solely with human brain power and resources. A tool like Knowledge Discovery is critical in the need to accurately and effectively complete the data analysis tasks - thereby giving back to agencies the time to successfully execute against the other portions of their missions.



MFGS, Inc. Mission Dominance Model

MFGS, Inc.'s Knowledge Discovery fits into the foundation of the MFGS, Inc. Mission Dominance Model. The model aims to illustrate how each area of enterprise software architecture interconnects to support the software development lifecycle. The Foundations outer ring bolsters each of the phases of Plan, Build, Deliver, and Run in driving Mission Dominance.

MFGS INC
Plan. Build. Deliver. Run.

About MFGS, Inc.

MFGS, Inc. is the trusted advisor to the U.S. Government, its partners, and system integrators for achieving optimal efficiency throughout an agency's enterprise software architecture. We bring a comprehensive portfolio of enterprise-grade software solutions and a deep understanding of how DOD agencies operate to support your entire software development lifecycle, enabling you to securely plan, build, deliver, and run your mission.

Learn more:

mfgsinc.com
linkedin.com/company/mfgs-inc
@MFGSInc