

8 Jul 2008

Putting the brain back into intelligence



Fingers on keyboard

As intelligence agencies fail to keep up with massive volumes of data and shrinking reaction times through automation - enter the human analyst.

<http://www.isn.ethz.ch/isn/Current-Affairs/Security-Watch/Detail/?id=88550&lng=en>

By Peter Buxbaum in Washington, DC

Many defense and intelligence organizations around the world collect and analyze data. One such agency, the US National Geospatial-Intelligence Agency in Bethesda, Maryland, expects to gather four petabytes of data annually in coming years. (One petabyte equals 1,000 terabytes or one-million gigabytes.) That is equivalent to the volume of data of every movie ever made since movies began on DVDs.

Every computer user knows how to extract information from large volumes of data: simply use a search engine like Google. But the service doesn't quite fit the bill for intelligence applications. The data volume is too large and yields too broad a result with a simple keyword search. Narrowing the search results doesn't work either because it takes too long. Intelligence analysts often want to scrutinize and analyze data as it arrives.

"We find a real sense of frustration across the intelligence community," Guljit Khurana, CEO of Centrifuge Systems, a technology company based in McLean, Virginia, told ISN Security Watch. "The volume and velocity of data is growing while the time frame in which to act on the data is shrinking."

The solution until now for intelligence agencies has been to develop enormously complex Boolean queries which were tested to achieve specific search results. But this approach, too, has its weaknesses. For one, it requires computing expertise to develop the queries; additionally, the queries themselves tend to remain static while the search needs of the intelligence community tend to be dynamic.

"With the technology available today, an analyst spends more time figuring out what is the right query than anything else," Brian Futchey, a solutions architect with Endeca Technologies in Cambridge, Massachusetts, told ISN Security Watch. "Queries for searching on current applications can run to three or five pages long."



7926 JONES BRANCH DR SUITE 210 MCLEAN VA 22102 / P 571.830.1300 / F 571-830-1301 / CENTRIFUGESYSTEMS.COM

"Automated applications can rapidly scan data but they can't replace the carbon-based unit in the equation when it comes to uncovering insights," added Khurana. "The single most important component remains human judgment."

The US Army's Special Operations Command 95th Civil Affairs Brigade has run into similar problems when it comes to analyzing data. The unit, which assists civilian populations in battle zones, provides humanitarian relief in disaster areas, develops civil infrastructures, and advises governments on their relationships with local populations. It collects and analyzes large amounts of data to assist with these tasks.

Technology's dark underbelly

One of the applications used by the Brigade, Analyst's Notebook by i2 Technologies, Inc in McLean, Virginia, provides an environment for creating information linkages and timeline analyses. The software enables insights into factors contributing to an issue, evaluating their causes and effects, and understanding the relationships between them.

"Analyst's Notebook helps us see linkages between things and to peel back layers of information to see the connections between people and places," Colonel Ferdinand Irizarry, the units commander, told ISN Security Watch.

But Irizarry also complained his unit has confronted a "dark underbelly of technology." "What we have found is that most of this software is not all that user friendly," he said. "It takes a lot of training to teach how to use it. The software is flexible and can serve a lot of different requirements but it is inherently complex. It requires the operator to memorize many commands and be familiar with the full breadth and depth of the software's functionality."

"It is like driving a car that requires you to crank the engine and adjust the choke," he added. "Most people these days are used to point and click functionality in software."

Technology companies are trying to solve these problems by "focusing on human cognition rather than technology," said Frutchey. "We want to commoditize information searching so that the user doesn't have to think about how it works any more than the average person has to think about how a refrigerator works. All you know is that it is keeping your food cold."

Endeca is working on technologies that organize masses of information in ways that enable them to question the user about his intentions, leading to better information extraction results. Frutchey analogizes the process to a hotel guest approaching the concierge. "The guest might ask, 'What is there to do around here?' In response, the concierge is not going to whip out a list of everything," Frutchey explained. "He is going to respond with questions of his own, by asking the guest what he is interested in doing."

Current search engines use a linear approach to searching hierarchies of data. The next generation of search is able to better put information into context by using a number of techniques which find relationships among data and user intent.

A faceted search is a model in which documents are associated with a variety of independent information facets. "For example," Frutchey explained, "the word 'Princeton' might refer to a university or to a geographical area. A faceted search allows the user to restrict the meaning of search terms and obtain a better result. The facets create a navigation space that allows users to meet initial search goals while also learning from the information to discover new goals."



7926 JONES BRANCH DR SUITE 210 MCLEAN VA 22102 / P 571.830.1300 / F 571-830-1301 / CENTRIFUGESYSTEMS.COM

Another advanced information extraction system is called dynamic clustering, which organizes the metadata tags associated with documents into distinctive groupings that can be used to refine a query. Metadata refers to searchable tags appended to documents, indicating what data knows about itself.

"The point is that the feedback has to be contextual," said Frutchey. "Instead of producing a massive list of documents, they are clustered around topics. The key is to move beyond query response to empower users through dialogue."

Human cognition

Another method that allows human cognition to interact with automated information extraction involves data visualization, an area in which Centrifuge Systems specializes. "Visual metaphors allow humans to grasp massive amounts of data and to draw inferences," said Khurana. "The human brain is enormously effective at detecting patterns in that way."

In this context, visualization involves more than producing a static chart that appears in a PowerPoint slide. "It is part of the analysis process itself," explained Khurana. "We humans have evolved a remarkable pattern recognition capability over a long development cycle. Visual representations of data allow us to draw inferences almost instantly. Visualization allows analysts to find subsets within masses of data and to identify specifically what is relevant."

Visualization has also been merged with link analysis, allowing analysts to click on an aspect of a chart that is of interest to expose additional connections or to drill down to get more details. The technique has also been used to model social networks, through which analysts can identify leaders and non-obvious interconnections among people and groups.

Another popular application of visualization involves temporal timeline analysis. "This maps sequences of events over time," said Khurana. "It tends to yield insights that jump right out at you."

"Visualization allows data analysis to take place at the speed of the human brain," he added. "It is a highly interactive model for deriving insights from data."

The next step in improving intelligence analysis will likely also come on the human, rather than the technology, front, according to Khurana. "Intelligence analysts too often work in isolation without sharing their insights with their colleagues," he said. "The next big push will be to develop a collaborative framework for intelligence analysis."

Peter Buxbaum, a Washington-based independent journalist, has been writing about defense, security, business and technology for 15 years. His work has appeared in publications such as Fortune, Forbes, Chief Executive, Information Week, Defense Technology International, Homeland Security and Computerworld. His website is www.buxbaum1.com.

To learn more about Centrifuge Systems, please visit www.centrifugesystems.com