

By Robert Popp,  
Thomas Armour,  
Ted Senator, and  
Kristen Numrych

# COUNTERING TERRORISM INFORMATION TECHNOLOGY

**S**eptember 11, 2001 might have been just another day if the U.S. intelligence agencies had been better equipped with information technology, according to the report of Congress's Joint Inquiry into the events leading up to the Sept. 11 attacks [9]. The report claims that enough relevant data was resident in existing U.S. foreign intelligence databases that had the "dots" been connected—that is, had intelligence analysts had IT at their disposal to access and analyze all of the available pertinent information—the worst foreign terrorist attack to ever occur on U.S. soil could have been exposed and stopped.<sup>1</sup>

In the aftermath of the Sept. 11th terrorist attack, the U.S. Defense Advanced Research Projects Agency (DARPA)—the U.S. Defense Department agency that engages in high-risk/high-payoff research for the defense department and national security community—focused and accelerated its counterterrorism thrust. The overarching goal was to empower users within the foreign intelligence and counterterrorism communities with IT so they could anticipate and ultimately preempt terrorist attacks by allowing them to find and share information faster, collaborate across multiple agencies in a more agile manner, connect the dots better, conduct quicker and better analyses, and enable better decision making [5, 8].

<sup>1</sup>It is important to point out the implicit assumption—and the resultant misconception—that all the dots were unambiguously preexisting; a significant portion of them were created by some a priori context, which may or may not have been correct. Moreover, when the Sept. 11 terrorist event shifted into the past, that ambiguous context became much clearer because the event then became part of the historical record [6].

# THROUGH

# GY

*Developing the information-analysis tools for an effective multi-agency information-sharing effort.*

**T**he world has changed dramatically since the Cold War era, when there were only two superpowers (see Table 1). During those years, the enemy was clear, the U.S. was well postured around a relatively long-term stable threat, and it was fairly straightforward to identify the intelligence collection targets. Today, we are faced with a new world in which change occurs very rapidly, and the enemy is asymmetric and poses a very different challenge; the most significant threat today is foreign terrorists and terrorist networks whose identities and whereabouts we do not always know.

What is the nature of the terrorist threat? Historically, terrorism has been a weapon of the weak characterized by the systematic use of actual or threatened physical violence, in pursuit of political objectives, against innocent civilians. Terrorist motives are to create a general climate of fear to coerce governments and the broader citizenry into ceding to the terrorist group's political objectives [7]. Terrorism today is transnational in scope, reach, and presence, and this is perhaps its greatest source of power. Terrorist acts are planned and perpetrated by collections of loosely organized people operating in shadowy networks that are difficult to define and identify. They move freely throughout the world, hide when necessary, and exploit safe harbors proffered by rogue entities. They find unpunished and oftentimes unidentifiable sponsorship and support, operate in small independent cells, strike infrequently, and utilize weapons of mass effect and the media's response in an attempt to influence governments [1].



Information Technology	Description
<b>Biometrics</b>	Identify and/or verify human terrorist (or watchlist) subjects using 2D and 3D modeling approaches over a variety of biometric signatures: face, gait, iris, fingerprint, voice. Also exploit multiple sensor modalities: EO, IR, radar, hyper-spectral.
<b>Categorization, Clustering</b>	Employ numerous technical approaches (natural language processing, AI, machine learning, pattern recognition, statistical analysis, probabilistic techniques) to automatically extract meaning and key concepts from (un)structured data and categorize via an information model (taxonomy, ontology). Cluster documents with similar contents.
<b>Database Processing</b>	Ensure platform, syntactic and semantic consistency and interoperability of multiple types of data stored on multiple storage media (disk, optical, tape) and across multiple database management systems. Desirable aspects include flexible middleware for: data location transparency and uncertainty management, linguistically relevant querying tuned for knowledge discovery and monitoring, scalability and mediation, schema evolution and metadata management, and structuring unstructured data.
<b>Event Detection and Notification</b>	Monitor simple and complex events and notify users (or applications) in real time of their detection. Monitoring can be scheduled a priori, or placed on an ad hoc basis driven by user demands. When an event is detected, automatic notifications can range from simple actions (sending an alert, page, or email) to more complex ones (feeding information into an analytics system).
<b>Geospatial Information Exploitation</b>	Fuse, overlay, register, search, analyze, annotate, and visualize high-resolution satellite and aerial imagery, elevation data, GPS coordinates, maps, demographics, land masses, political boundaries to deliver a streaming 3D map of the entire globe.
<b>Information Management and Filtering</b>	Collect, ingest, index, store, retrieve, extract, integrate, analyze, aggregate, display, and distribute semantically enhanced information from a wide variety of sources. Allow for simultaneous search of any number of information sources, sorting and categorizing various items of information according to query relevance. Provide an overall view of the different topics related to the request, along with the ability to visualize the semantic links relating the various items of information to each other.
<b>Infrastructure</b>	Provide comprehensive infrastructure for capturing, managing, and transferring knowledge and business processes that link enterprise software packages, legacy systems, databases, workflows, and Web services, both within and across enterprises. Important technologies include Web services, service-oriented grid-computing concepts, extensible component-based modules, P2P techniques, and platforms ranging from enterprise servers to wireless PDAs, Java, and Microsoft, .NET implementations.
<b>Knowledge Management, Context Development</b>	Use Semantic Web, associative memory, and related technologies to model and make explicit (expose via Web services) an analyst's personal preferences, intellectual capital, multidimensional knowledge, and tacit understanding of a problem domain.
<b>Predictive Modeling</b>	Predict future terrorist group behaviors, events, and attacks, based on past examples and by exploiting a variety of promising approaches, including neural networks, AI, and behavioral sciences techniques, subject matter expertise, and red teams.
<b>Publishing</b>	Generate concise accurate summaries of recent newsworthy items, ensuring users see topics only once, regardless how many times the item appears in data or in the press.
<b>Searching</b>	Allow users to perform more complete and meaningful searches (free text, semantic, similarity, partial or exact match) across a multitude of geographically dispersed, multilingual and diverse (un)structured information repositories within and across enterprises (any document type located on file servers, groupware systems, databases, document management systems, Web servers).
<b>Semantic Consistency, Resolving Terms</b>	Exploit ontologies, taxonomies, and definitions for words, phrases, and acronyms using a variety of schemes so users have a common and consistent understanding of the meaning of words in a specific context. Resolve semantic heterogeneity by capitalizing on Semantic Web technologies.
<b>Video Processing</b>	Analyze, detect, extract, and digitally enhance (reduce noise, improve image color and contrast, and increase resolution in selected areas) user-specified behaviors or activities in video (suspicious terrorist-related activities).
<b>Visualization</b>	Provide graphical displays, information landscapes, time-based charts, and built-in drill-down tools to help analysts and investigators discover, discern, and visualize networks of interrelated information (associations between words, concepts, people, places, or events) or visually expose non-obvious patterns, relationships, and anomalies from large data sets.
<b>Workflow Management</b>	Create optimized workflows and activities-based business process maps using techniques, such as intelligent AI engines by watching, learning, and recording/logging the activities of multiple users using multiple applications in multiple sessions.

**Table 2. Other information technologies considered important for counterterrorism.**

results recently obtained through experiments via partnerships that DARPA conducted with several entities within the U.S. intelligence and counterterrorism communities.

The purpose of the experiments was for analysts to assess the merits of several IT tools developed and integrated under DARPA sponsorship applied to various foreign intelligence problems. The experiments involved real analysts solving real foreign intelligence problems using their own lawfully collected foreign intelligence data. The tools provided by DARPA spanned the three core IT areas: peer-to-peer collaboration

however, there are numerous other information technologies that are equally important. Table 2 identifies and describes some of these core areas.

Figure 2 shows how the three core IT areas map onto a typical intelligence analysis process. These technologies will allow users to: search, query, and exploit vastly more foreign speech and text than would otherwise be possible by human translators alone; automatically extract entities and relationships from massive amounts of unstructured data and discover terrorist-related relationships and patterns of activities among those entities; and collaborate, reason, and share information, so analysts can hypothesize, test, and propose theories and mitigating strategies about possible futures, and to enable decision- and policymakers to effectively evaluate the impact of current or future policies and prospective courses of action.

We discuss each of these areas in more detail later in this article. Before doing so, however, we first underscore their critical importance by describing some promising

tools, structured argumentation and analytical tools, foreign language tools for audio searching/indexing and text and audio filtering/categorization, and statistical graph-based pattern analysis tools.

As Figure 3 shows, when doing traditional intelligence analysis, an analyst spends the most time on the major processes broadly defined as research, analysis, and production. The pink “bathtub curve” represents the distribution of time one typically sees.<sup>2</sup> This shows that analysts spend too much time doing research (searching, harvesting, reading, and preprocessing data for analysis), too much time doing production (turning analytical results into reports and briefings for the decision maker), and too little time doing analysis (thinking about the problem). The objective of the experiment was to see if intelligence analysis could be improved through

<sup>2</sup>The three numeric values associated with the pink curve are based on a real-world problem that will be described in more detail here. To interpret the curve, each value denotes the percentage of time—out of 100% total—an analyst spends on research, analysis, or production.

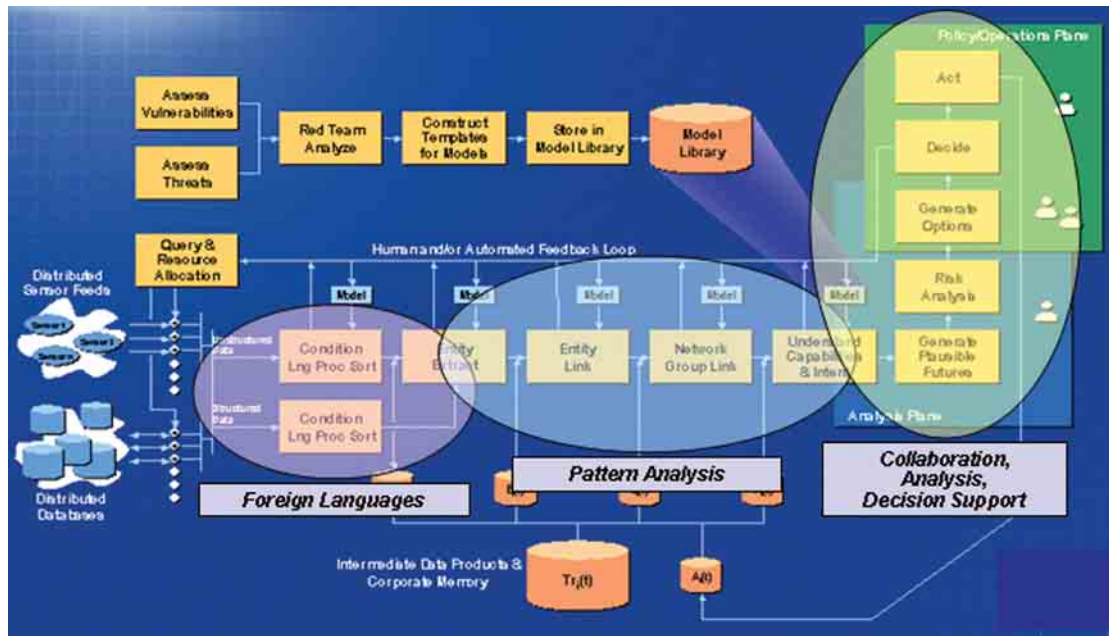


Figure 2. Three core information technology thrust areas mapped onto a typical intelligence analysis process.

IT by reversing this trend and inverting the bathtub curve.

In this experiment, the intelligence question the analysts were asked to analyze was “What is the threat posed by Al Qaeda’s Weapons of Mass Destruction capabilities to several cities in the U.S.?” The data was drawn from a variety of classified intelligence sources, foreign news reports, and Associated Press wire service reports.

The results of the experiment were impressive. As the yellow curve in Figure 3 shows, an inverted bathtub curve, allowing for more and better analysis in a shorter period of time, resulted when analysts used IT to aid their analysis. Results included an impressive savings in analyst labor (half as many analysts were used for the analysis), and five reports were produced in the time it ordinarily took to produce one. Moreover, the time spent in the research phase was dramatically reduced due mainly to using collaboration and foreign language tools to share and preprocess the foreign news and AP wire service data in 76 hours versus the 330 hours it took previously using more traditional manually driven methods.

## Collaboration, Analysis, and Decision Support

**C**ollaboration, analysis, and decision support tools allow humans and machines to analyze (think) and solve complicated and complex problems together more efficiently and effectively. These tools are what transform the massive amounts of data flowing into the government’s intelligence and counterterrorism communities into intelligence. Specifi-

cally, tools are needed to address each element of the “cognitive hierarchy,” namely, tools to transform data (discriminations between states of the world) into information (dots, or evidence, which is data put into context), and information into knowledge (useful and actionable information to decision makers).

*Enable center-edge collaboration.* Combating the terrorist threat requires all elements of the government to share information and coordinate operations. No one organization now has nor will ever have all the needed information or responsibility for counterterrorism. In addition to breaking down organizational barriers and sharing data, collaboration is also about sharing the thinking processes. Sharing of the thinking processes is about multiple perspectives and conflictive argument, and embracing paradox—all which enable humans to find the right perspective lenses in which to properly understand the contextual complexity through which correct meaning is conveyed to data. Collaboration tools permit the formation of high-performance agile teams from a wide spectrum of organizations. These tools must support both top-down, hierarchically organized and directed, “center-based” teams, as well as bottom-up, self-organized and directed ad-hocracies—“edge-based” collaboration. These two modes of operation must also be able to interoperate: “center-edge” coexistence.

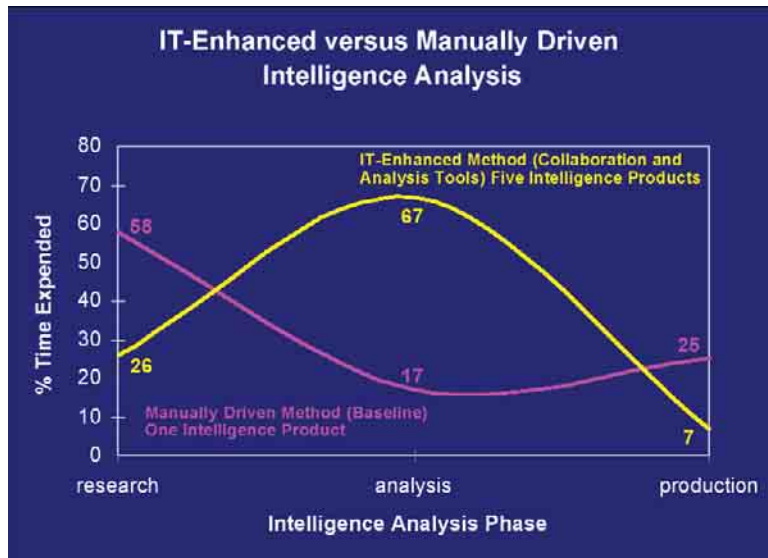
*Manage policies.* The U.S. is a nation of laws, and all activities of government are conducted within the bounds of existing laws, policies, and regulations. But the policies and regulations vary tremendously across the variety of organizations that must collaborate to counter today’s threats. Tools are needed to

allow policy and regulation to be unambiguously defined and understood at all levels, to permit teams to reconcile their differing policies and regulations into a single coherent regime, to consistently and reliably apply that regime to its operations, and to identify any deviations from policy and regulation to prevent abuses.

*Supporting process.* Teams, especially so-called ad-hocracies, need support in designing and executing strategies embodied in procedures to accomplish their goals. Tools are needed to allow them to develop and execute appropriate processes and procedures throughout their life cycles, and to ensure these processes and procedures are consistent with applicable policy.

*Amplify human intellect.* To effectively deal with the terrorist threat, it is not sufficient for well-informed experts to simply be able to communicate and share information. The counterterrorism problem is an intrinsically difficult one that is only compounded by unaided human intellect. Humans as individuals and in teams are beset by cognitive

**Figure 3. Early results are promising, showing more and better analysis in a shorter period of time by way of collaboration, modeling, and analysis tools.**



biases and limitations that have been partially responsible for some intelligence failures [3]. Analysts must be given assistance in the form of structured analytic approaches and methodologies to amplify their cognitive abilities and allow them to think together [11]. Additionally, rich toolsets are needed to allow users to understand the present, imagine the future, and generate actionable options for the decision maker.

*Reinvent policy/intelligence interface.* As U.S. Secretary of Defense Donald Rumsfeld has indicated, pol-

icymakers must not be simply passive consumers of intelligence. Instead, senior policymakers must “engage analysts, question their assumptions and methods, seek from them what they know, what they don’t know, and ask them their opinions [10].” Also, because the current policy/intelligence interface model was developed in the Cold War era during a time of information scarcity (unlike today’s information-abundant environment), some of the basic assumptions that underlie it are no longer optimal. Novel technology is needed to reinvent the interface between the worlds of policy and intelligence, allowing for intelligence that is, for example: aggressive, not necessarily cautious; intuitive, not simply fact-based; metaphor-rich, as opposed to concrete; collaborative, in addition to hierarchical; precedent-shattering, not precedent-based; and opportunistic, as well as warning-based.

*Explanation generation.* It is not enough to simply connect the dots. The fact that the dots are connected must be persuasively explained and communicated to decision- and policymakers. Traditional methods, such as briefings and reports, lack on both counts and also demand a significant amount of analysts’ time to produce (recall the bathtub curves in Figure 3). Explanation-generation technology is critical to producing traditional products, as well as making possible newer forms of intelligence products.

## Foreign Languages

**F**oreign language speech and text are indispensable sources of intelligence, but the vast majority is unexamined: Volumes are huge and growing; processing is labor intensive; and the U.S. intelligence and counterterrorism communities have too few people with suitable language skills. Because it would be impossible to find, train, or pay enough people, creating effective foreign language technology is the only feasible solution. New

and powerful foreign language technology is needed to allow English-speaking analysts to exploit and understand vastly more foreign speech and text than is possible today.

*Transcription.* Automatic transcription technology is needed to produce rich, readable transcripts of foreign news broadcasts—despite widely varying pronunciations, speaking styles, and subject matter. The two basic components of transcription are speech-to-text conversion (finding the words) and metadata extraction (pulling out more information). Inter-

Arabic	Human Translation	Machine Translation
مصر للطيران قد تعاود عدا الإرتقاء رحلاتها الى ليبيا اطن - (اف ب) 4 - القاهرة مسؤول في شركة الخطوط المصرية مصر " للطيران اليوم الثلاثاء ان شركة قد ستألف اعتبارا من يوم غد " للحدوث الإرتقاء رحلاتها الى ليبيا اثر قرار مجلس الامن الدولي تطبيق المظالمهروض على ليبيا.	<b>Egypt Air May Resume its Flights to Libya Tomorrow</b> Cairo, April 6 (AFP) - An Egypt Air official announced, on Tuesday, that Egypt Air will resume its flights to Libya as of tomorrow, Wednesday, after the UN Security Council had announced the suspension of the embargo imposed on Libya.	<b>Egyptair Has Tomorrow to Resume Its Flights to Libya</b> Cairo 4-6 (AFP) - said an official at the Egyptian Aviation Company today that the company Egyptair may resume as of tomorrow, Wednesday its flights to Libya after the International Security Council resolution to the suspension of the embargo imposed on Libya.

Table 3. Recent machine translation results of Arabic news text show great promise.

ested readers can find more information on basic speech-to-text technology in [12]. Recent achievements include word error rates of 26.3% and 19.1% at processing speeds seven and eight times slower than real-time rates on Arabic and Chinese news broadcasts. The goal is 10% or less at real-time rates.

*Translation.* A key finding in [9] was that the intelligence community was not adequately prepared to handle the challenge it faced in translating the multitude of foreign language intelligence data it collected. The challenges to contend with are numerous, including massive amounts of foreign text from an ever-growing number of foreign data sources, large unconstrained vocabularies, and numerous domains and languages with limited linguistic resources. Although the problem is far from solved, researchers are making considerable progress on the automatic translation of text. Table 3 shows the promising results obtained recently in translating an Arabic news article.

*Detection.* Advanced techniques to detect and discover the exact information a user seeks quickly and effectively and to flag new information that may be of interest are needed. Cross-language information retrieval is the current focus of the research community, with recent results showing it works approximately as well as monolingual retrieval.

*Extraction.* More sophisticated ways to extract key facts from documents are needed. Although name translation remains problematic, automatic name extraction (or tagging) works reasonably well in English, Chinese, and Arabic. Researchers are now focusing on sophisticated techniques for extracting information about entities, relationships, and events.

*Summarization.* Substantially reducing the amount of text that people must read in order to perform analysis is absolutely critical. Researchers are now working on techniques for automatic headline generation (for single documents) and for multi-document summaries (of clusters of related documents).

*Language independence.* Researchers are pursuing a wide variety of approaches that are substantially

language-independent and empirically driven. Algorithms are exploiting the continuing advances in computational power plus the large quantities of electronic speech and text now available. The ultimate goal is to create rapid, robust technology that can be ported cheaply and easily to other languages and domains.

## Pattern Analysis

**M**any terrorist activities consist of illegitimate combinations of otherwise legitimate activities. For example, acquisition of explosives, selection of a location, and financing of the acquisition by external parties are all legitimate activities in some contexts, but when combined or when performed by individuals known to be associated with terrorist groups or when the location is not, for example, a demolition/construction site but a landmark or other public building, suggest that further investigation may be warranted. While examples of terrorist activities are rare, examples of the component activities are not. Pattern analysis tools, therefore, must be able to detect instances of the component activities involving already suspicious people, places, or things and then determine if the other components are present to separate situations warranting further investigation from the majority that do not. Comprehensive overviews of some of the key technologies are available in [2, 4].

*Graphical representations.* One key idea that enables connecting the dots is representing both data and patterns as graphs. Patterns specified as graphs with nodes representing entities, such as people, places, things, and events; edges representing meaningful relationships between entities; and attribute labels amplifying the entities and their connecting links are matched to data represented in the same graphical form. These highly connected evidence and pattern graphs also play a crucial role in constraining the combinatorics of the iterative graph processing algorithms, such as directed search, matching, and hypothesis evaluation.

*Relationship extraction.* The initial evidence graph is comprised of entities and their relationships extracted from textual narratives about suspicious activities, materials, organizations, or people. Advanced techniques are needed to efficiently and accurately discover, extract, and link sparse evidence contained in large amounts of unclassified and clas-

sified data sources, such as public news broadcasts or classified intelligence reports.

*Link discovery.* Starting from known or suspected suspicious entities, patterns are used to guide a search through the evidence graph. Patterns can be obtained from intelligence analysts, subject matter experts, and intelligence or law enforcement tips, and are subject to extensive verification and testing before use. Statistical, knowledge-based, and graph-theoretic techniques are used to infer implicit links and to evaluate their significance. Search is constrained by expanding and evaluating partial matches from known starting points, rather than the alternative of considering all possible combinations. The high probability that linked entities will have similar class labels (often called autocorrelation or homophily) can be used to increase classification accuracy.

*Pattern learning.* Pattern learning techniques can induce a pattern description from a set of exemplars. Such pattern descriptions can assist an analyst to discover unknown terrorist activities in data. These patterns can then be evaluated and refined before being considered for use to detect potential terrorist activity. Pattern learning techniques are also useful to enable adaptation to changes in terrorist behavior over time.

## Conclusion

The results shown in Figure 3, based on the three core IT areas discussed in this article, represent the tip of the iceberg. Many other information technologies are important for successfully conducting the global war on terror (see Table 2). Experiments, such as the one described here, will help validate the merits and utility of these tools. Ultimately, such tools will create a seamless environment where analysts and decision- and policymakers can come together to collaborate, translate, find, link, evaluate,

share, analyze, and act on the right information faster than ever before to detect and prevent terrorist attacks against the U.S. **C**

## REFERENCES

1. Benjamin, D. and Simon, S. *The Age of Sacred Terror*. Random House, New York, 2002.
2. Goldszmidt, M. and Jensen, D. EELD recommendations report. In *Proceedings of the DARPA Workshop on Knowledge Discovery, Data Mining, and Machine Learning (KDD-ML)*, Arlington, VA, 1998.
3. Heuer, R. *Psychology of Intelligence Analysis*. Center for the Study of Intelligence, Central Intelligence Agency, 1999.
4. Jensen, D. and Goldberg, H., Eds. *Artificial Intelligence and Link Analysis: Papers from the 1998 AAAI Fall Symposium*, AAAI Press, Menlo Park, CA, 1998.
5. Jonietz, E. Total information overload. *MIT Technology Review* (Aug. 2003).
6. Lazaroff, M. and Sickels, S. *Human Augmentation of Reasoning Through Patterning (HARP)*. DARPA Genoa II PI Meeting, Austin, TX, May 2003.
7. *National Strategy for Combating Terrorism*. Submitted by the White House, Feb. 2003.
8. *Report to Congress Regarding the Terrorism Information Awareness (TIA) Program*. Submitted by the Secretary of Defense, Director of Central Intelligence and Attorney General, May 2003.
9. *Report of the Joint Inquiry into the Terrorist Attacks of September 11, 2001*. Submitted by the House Permanent Select Committee on Intelligence (HPSCI) and the Senate Select Committee on Intelligence (SSCI), July 2003.
10. Shanker, T. For military intelligence, a new favorite commando. *New York Times* (Apr. 11, 2003).
11. Schum, D. *The Evidential Foundations of Probabilistic Reasoning*. Wiley, New York, 1994.
12. Young, S. *Large Vocabulary Continuous Speech Recognition: A Review*. Technical Report, Cambridge University Engineering Department, Cambridge, U.K., 1996.

---

**ROBERT POPP** (rpopp@darpa.mil) is a special assistant to the DARPA Director for Strategic Matters and was formerly the deputy director of the Information Awareness Office.

**THOMAS ARMOUR** (tarmour@darpa.mil) is a program manager with DARPA's Information Processing Technology Office and was formerly a program manager with the Information Awareness Office.

**TED SENATOR** (tsenator@darpa.mil) is a program manager with DARPA's Information Processing Technology Office and was formerly a program manager with the Information Awareness Office.

**KRISTEN NUMRYCH** (knumrych@darpa.mil) is an assistant director for Program Management at DARPA.

---