

## **Special Notice 13-SN-0002**

### **Special Program Announcement for the Office of Naval Research Research Opportunity: Text Analytics for Data to Decisions (D2D)**

#### **I. INTRODUCTION:**

On behalf of the Office of the Secretary of Defense (OSD), Office of the Assistant Secretary of Defense, Research and Engineering (ASD (R&E)), this announcement describes a research thrust, *Text Analytics*, to support the OSD Data to Decisions (D2D) research program. The Office of Naval Research (ONR) is releasing this notice on behalf of OSD to be launched under the ONRBAA13-001, Long Range Broad Agency Announcement for Navy and Marine Corps Science and Technology which can be found at <http://www.onr.navy.mil/Contracts-Grants/Funding-Opportunities/Broad-Agency-Announcements.aspx>. The research opportunity described in this announcement specifically falls under the following sections of ONRBAA 13-001.

The research opportunity described in this announcement specifically falls under the “Research Opportunity Description”, Command, Control, Communications, Computers, Intelligence, Surveillance and Reconnaissance (C4ISR) (Code 31), (a) Applied Computational Analysis, (b) Command and Control, and (c) Intelligent and Autonomous sub-sections. The submission of proposals, their evaluation, and the placement of research grants and contracts will be carried out as described in that Broad Agency Announcement.

#### **II. TOPIC DESCRIPTION:**

The proposed topic will explore and exploit advances in the analysis of unstructured text. The program will pursue research in the areas of contextual understanding, event prediction, as well as machine translation and processing.

#### **BACKGROUND**

The Office of Naval Research (ONR) is seeking innovative solutions for the Data to Decisions (D2D) research program in the area of text analytics. The D2D program is an Office of the Secretary of Defense (OSD), Office of the Assistant Secretary of Defense, Research and Engineering (ASD (R&E)) multi-year, funded, and directed program that is competitively awarded. The D2D program provides an opportunity to conduct objective assessments of analytics for rapid maturation to support understanding and develop improvements in mission effectiveness in time constrained and data intensive environments. The D2D program is comprised of an applied research component (6.2 program element) and a demonstration and evaluation component (6.3 program element). This Special Notice (13-SN-0003) will focus specifically on text analytics techniques and methods that improve mission effectiveness in a time constrained environment where increasing amounts of text-based data and information are present.

Military decision making requires supporting data from a variety of sources and in a wide range of formats to achieve success in its many missions. Text and language play an important part in many military missions by providing the key data required for understanding cultures, attitudes, events, and relationships that serve as the basis of many of the missions conducted around the world. With advances and increased use of Internet and mobile communications, text information is available in unprecedented amounts and formats, presenting a unique opportunity to gain understanding through text analytics. In the context of the D2D research program text analytics refers to identification of a set of linguistic, statistical, and machine learning techniques that model and structure the information content of textual sources for exploratory data analysis, research, and investigation. Text analytics involves information retrieval, lexical analysis to study word frequency, data mining techniques including link and association analysis, visualization, and predictive analytics. This solicitation for research in text analytics focuses on three primary thrust areas 1) contextual understanding, 2) event prediction, and 3) machine translation and processing.

The Department of Defense (DoD) recognizes the potential for text analytics to play a vital role in future capabilities that inform timely and accurate situational awareness in time-constrained, uncertain, and complex environments by developing capability in contextual understanding, event prediction, and machine translation and processing. The D2D program will achieve these capabilities through a spiral development process that begins with a scenario that exhibits several text analytics challenges that will be described in the following section. The D2D program will build and evolve an end-to-end demonstration prototype for each spiral that will enable us to 1) analyze performance enabling us to gain knowledge to identify critical science and technology needs, 2) provide data that can be distributed in support of science and technology development, 3) provide demonstrations of key technologies enabling assessment of technology maturity, 4) provide system solutions, system improvements, and alternative concepts of operations, and 5) develop a predictive capability for operations.

Our overarching end state scenario supports a group of mobile operators moving through the battlespace with support from a mixture of fixed and mobile sensors, intelligence data, and with connectivity to an in theater support infrastructure with minimal resources including personnel, computation, and bandwidth. Note that we are using the scenario to identify and address technology issues and solution recommendations.

Currently, the United States faces a significantly different threat than existed in the past. Modern threats are asymmetric and often do not represent nation states. The “battlefield” is wider and the opponent blends into the background culture and employs non-traditional tactics, techniques, and procedures (TTPs). Threat activities occur on a variety of time scales and the threat activity signatures are not well known. Developing capabilities to rapidly identify the signature of an opponent’s TTPs that represent a threat is essential.

To illustrate the challenges faced by the current threat, the following is a case scenario based upon the needs of the recently formed Africa Command (AFRICOM). Physically, Africa is approximately three times the size of the US with three times the population. Climatologically, much of Africa has natural fluctuations in weather that produce long periods of drought leading to greatly reduced agricultural products and food shortages. The environmental conditions,

combined with ineffective political/economic infrastructure, produce severe human conditions such as sustained food shortages, famine, illness, and migration. Each of these systematically places great strain on the political infrastructure and creates a downward spiral in the ability of leaders to address human needs. With the inability of the central governments to cope, feudal conflicts arise among clans and villages for control of diminished resources causing mass human migrations and conflicts within nations.

Conventional Intelligence, Surveillance, and Reconnaissance (ISR) technologies require additional information to help policy/decision makers respond to human needs and predict conflict in this area. Given the size of the AFRICOM theater, its geographic diversity, as well as limitations with respect to our ability to employ conventional sensors, it is unlikely that AFRICOM will be able to maintain persistent surveillance over important regions. In order to maintain situational awareness, a different approach to sensing and exploitation, that reasons over all of the available data must be employed. Much of this non-traditional ISR data is expected to be in the form of structured and unstructured text that is currently being produced.

Non-nation state actors and insurgents exploit the Internet as a backbone for their command and control that minimizes the visible footprint of the interactions among individuals. In addition, the amount of information contained within the web is tremendous; government press reports, official releases, and interviews; non-government organizations such as aid groups' needs and action statements for specific crises; and chat rooms populated by neutral observers and active participants. These same sources can support also support counterinsurgency and counterterrorism.

The ongoing D2D program depends heavily on collaboration. The D2D program has benefitted greatly from working with a wide variety of research partners from across Defense Science and Technology (S&T), the intelligence community, and academia. It is anticipated that work solicited under this announcement will build on and augment previous work and leverage lessons learned, community knowledge, and research relationships that have been established. D2D expects that the performers within the program will continue to collaborate and work directly with Defense S&T labs in order to meet and exceed program objectives.

## **OBJECTIVES**

The Office of Naval Research (ONR) is interested in receiving proposals on contextual understanding, event prediction, as well as machine translation and processing.

Today's warfighter has access to text-based information from a wider range and greater number of sources than ever before. The influx of information can potentially improve warfighters' situation understanding and decision-making. However, it is clear that there has been, and will be no, increase in the number of warfighters to process and interpret the growing volume of available data. The practical implication of this is that the DoD has access to more data than it can process to achieve actionable information in support of diverse military information needs. It is imperative that we create, harvest and exploit technologies that will help realize the potential for improved decision-making without imposing a need for increased warfighter numbers or their workload. On behalf of the Office of the Secretary of Defense (OSD), Office of the Assistant Secretary of Defense, Research and Engineering (ASD (R&E)), the Office of Naval Research

(ONR)) is interested in receiving full proposals for the Data to Decision Program. The program has three (3) primary thrust areas, (1) contextual understanding, 2) event prediction, and 3) machine translation and processing.

Together these thrusts areas seek to develop new technological capabilities that support military operations. The three (3) thrust areas are described below:

### **Thrust Area #1- Contextual Understanding**

The 2012 National Security Strategy has indicated that “for the foreseeable future, the United States will continue to take an active approach to countering [threats] by monitoring the activities of non-state threats worldwide.” With ever increasing decentralization of decision-making brought about by ubiquitous electronic communication, there is a need to identify threats in complex, uncertain, contradictory and incomplete large data sets available in the open source environment. Irregular warfare, non-state terrorism movements, and uncertain environmental patterns that trigger major weather disasters are examples of events that require military response. In responding, decision makers will use text analytics to develop the necessary contextual understanding of the region and key elements. Strategies for achieving contextual understanding can include observational data, a priori knowledge models, and inductive knowledge. Contextual understanding is generally achieved through a combination of human and computer processing techniques that take advantage of a person’s cognitive ability to fuse and assimilate multiple sources and types of information for new insights. Correlation and aggregation of open source data, such as agriculture, weather, terrain, demographics, economics, social patterns etc. is nontrivial but vital to effective military response.

This research thrust area seeks innovative approaches to the following aspects of contextual information 1) the discovery of specific events that are planned or may have occurred, 2) stated values and beliefs that motivate behaviors of interest, 3) discovery of topics and concepts developed in a shared community, 4) analysis of semantic relationships existing in a community, 5) strength of relationships, 6) community structure and clusters of social networks, and 7) semantic analysis and trending of emotional support expressed toward topics or persons.

### **Thrust Area #2- Event Prediction**

Intelligence analysts need the ability to rapidly monitor and analyze event information in large volumes of unstructured textual data, such as news articles or Human Intelligence (HUMINT) reports, in order to achieve and maintain persistent Situational Awareness (SA). For example, the ability to stay apprised of events that have already occurred, as well as events that is threatened or planned, would be a valuable contribution to an analyst’s SA; however, the amount of unstructured textual data available is well beyond what can be manually read and processed in the time available. A capability enabling analysts to rapidly extract event information from large volumes of unstructured text and store it in a structured form, such as a database, is needed to improve an analysts’ ability to maintain persistent SA.

The goal of this thrust is to advance the state-of-the art for extracting events with their attributes of modality, polarity, genericity, and tense from large volumes of unstructured text. Modality of an event indicates if the event was a real occurrence. Examples of event modality include asserted, i.e. “The bomb exploded on Sunday;” believed, i.e. “It is rumored he will be

sentenced;” hypothetical, i.e. “If he were arrested, he would be convicted of murder;” and threatened, i.e. “He threatened to attack the country.” Event polarity indicates whether the event actually occurred. For example, “The city was not attacked” is an event with negative polarity, and “The attack occurred on Sunday” is an event with positive polarity. Genericity indicates whether an event is specific, i.e. “The city was attacked on Saturday”, or generic, i.e. “They specialize in transporting weapons.” Tense indicates whether an event occurred in the past, is occurring in the present, or will occur in the future. Secondary challenges include, but may not be limited to, rapid customization to different sources/styles/formats of textual data, and rapid customization to various domains. While addressing other technology gaps that would contribute to the capability would be useful, it is of lower priority to the program since it should not happen at the expense of addressing the primary research challenge of extracting event attributes.

DOD text analytics will often be focused on social groups who have an interest in hiding behavior, such as terrorist networks. In such conditions, innovative methods are needed to identify proxy features of a network that may aid discovery goals to uncover potential events of interest. Temporal trends are one such category. Examples may include factors such as frequency of contacts between nodes or clusters, inter-contact time, recurrent contacts, time order of contacts along a path, and delay path of information diffusion. Methods to extract, characterize, and monitor social networks dynamically over time is a research challenge of interest that may support event prediction.

Scalability and predictability have been perennial problems in certain types of text-based analysis, such as social network analysis (SNA). As networks increase exponentially in size and complexity, it is harder to use graphical methods to represent, monitor, and understand network behavior. The representational graphs grow to unmanageable size, contain complex relationships among nodes, and often contain several varieties of nodes. Two promising approaches are being explored by the SNA community, visual analytics and semantic analysis. Visual analytic methods supported by ontology have been shown to reduce the visual complexity of these graphs to enable users to identify important structural and semantic aspects of networks. Research is needed to identify key actors and supported relationships, detect the presence of bridging nodes that can uncover hidden sub-networks, and determine the flow of resources (information, money, influence) within the social network.

This research thrust seeks innovative approaches to the following aspects of event prediction 1) identify proxy features of a network, 2) extract temporal trends, i.e. frequency of contacts between nodes or clusters, inter-contact time, recurrent contacts, time order of contacts along a path, and delay path of information diffusion, 3) extract, characterize, and monitor social networks dynamically over time, 4) evolve visual analytics and semantic analysis at scale, 5) identify key actors and supported relationships, and 6) detect the presence of bridging nodes that can uncover hidden sub-networks, and determine the flow of resources (information, money, influence) within the social network.

### **Thrust Area #3- Machine translation and processing**

Unfortunately, while there have been notable advances in information retrieval (e.g., intelligent, adaptive and ontology-based search engines), data mining, and (to a limited extent) development of cognitive aids and decision support tools, progress on the “ingestion” or understanding of information has not advanced as rapidly as for collection. This is due, in no small part, to the huge increase in unstructured information available to the warfighter. While text can be

extremely valuable, it is not readily amenable to automated processing. Instead text must, in large part, be handled, assessed and interpreted individually by humans. As a result, the sheer magnitude of these resources can overwhelm the very mission they are collected to support.

Many areas of the world where future military action may be required are rich in language or dialect diversity. To fully engage local populations and respond to humanitarian needs, language translation will become critical to text analytics efforts. Strategies that lead to computationally efficient algorithms are needed to develop and improve technologies for machine translation and processing, information extraction, and automated summarization. Also relevant, are the methods and algorithms to develop and improve technologies that “import” physical sources into electronic form such as optical character recognition (OCR) and speech recognition as input to machine translation and processing, information extraction, and automated summarization. Development of language data in support of building these technologies and development of metrics to evaluate underlying software algorithms are also needed.

Research in the areas of linguistics, natural language processing, mathematics, statistics, computational data analysis and visualization, computational sciences and computer science are of interest. In addition to the application of research methods and approaches, it is important to evaluate the impact of these efforts areas with regards to the way they change how data is collected, analyzed and assessed to meet a prescribed time for operational necessity and efficiency. It is of value to use open standards to reduce costs. This research thrust seeks innovative approaches to the following aspects of machine translation and processing 1) intelligent, adaptive and ontology-based search engines, 2) improved data mining, 3) improved cognitive aids and decision support tools, 4) “ingestion” or understanding of information at scale, 5) improved information extraction, 6) improved automated summarization.

### **III. FULL PROPOSAL SUBMISSION AND AWARD INFORMATION**

Full proposals should be submitted under **ONRBAA13-001** by 15 January 2013. Full Proposals received after that date will be considered as time and availability of funding permit.

ONR anticipates that both grants and contracts will be issued for this effort. Full proposals for contracts should be submitted in accordance with the instructions in ONRBAA 13-001, Section IV, Application and Submission Information, item 2.b, Full Proposals.

Full proposals for grants should be submitted in accordance with the instructions at Section IV, Application and Submission Information, item 5, “Submission of Grant Proposals” through Grants.gov. All full proposals for grants must be submitted through [www.grants.gov](http://www.grants.gov). The following information must be completed as follows in the SF 424 to ensure that the application is directed to the correct individual for review: Block 4a, Federal Identifier: Enter N00014; Block 4b, Agency Routing Number, Enter the three (3) digit Program Office Code (311) and the Program Officer’s name, last name first, in brackets (Schwartz Carey). All attachments to the application should also include this identifier to ensure the proposal and its attachments are received by the appropriate Program Office.

ONR plans to fund five (5) to ten (10) individual awards with a value of approximately \$500,000 per year, using Applied Research (6.2/6.3) funds. However, lower and higher cost proposals will be considered.

The period of performance for projects may be from one (1) to three (3) years.

Although ONR expects the above described program plan to be executed, ONR reserve the right to make changes.

Funding decisions should be made by 15 February 2013. Projects will have an estimated contract award on or about 01 July 2013. Projects will have an estimated grant award on or about 01 April 2013.

#### **IV. POINTS OF CONTACT**

In addition to the points of contact listed in ONR Long Range Broad Agency Announcement, ONRBAA13-0001 the specific points of contact for this announcement are listed below:

##### Technical Points of Contact:

Office of Naval Research  
875 North Randolph Street, Suite 1179  
ATTN: Dr. Carey Schwartz, Program Officer, Code 311  
Email Address: [carey.schwartz@navy.mil](mailto:carey.schwartz@navy.mil)

##### Business Point of Contact(s):

##### Primary Point of Contact

Office of Naval Research  
875 North Randolph Street, Suite W1275A  
ATTN: Kenesha Y. Hargrave, Contract Specialist, Code BD251  
Email Address: [kenesha.y.hargrave@navy.mil](mailto:kenesha.y.hargrave@navy.mil)

##### Secondary Point of Contact

Office of Naval Research  
875 North Randolph Street, Suite 1279  
Vera M. Carroll, Acquisition Branch Head, Code BD 251  
Email Address: [vera.carroll@navy.mil](mailto:vera.carroll@navy.mil)

#### **V. SUBMISSION OF QUESTIONS**

Any questions regarding this announcement must be provided to the Technical Points of Contact and/or the Business Point of Contact listed above. All questions shall be submitted in writing by electronic mail.

Answers to questions submitted in response to this Special Notice will be addressed in the form of an Amendment and will be posted to the following web pages:

Special Notice 13-SN-0002

- Federal Business Opportunities (FEDBIZOPPS) Webpage – <https://www.fbo.gov/>
- Grants.gov Webpage – <http://www.grants.gov/>
- ONR Special Notice Webpage - <http://www.onr.navy.mil/Contracts-Grants/Funding-Opportunities/Special-Notices.aspx>

Questions regarding **Full Proposals** should be submitted NLT (2) two weeks before the dates recommended for receipt of Full Proposals. Questions after this date may not be answered.