

Department of the Navy SBIR/STTR Transition Program

DISTRIBUTION STATEMENT A. Approved for public release. Distribution is unlimited.

ONR Approval #43-3252-17

Topic # N102-176

RDF Entity and Association Disambiguation (READ) Enhancement

DECISIVE ANALYTICS Corporation

WHO

SYSCOM: ONR

Sponsoring Program:

Expeditionary Maneuver Warfare and Combating Terrorism Department - ISR Programs

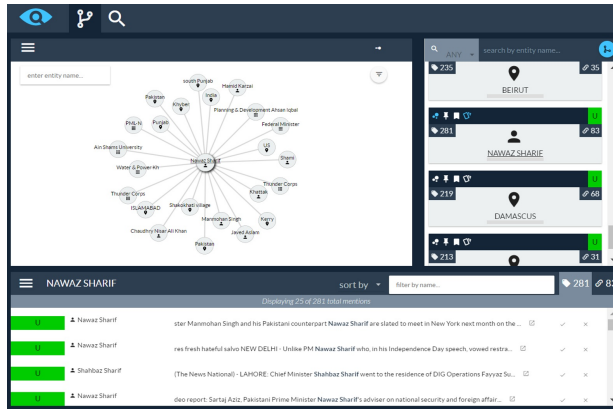
Transition Target: Distributed Common Ground System-Navy (DCGS-N)

TPOC:

Mr. Martin Kruger
martin.kruger1@navy.mil

Other transition opportunities: DCGS-A, AF DCGS, Securities and Exchange Commission, Special Operations Command, FBI

Notes: Image is screenshot for the technology application



Copyright 2017 Decisive Analytics Corporation (DAC)

WHAT

Operational Need and Improvement: It is difficult if not impossible for Warfighters to make sense out of today's ever-growing sea of intelligence data. To have a crisp understanding of all of this data, Warfighters must often go through tedious rounds of analysis to identify the organizations, products, people, and other entities identified in the data. Furthermore, this data comes from multiple disparate data sources. Intelligence sources, open source data, and other public/private data stores each contribute information to the overall intelligence picture. Each data source provides different types of information, in different formats, with varying levels of quality and detail. The automation of the "grunt work" of data acquisition, manual extraction of the relevant entities, and normalization and structuring of the relevant data enables the Warfighter to better put their cognitive skills towards the critical analytical problems.

Specifications Required: Advances have been made with regard to our ability to express large disparate unstructured data sources (e.g. text, images, audio) as connected entity graphs. Due to entity and association uncertainty, current implementations become filled with redundant statements, preventing the expression of a large data corpus as one connected graph. The objective of this topic is to develop algorithms and techniques for level one fusion of association statements regarding entities.

Technology Developed: Our suite of Natural Language Processing (NLP) analytics provides the following:

- Statistical Topic Modeling (STM) provides analysts with an categorized overview of large data sets.
- Semantic Role Labeling (SRL) automatically identifies relationships between mentioned entities.
- Entity Disambiguation and Resolution providing analysts with collapsed entity lists which are easily vetted.

Warfighter Value: The capability allows Warfighters to more accurately and quickly extract entities and understand relationships extracted from unstructured data repositories faster and more accurately than ever before.

WHEN

Contract Number: N00014-15-C-0064 **Ending on:** September 30, 2017

| Milestone | Risk Level | Measure of Success | Ending TRL | Date |
|--|------------|--|------------|---------------|
| Operational Testing at Fort Gordon | Low | Successful user testing by intel analysts at Fort Gordon | 6 | April 2015 |
| Exercise Testing within Trident Warrior 16 | Low | Transition to DCGS-N in support of Trident Warrior | 6 | July 2016 |
| Expansion maritime data sources and entity types | Med | Demo of all DCGS-N data sources and entity types | 7 | October 2017 |
| Operational test within DCGS-N | Med | Completion of all testing and accreditation requirements | 8 | December 2017 |
| Deployment within Operational DCGS-N Inc 2 | Med | Full capability within the deployed DCGS-N System | 9 | June 2018 |

HOW

Projected Business Model: The business model for this effort is a mix of license fees and custom development for various user communities. Through our experience, we know that this capability requires some custom development for different domains to produce effective results.

Company Objectives: Below is a list of transition partners that can benefit from this technology. DCGS-N / DCGS-A / AF DCGS: Entity network graphs are a powerful and flexible tool to help analysts understand the human terrain of a region or organization of interest. DAC's automated capability results in a large reduction in the labor required to perform network analysis. Inclusion these developed methods of processing large scale data repositories of documents will be important to these customers.

The Intelligence Community including the CIA, DIA, NSA, NASIC, MSIC, NGA, and others have requirements for analysts to automatically make senses out of large amounts of unstructured text data. The entity disambiguation capability is a analytic that identifies different mentions of entities and collapses them into one. Furthermore, the capability enables analysts to see relationships with other entities.

Potential Commercial Applications: Financial Analysts: Investment management and other financial market analysts have a need to analyze complex data sets about companies and industries. Many of these problems mimic the problems facing intelligence analysts: analysis must draw from huge amounts of data, a variety of expertise is required, and understanding and managing the data is crucial to success. Through a self-funded marketing effort, DAC has past performance and relationships with banks, hedge funds, and investment advisers in the financial markets and has validated this requirement.

Contact: James J. Nolan, Ph.D., Vice President, Analytic Technologies
jim.nolan@dac.us 703-414-5002