Conclusion
Michelle Albert

The IDATA Team

Recently, the open-source community and the commercial marketplace have exploded with data triage and discovery capabilities. Most of these tools, however, are not designed for DoD or federal government use. Commercially available tools rarely provide data pre-processing or post-processing. They have no knowledge of the context in which DoD and other government agencies operate or where the data come from.

The IDATA capability, on the other hand, uses cutting-edge open-source libraries and a modular approach that enables it to be customized to suit a sponsor's needs and work flow. The other major part of IDATA's value is the IDA analysts, subject matter experts, and researchers who work with the capability. They come from a variety of backgrounds – academia, industry, military – and have diverse experience and education. They include the expected computer scientists and mathematicians, but also include the expertise of English majors, political scientists, economists, and physicists. Just as non-correlated algorithms can compensate for each other's weaknesses when used together, a well-rounded team can address myriad disparate tasks and issues.

This approach and the IDATA team's results have been recognized by IDA and the scientific community at large. The team has published multiple papers and has received the IDA Welch Award, IDA's W.Y. Smith Award, and awards from the Association of Enterprise Information (AFEI, an affiliate of the National Defense Industrial Association) for Automated Information Triage for Rapid Decision Making, in 2015, and State Cyber Operations Framework, in 2016.

Future Opportunities for the IDATA Capability

The IDATA team continues to grow, evolve, and incorporate new tools and techniques to stay on the cutting edge of natural language processing and

2016 Larry D. Welch Award for Best External Publication winners:
Andrew Wan, Arun Maiya, and Dale Visser (left to right).

Miranda Seitz-McLeese accepts the W.Y. Smith Award.

The IDATA Team receives the AFEI Award.

From left to right: Andrew Wan, Edna Jordan, Anna Vasilyeva, Thomas Barth, Andrew Ferguson, Tristian Weir, Dave Chesebrough (AFEI), Laura Odell, Miranda Seitz-McLeese, Cameron DePuy, and Corbin Fauntleroy

provide sponsors with tailored, non-biased, and actionable results.

The most fertile ground for the IDATA capability lies in areas with large quantities of public data, especially the growing amount of information available online that is released both by private individuals on social media and by the U.S. Government as part of the Open Data Policy.[1]

Sentiment analysis could provide a new way for sponsors to look at this available data. Tracking connections via LinkedIn, geotags on Facebook, or hashtag use on Twitter may provide sponsors with valuable insight. Recruiters, for example, could use these data to more easily find and target potential recruits with desired skill sets.

Also, the Office of Legislative Affairs could track Twitter posts from a particular location to help anticipate the concerns of a particular Congress member's constituents. There are algorithms that would work with efforts of this kind, although they have not been fully employed in a national security context. The IDATA capability's modular nature makes it fully compatible with them.

Another avenue for potential growth comes from the discovery that existing entity-extraction algorithms, which attempt to extract the names of people and organizations from natural language documents, do not adapt well to the national security context. Training and modifying these algorithms would make the task of sifting through and sorting a large amount of government documents less daunting. For example, users could filter for documents containing a particular office's name or filter for memoranda signed by a particular official. Contracts could also be filtered by company, which would potentially enable DoD to consolidate purchases and eliminate duplication.

These are just a few examples. Many more may become available as research continues in this area and analysts apply new techniques.

Summary

By covering a wide range of topic areas, IDA has proven that the capability underlying the text analytics concept can be rapidly customized to produce results in a relatively short time. Once sponsors have the opportunity to use automated information triage to solve a problem, they quickly see the benefit and bring in other problems to solve. This approach has also given sponsors in DoD and other federal agencies the opportunity to reduce costs due to the IDATA capability's ability to minimize the time needed to find, search through, and analyze information across a variety of documents and file types. It has also enabled DoD and other federal agencies to find and process existing information, eliminating duplicative activities that result from organizations' inability to find or manipulate data from earlier efforts.

---

[1] Executive Order (EO) 13642, "Making Open and Machine Readable the New Default for Government Information," states that making information accessible and usable can promote job growth, innovation, and scientific discovery. It establishes a default in which data are released to the public whenever possible and legally permissible.

Reference

The White House. Executive Order 13642, *Making Open and Machine Readable the New Default for Government Information.* May 9, 2013.

Postscript

The articles in this *IDA Research Notes* describe examples of how the IDATA capability has been used to solve real-world problems. Those who wish to know more about the research behind IDATA should review the following papers:

- *Mining Measured Information from Text* (published at SIGIR '15)
  (https://doi.org/10.1145/2766462.2767789)

- *A Framework for Comparing Groups of Documents* (published at EMNLP '15)
  (https://www.ida.org/idamedia/Corporate/Files/Publications/IDA_Documents/ITSD/2015/D-5543.pdf)

- *Topic Similarity Networks: Visual Analytics for Large Document Sets* (published at IEEE BigData '14)
  (https://ieeexplore.ieee.org/document/7004253/)

- *Exploratory Analysis of Highly Heterogeneous Document Collections* (published at KDD '13)
  (https://doi.org/10.1145/2487575.2488195)

- *Supervised Learning in the Wild: Text Classification for Critical Technologies* (published at IEEE MILCOM '12)
  (https://ieeexplore.ieee.org/document/6415660/)

Contributors

Ms. Michelle Albert is a Research Associate in IDA's Information Technology and Systems Division. She holds a Master of Arts in journalism from the University of Missouri.

Ms. Katharine Burton is an Adjunct Research Staff Member in IDA's Information Technology and Systems Division. She holds a Master of Arts in administrative science from George Washington University and a Master of Science in national security strategies from the National Defense University.

Dr. Forrest Frank is a Consultant in IDA's Information Technology and Systems Division. He holds a Doctor of Philosophy in political science from Stanford University.

Dr. Arun Maiya is a Research Staff Member in IDA's Information Technology and Systems Division. He holds a Doctor of Philosophy in computer science from the University of Illinois at Chicago.

Dr. Daniel Nakada is a former Research Staff Member in IDA's System Evaluation Division. He holds a Doctor of Philosophy in electrical engineering from the Massachusetts Institute of Technology.

Mr. James O'Connor is a former Research Assistant in IDA's Information Technology and Systems Division. He holds a Master of Science in mechanical engineering from Worcester Polytechnic Institute.

Ms. Laura Odell is an Assistant Director in IDA's Information Technology and Systems Division. She holds a Bachelor of Science in electrical engineering from the University of Minnesota, a Master of Science in contract acquisition management, and a Master of Science in management from the Florida Institute of Technology.

Dr. Robert Rolfe is an Adjunct Research Staff Member in IDA's Information Technology and Systems Division. He holds a Doctor of Philosophy in physics from the University of California, Los Angeles.

Ms. Miranda Seitz-McLeese is a Research Associate in IDA's Information Technology and Systems Division. She holds a Master of Science in applied mathematics and statistics from Georgetown University.

Ms. Thi Uyen Tran is a Research Staff Member in IDA's System Evaluation Division. She holds a Bachelor of Science in electrical engineering from George Mason University.

Ms. Anna Vasilyeva is a former Research Associate in IDA's Information Technology and Systems Division. She holds a Master of Science in aeronautics and astronautics from the Massachusetts Institute of Technology and a Master of Science in technology policy from the Judge Business School, University of Cambridge.

Dr. Dale Visser is a Research Staff Member in IDA's Information Technology and Systems Division. He holds a Doctor of Philosophy in physics from Yale University.

Dr. Andrew Wan is an Adjunct Research Staff Member in IDA's Information Technology and Systems Division. He holds a Doctor of Philosophy in computer science from Columbia University.

Past Issues

Challenges in Cyberspace

- Cyberspace – The Fifth and Dominant Operational Domain
- Transitioning to Secure Web-Based Standards
- Information Assurance Assessments for Fielded Systems During Combat Command Exercises
- Supplier-Supply Chain Risk Management
- Internet-Derived Targeting: Trends and Technology Forecasting
- Training the DoD Cybersecurity Workforce

Multidisciplinary Research for Securing the Homeland – IDA and DHS: Beyond 15

- Countering Terrorism One Technology at a Time
- Does Imposing Consequences Deter Attempted Illegal Entry into the United States?
- Improving Shared Understanding of National Security and Emergency Preparedness Communications
- Foreign Counter-Unmanned Aerial Systems: Developments in the International Arms Market
- Operationalizing Cyber Security Risk Assessments for the Dams Sector
- Understanding the Juvenile Migrant Surge from Central America
- Implementing a Roadmap for Critical Infrastructure Security and Resilience
- Baselining: Application of Qualitative Methodology for Quantitative Assessment
- Analysis, Analysis Practices, and Implications for Modeling and Simulation
- Test and Evaluation for Reliability

Acquisition, Part 1: Starting Viable Programs

- Defining Acquisition Trade Space Through "DERIVE"
- Supporting Acquisition Decisions in Air Mobility
- Assessing Reliability with Limited Flight Testing
- Promise and Limitations of Software Defined Radios
- Implications of Contractor Working Capital on Contract Pricing and Financing
- The Mechanisms and Value of Competition
- Early Management of Acquisition Programs

Acquisition, Part 2: Executing and Managing Programs

- Cost Growth, Acquisition Policy, and Budget Climate
- Improving Predictive Value of Poor Performance

- Root Cause Analysis of VTUAV Fire Scout's Nunn-McCurdy Breach
- Evaluating Solid Rocket Motor Industrial Base Consolidation Scenarios
- Managing Supply Chain Cyber Risks To DoD Systems and Networks
- Looking Back at PortOpt: An Acquisition Portfolio Optimization Tool
- Predicting the Effect of Schedule on Cost
- Recent Developments in the Joint Strike Fighter Durability Testing

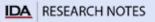## Test and Evaluation: Statistical Methods for Better System Assessments

- Assessing Submarine Sonar Performance Using Statistically Designed Tests
- Applying Advanced Statistical Analysis to Helicopter Missile Targeting Systems
- Tackling Complex Problems: IDA's Analyses of the AN/TPQ-53 Counterfire Radar
- Improving Reliability Estimates with Bayesian Hierarchical Models
- Managing Risks: Statistically Principled Approaches to Combat Helmet Testing
- Validating the Probability of Raid Annihilation Test Bed Using a Statistical Approach

Technological Innovation for National Security

- Acquisition in a Global Technology Environment
- Lessons on Defense R&D Management
- Commercial Industry R&D Best Practices
- Strengthening Department of Defense Laboratories
- Policies of Federal Security Laboratories
- The Civilian Science and Engineering Workforce in Defense Laboratories
- Technology Transfer: DoD Practices

Security in Africa

- Trends in Africa Provide Reasons for Optimism
- China's Soft Power Strategy in Africa
- Sudan on a Precipice
- A New Threat: Radicalized Somali-American Youth
- Chinese Arms Sales to Africa