NPS-AM-11-C8P13R01-051

EXCERPT FROM THE
PROCEEDINGS
OF THE
EIGHTH ANNUAL ACQUISITION RESEARCH SYMPOSIUM WEDNESDAY SESSIONS VOLUME I

A Web Service Implementation for Large-Scale Automation, Visualization and Real-Time Program Awareness via Lexical Link Analysis

Ying Zhao, Shelley Gallup, and Douglas MacKinnon, NPS

Published: 30 April 2011

Prepared for the Naval Postgraduate School, Monterey, California 93943

ACQUISITION RESEARCH PROGRAM
GRADUATE SCHOOL OF BUSINESS & PUBLIC POLICY
NAVAL POSTGRADUATE SCHOOL

The research presented at the symposium was supported by the Acquisition Chair of the Graduate School of Business & Public Policy at the Naval Postgraduate School.

To request Defense Acquisition Research or to become a research sponsor, please contact:

NPS Acquisition Research Program
Attn: James B. Greene, RADM, USN, (Ret.)
Acquisition Chair
Graduate School of Business and Public Policy
Naval Postgraduate School
555 Dyer Road, Room 332
Monterey, CA 93943-5103
Tel: (831) 656-2092
Fax: (831) 656-2253
E-mail: jbgreene@nps.edu

Copies of the Acquisition Sponsored Research Reports may be printed from our website www.acquisitionresearch.net

Preface & Acknowledgements

During his internship with the Graduate School of Business & Public Policy in June 2010, U.S. Air Force Academy Cadet Chase Lane surveyed the activities of the Naval Postgraduate School's Acquisition Research Program in its first seven years. The sheer volume of research products—almost 600 published papers (e.g., technical reports, journal articles, theses)—indicates the extent to which the depth and breadth of acquisition research has increased during these years. Over 300 authors contributed to these works, which means that the pool of those who have had significant intellectual engagement with acquisition issues has increased substantially. The broad range of research topics includes acquisition reform, defense industry, fielding, contracting, interoperability, organizational behavior, risk management, cost estimating, and many others. Approaches range from conceptual and exploratory studies to develop propositions about various aspects of acquisition, to applied and statistical analyses to test specific hypotheses. Methodologies include case studies, modeling, surveys, and experiments. On the whole, such findings make us both grateful for the ARP's progress to date, and hopeful that this progress in research will lead to substantive improvements in the DoD's acquisition outcomes.

As pragmatists, we of course recognize that such change can only occur to the extent that the potential knowledge wrapped up in these products is put to use and tested to determine its value. We take seriously the pernicious effects of the so-called "theory–practice" gap, which would separate the acquisition scholar from the acquisition practitioner, and relegate the scholar's work to mere academic "shelfware." Some design features of our program that we believe help avoid these effects include the following: connecting researchers with practitioners on specific projects; requiring researchers to brief sponsors on project findings as a condition of funding award; "pushing" potentially high-impact research reports (e.g., via overnight shipping) to selected practitioners and policy-makers; and most notably, sponsoring this symposium, which we craft intentionally as an opportunity for fruitful, lasting connections between scholars and practitioners.

A former Defense Acquisition Executive, responding to a comment that academic research was not generally useful in acquisition practice, opined, "That's not their [the academics'] problem—it's ours [the practitioners']. They can only perform research; it's up to us to use it." While we certainly agree with this sentiment, we also recognize that any research, however theoretical, must point to some termination in action; academics have a responsibility to make their work intelligible to practitioners. Thus we continue to seek projects that both comport with solid standards of scholarship, and address relevant acquisition issues. These years of experience have shown us the difficulty in attempting to balance these two objectives, but we are convinced that the attempt is absolutely essential if any real improvement is to be realized.

We gratefully acknowledge the ongoing support and leadership of our sponsors, whose foresight and vision have assured the continuing success of the Acquisition Research Program:

- Office of the Under Secretary of Defense (Acquisition, Technology & Logistics)
- Program Executive Officer SHIPS
- Commander, Naval Sea Systems Command
- Army Contracting Command, U.S. Army Materiel Command
- Program Manager, Airborne, Maritime and Fixed Station Joint Tactical Radio System

- Program Executive Officer Integrated Warfare Systems
- Office of the Assistant Secretary of the Air Force (Acquisition)
- Office of the Assistant Secretary of the Army (Acquisition, Logistics, & Technology)
- Deputy Assistant Secretary of the Navy (Acquisition & Logistics Management)
- Director, Strategic Systems Programs Office
- Deputy Director, Acquisition Career Management, US Army
- Defense Business Systems Acquisition Executive, Business Transformation Agency
- Office of Procurement and Assistance Management Headquarters, Department of Energy

We also thank the Naval Postgraduate School Foundation and acknowledge its generous contributions in support of this Symposium.

James B. Greene, Jr.
Rear Admiral, U.S. Navy (Ret.)

Keith F. Snider, PhD
Associate Professor

Panel 13 – Harnessing Program Interdependencies for Efficient Acquisition

Wednesday, May 11, 2011	
3:30 p.m. – 5:00 p.m.	**Chair: Patrick M. Sullivan,** Executive Director, Command, Control, Communications, Computers, & Intelligence **Discussant: Mark Krzysko,** Deputy Director, Acquisition Resources & Analysis; Enterprise Information ***A Web Service Implementation for Large-Scale Automation, Visualization and Real-Time Program Awareness via Lexical Link Analysis*** Ying Zhao, Shelley Gallup, and Douglas MacKinnon, NPS ***Programmatic Complexity & Interdependence: Emerging Insights and Predictive Indicators of Development Resource Demands*** Maureen Brown, Robert Kravchuk, and Graham Owen, UNC Charlotte

Patrick M. Sullivan—Executive Director, Program Executive Office for Command, Control, Communications, Computers and Intelligence (PEO C4I), San Diego, CA. Mr. Sullivan is responsible for integrating, executing and delivering capability in a $2.5 billion portfolio supporting information needs for naval, joint, and coalition warfighters. Mr. Sullivan received a bachelor's degree in electrical and computer engineering from the University of California, San Diego (UCSD) in 1989 and continued at the university to earn his master's degree in electrical engineering and applied physics in 1991.

Mr. Sullivan began his government career at the Naval Ocean Systems Center a predecessor of the Space and Naval Warfare Systems Center Pacific (SSC-Pacific). From 1991 to 1996, he was a project manager for the Design and Development Branch, working with the Defense Advanced Research Projects Agency Electronics Technology Office to develop new initiatives in the area of advanced electronic packaging. From September 1996 to January 2000, Mr. Sullivan was a project manager for the Integrated Circuit Research and Fabrication Branch, responsible for developing, managing, and performing as principal investigator for several advanced microelectronic research and development projects.

In January 2000, Mr. Sullivan assumed responsibilities as the head of the Integrated Circuit Research and Fabrication Branch, where he focused on microelectronic technology development for the strategic space and intelligence communities. From August 2002 through June 2006, Mr. Sullivan led the Joint and National Systems Division, supplying advanced technology to the intelligence and special operations communities. In March 2006, Mr. Sullivan was selected to lead the Intelligence, Surveillance and Reconnaissance Department and entered the Senior Executive Service later that year. His responsibilities in this position at SSC-Pacific included managing a broad set of programs to develop capabilities in the areas of maritime surveillance and ocean systems, joint and national information systems, intelligence systems, signal exploitation and cryptologic systems, and systems to support information operations and battlespace awareness. He also served as SPAWAR Engineering's National Competency Lead for ISR and Information Operations. He assumed his current position with PEO C4I in October 2010. Mr. Sullivan is a member of the UCSD Electrical and Computer Engineering Advisory Board, the National Defense Industrial Association, Armed Forces Communications and Electronics Association, and the Acquisition Professional Community.

Mark Krzysko—Deputy Director, Enterprise Information and Office of the Secretary of Defense Studies. In his senior leadership position, Mr. Krzysko oversees Federally Funded Research and Development Centers and directs data governance, technical transformation and shared services efforts to make timely, authoritative acquisition information available to support oversight of the Department of Defense's major programs, a portfolio totaling more than $1.6 trillion of investment funds over the lifecycle of the programs.

Preceding his current position, Mr. Krzysko served as ADUSD for Business Transformation, providing strategic guidance for re-engineering the Department's business system investment decision-making processes. He also served as ADUSD for Strategic Sourcing & Acquisition Processes and as Director of the Supply Chain Systems Transformation Directorate, championing and facilitating innovative uses of information technologies to improve and streamline the supply chain process for the Department of Defense. As the focal point for supply chain systems, he was responsible for transformation, implementation and oversight of enterprise capabilities for the acquisition, logistics and procurement communities. In addition, Mr. Krzysko served as advisor to the Deputy Under Secretary of Defense for Business Transformation on supply chain matters and as the functional process proponent to the Department's Business Transformation efforts, resulting in the establishment of the Business Transformation Agency.

In March of 2002, Mr. Krzysko joined the Defense Procurement and Acquisition Policy office as Deputy Director of e-Business. As the focal point for the Acquisition Domain, he was responsible for oversight and transformation of the acquisition community into a strategic business enterprise. This included driving the adoption of e-business practices across the Department, leading the move to modernize processes and systems, and managing the investment review process and portfolio of business systems.

Mr. Krzysko served as the Division Director of Electronic Commerce Solutions for the Naval Air Systems Command from June 2000 to March 2002. From April 1991 until March 2000, Mr. Krzysko served in various senior-level acquisition positions at the Naval Air Systems Command, including Contracting Officer of F/A-18 Foreign Military Sales, F/A-18 Developmental Programs, and the F-14. In addition, he served as Program Manager of Partnering, the Acquisition Business Process Re-engineering Effort, and as Acquisition Program Manager for the Program Executive Office for Tactical Aircraft.

Mr. Krzysko began his career in the private sector in various executive and managerial positions including Assistant Managing Director for Lord & Taylor Department Stores and Operations Administrator for Woodward & Lothrop Department Stores.

Mr. Krzysko holds a Bachelor of Science Degree in Finance from the University of Maryland, University College, College Park, MD, and a Master of General Administration in Financial Management from the same institution.

A Web Service Implementation for Large-Scale Automation, Visualization and Real-Time Program Awareness via Lexical Link Analysis

Ying Zhao—Research Associate Professor, Naval Postgraduate School. Dr. Zhao joined NPS in May 2009. Her research is focused on knowledge management approaches such as data text mining using lexical link analysis, search and visualization for system self-awareness, decision-making, and collaboration. She received her PhD in Mathematics from MIT and co-founded Quantum Intelligence, Inc. She has been Principal Investigator (PI) for six DoD Small Business Innovation Research (SBIR)-–awarded contracts and is a co-author of two patents in knowledge pattern search from networked agents, fusion, and visualization for multiple anomaly detection systems. [yzhao@nps.edu]

Shelley Gallup—Research Associate Professor, Department of Information Sciences, Naval Postgraduate School, and Director of Distributed Information and Systems Experimentation (DISE). Dr. Gallup has a multi-disciplinary science, engineering, and analysis background, including microbiology, biochemistry, space systems, international relations, strategy and policy, and systems analysis. He returned to academia after retiring from Naval service in 1994 and received his PhD in Engineering Management from Old Dominion University in 1998. Dr. Gallup joined NPS in 1999, bringing his background in systems analysis, Naval operations, military systems, and experimental methods first to the Fleet Battle Experiment series (1999–2002), then to the FORCEnet experimentation in the Trident Warrior series of experiments (2003–present). [spgallup@nps.edu]

Douglas MacKinnon—Research Associate Professor, Naval Postgraduate School. Dr. MacKinnon led an NPS research team to assess new MDA, spiral-1 technologies being fielded by PEO C4I developing original decision matrix structures and metrics structures to leverage the new technology. He has also led the assessment of the Tasking, Planning, Exploitation, and Dissemination (TPED) process during field experiments Empire Challenge 2008 and 2009 (EC08/09). He holds a PhD from Stanford University, conducting theoretic and field research in Knowledge Management (KM). He has served as the Program Manager for two major government projects of over $50 million each, implementing new technologies while reducing manpower requirements. He has served over 20 years as a Naval Surface Warfare Officer, amassing over eight years at sea, serving in four U.S. Navy warships with five major, underway deployments. [djmackin@nps.edu]

Abstract

DoD acquisition is an extremely complex system, comprised of myriad stakeholders, processes, people, activities, and organizational structures. Processes within this complex system are encumbered by the continuous development of large amounts of unstructured and unformatted acquisition program data, which is narrowly useful but difficult to aggregate across the "enterprise." Yet, acquisition analysts and decision-makers must analyze all types and spectrums of the available data to obtain a complete and understandable picture. This is a kind of systems *non-congruence* that has been difficult to overcome. For those embedded within the complexities of the acquisition community, this can be a daunting, if not impossible, task. We will apply a data-driven automation system, namely, Lexical Link Analysis (LLA), to facilitate acquisition researchers and decision-makers to recognize important connections (concepts) that form patterns derived from dynamic, ongoing data collection. The LLA technology and methodology is used to uncover and display relationships among competing programs and Navy-driven requirements. In the past year, we tested our method using samples of acquisition data for visualization and validity. LLA was demonstrated to discover statistically significant correlations and automatically extract the links that might require expensive manpower to perform otherwise (imagine use of many contractors, continually

looking through documentation and adding excerpts to categories of interest in various spreadsheets). This year, we started to develop LLA from a demonstration to an operational capability and facilitate a wider range of acquisition research applications. If successful, the resulting system could facilitate real-time awareness, reduce the workload of decision-makers, and make a profound impact on the long-term success of acquisition strategies—by revealing the current status of acquisition programs and connections within and external to contributing or competing interests, as well as inform potential strategic choices available to decision-makers.

Significance of the Research

Acquisition research has increased in component, organizational, technical, and management complexity. It is difficult for acquisition professionals to remain continuously aware of their decision-making domains because information is overwhelming and dynamic. According to the *Chairman of the Joint Chiefs of Staff Instruction for Joint Capabilities Integration and Development System (JCIDS)* (CJCS, 2009), there are three key processes in the DoD that must work in concert to deliver the capabilities required by the warfighters: the requirements process; the acquisition process; and the Planning, Programming, Budget, and Execution (PPBE) process.

Each process produces a large amount of data in an unstructured manner; for example, the warfighters' requirements are documented in Universal Joint Task Lists (UJTLs), Joint Capability Areas (JCAs), and Urgent Need Statements (UNSs). These requirements are processed in the JCIDS to become projects and programs, which should result in products such as weapon systems that meet the warfighters' needs. Program data are stored in the Defense Acquisition System (DAS). Programs are divided into Major DoD Acquisition Programs (MDAP), Acquisition Category II (ACATII), etc. Program Elements (PE) are the documents used to fund programs yearly through the congressional budget justification process. Data is too voluminous, too unformatted, and too unstructured to be easily digested and understood—even by a team of acquisition professionals. There is a critical need for automation to help reveal to decision-makers and researchers the interrelationships within these processes (see Figure 1).

We have attempted to develop and frame our research efforts around research questions in the following categories: conceptual, focused, theory development, and methodology.

Conceptual

- How can the information that emerges from the acquisition process be used to produce overall awareness of the fit between programs, projects, and systems and of the needs for which they were intended?
- If a higher level of awareness is possible, how will that enable system-level regulation of programs, projects, and systems for improvement of the acquisition system?

Focused

- Based on the normal evolution of documentation and on the current data-based program information, how can requirements (needs) be connected to system capabilities via automation of analysis?
- Can requirements gaps be revealed?

Theory Development

- How can a correlation between system interdependency (links/relationships) and development costs be shown if present?

Methodology

- How can we use natural language and other documentation (roughly, unformatted data) to produce visualization of the internal constructs useful for management through Lexical Link Analysis (LLA)?

Lexical analysis ("Lexical Analysis," 2010) is a form of text mining in which word meanings are developed from the context from which they are derived. Link analysis, a subset of network analysis that explores associations between objects, reveals the crucial relationships between objects when collected data may not be complete. Lexical Link Analysis (LLA) is an extended lexical analysis and link analysis. LLA can also be used in a learning mode in which such features and context associations are initially unknown and are constantly being learned, updated, and improved as more data become available.

We consider that the cognitive interface between decision-makers and a complex system may be expressed in a range of terms or features (i.e., a specific vocabulary or lexicon) to describe attributes and the surrounding environment of a system. Here, system self-awareness, or program awareness (Gallup, MacKinnon, Zhao, Robey, & Odell, 2009), allows decision-makers to be aware of what systems, programs, and products are available for acquisition; to understand how the systems match warfighters' needs and requirements; to recognize relationships among them; to improve efficiency of available collaboration; to reduce duplication of effort; and to reuse components to support cost-effective management with greater immediacy, possibly in real-time.

In the past year, we began at the Naval Postgraduate School (NPS) by using Collaborative Learning Agents (CLA; QI, 2009) and expanded to other tools, including AutoMap (CASOS, 2009) for improved visualizations. Results from these efforts arose from leveraging intelligent agent technology via an educational license with Quantum Intelligence, Inc. CLA is a computer-based learning agent, or agent collaboration, capable of ingesting and processing data sources.

This approach is related to a number of extant tools for text mining, including Latent Semantic Analysis (LSA; Dumais et al., 1988), keyword analysis and tagging technology (Foltz, 2002), and intelligence analysis ontology for cognitive assistants (Tecuci et al., 2007). What results from this process is a learning model—like an ethnographic *code book* (Schensul et al., 1999).

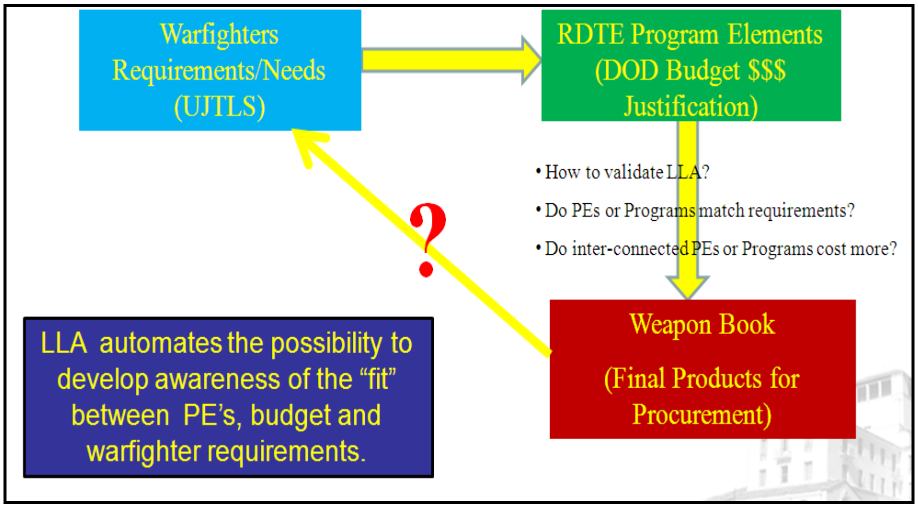


Figure 1. LLA Seeks to Inform the Business Processes Links (e.g., From Requirements to DoD Budget Justification to Final Products) That Are Critical for DoD Acquisition Research

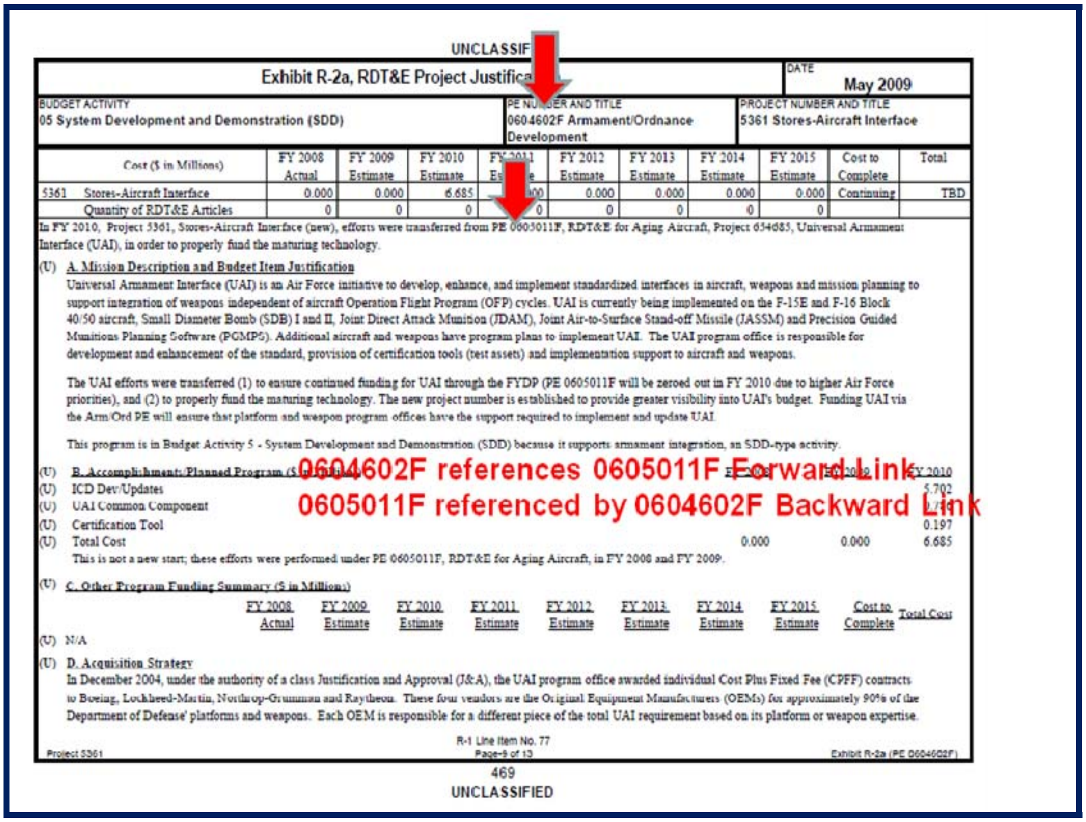
In precise terms, we observed that there were three important processes that seem fundamentally disconnected. They were the congressional budgeting justification process (such as information contained within the PEs), the acquisition process (such as information in the MDAP and ACATII), and the warfighters' requirements (such as information in UNSs and in UJTLs). They were not analyzed and compared together in a dynamic, holistic methodology that could keep up with changes and reflect patterns of relationships.

There had been little previous effort to integrate the data in these three components. For example, the Matrix Mapping Tool (MMT; Dahmann et al., 2005) included MDAP, UJTL, and JCA, yet did not include PE. Furthermore, in MMT, the links among programs and the matches to UJTL were extracted manually and were therefore not updated in a timely fashion. We employed the LLA automation to analyze more data, and we achieved a better outcome and provided dynamic, real-time integration. We focused our efforts on demonstrating validation and visualization and on providing insights for decision-makers on the large-scale data, as described in the next section.

The Validation for Using Large-Scale Data

To realize the potential of the LLA method, an important first step was to establish the validity of the method in the context of realistic, large-scale data sets. In the past year, we started to work on larger scale, open-source acquisition data sets. We obtained the Research, Development, Test and Evaluation (RDT&E) congressional budget justification documents (e.g., PEs from the DoD Comptroller website, http://comptroller.defense.gov/defbudget/). We also obtained program data including MDAPs data and ACATII data, UJTLs data, and Weapon Books data from the DoD open-source websites and our OSD contacts.

We first applied LLA to extract the links based on PEs for the RDT&E congressional budget justification process. PEs were observed at the center of many documents because each PE listed all the programs that the PE funded and their costs for the one- and five-year projections. Specifically, we compared the trends of LLA with what human analysts had identified manually. As shown in Figure 2, in each PE exhibition, another PE might be referenced, indicted as directionally linked PEs. For example, in Figure 2, PE 0604602F referenced PE 0605011F, which we defined as a *forward link* for PE 0604602F, while PE 0605011F was referenced by PE 0604602F, which we defined as a *backward link* for PE 0605011F. A backward link was usually a stronger indicator of the importance of a PE than was a forward link. This indicator was similar to the page ranking in a search engine (Gerber, 2005). In this research, we combined the total number of forward and backward links identified by human analysts as the attributes to correlate with the total number of machine-discovered lexical links. The Pearson correlation between the links identified by human analysts and by the LLA method was 0.39 with a p value < 0.0000001 (bidirectional t test with a sample size N = 461). This was an earlier validation for the LLA method that was achieved in Phase I of the research (Zhao et al., 2010).



UNCLASSIF[illegible]

Exhibit R-2a, RDT&E Project Justifica[illegible]	DATE May 2009

BUDGET ACTIVITY	PE NUMBER AND TITLE	PROJECT NUMBER AND TITLE
05 System Development and Demonstration (SDD)	0604602F Armament/Ordnance Development	5361 Stores-Aircraft Interface

Cost ($ in Millions)	FY 2008 Actual	FY 2009 Estimate	FY 2010 Estimate	FY 2011 Estimate	FY 2012 Estimate	FY 2013 Estimate	FY 2014 Estimate	FY 2015 Estimate	Cost to Complete	Total
5361 Stores-Aircraft Interface	0.000	0.000	6.685	[illegible]	0.000	0.000	0.000	0.000	Continuing	TBD
Quantity of RDT&E Articles	0	0	0	0	0	0	0	0		

In FY 2010, Project 5361, Stores-Aircraft Interface (new), efforts were transferred from PE 0605011F, RDT&E for Aging Aircraft, Project 654685, Universal Armament Interface (UAI), in order to properly fund the maturing technology.

(U) A. Mission Description and Budget Item Justification

Universal Armament Interface (UAI) is an Air Force initiative to develop, enhance, and implement standardized interfaces in aircraft, weapons and mission planning to support integration of weapons independent of aircraft Operation Flight Program (OFP) cycles. UAI is currently being implemented on the F-15E and F-16 Block 40/50 aircraft, Small Diameter Bomb (SDB) I and II, Joint Direct Attack Munition (JDAM), Joint Air-to-Surface Stand-off Missile (JASSM) and Precision Guided Munitions Planning Software (PGMPS). Additional aircraft and weapons have program plans to implement UAI. The UAI program office is responsible for development and enhancement of the standard, provision of certification tools (test assets) and implementation support to aircraft and weapons.

The UAI efforts were transferred (1) to ensure continued funding for UAI through the FYDP (PE 0605011F will be zeroed out in FY 2010 due to higher Air Force priorities), and (2) to properly fund the maturing technology. The new project number is established to provide greater visibility into UAI's budget. Funding UAI via the Arm/Ord PE will ensure that platform and weapon program offices have the support required to implement and update UAI.

This program is in Budget Activity 5 - System Development and Demonstration (SDD) because it supports armament integration, an SDD-type activity.

(U) B. Accomplishments/Planned Program ($ in Millions)	FY 2008	FY 2009	FY 2010
(U) ICD Dev/Updates			5.702
(U) UAI Common Component			[illegible]
(U) Certification Tool			0.197
(U) Total Cost	0.000	0.000	6.685

This is not a new start; these efforts were performed under PE 0605011F, RDT&E for Aging Aircraft, in FY 2008 and FY 2009.

(U) C. Other Program Funding Summary ($ in Millions)

	FY 2008 Actual	FY 2009 Estimate	FY 2010 Estimate	FY 2011 Estimate	FY 2012 Estimate	FY 2013 Estimate	FY 2014 Estimate	FY 2015 Estimate	Cost to Complete	Total Cost
(U) N/A										

(U) D. Acquisition Strategy

In December 2004, under the authority of a class Justification and Approval (J&A), the UAI program office awarded individual Cost Plus Fixed Fee (CPFF) contracts to Boeing, Lockheed-Martin, Northrop-Grumman and Raytheon. These four vendors are the Original Equipment Manufacturers (OEMs) for approximately 90% of the Department of Defense' platforms and weapons. Each OEM is responsible for a different piece of the total UAI requirement based on its platform or weapon expertise.

Project 5361 R-1 Line Item No. 77 Page-9 of 13 Exhibit R-2a (PE 0604602F)

469

UNCLASSIFIED

Figure 2. PE Cross-References Identified by Human Analysts

Figure 3 shows the accuracy of using the LLA method to predict the links between PEs. The x axis showed sorted PEs by three methods: random (green), LLA (red), and human (blue). The y axis showed the corresponding accumulative percentages of the PEs that were predicted correctly and that were linked to other PEs. The x-y curve was called a *Gains Chart*. As shown in Figure 3, there were about 225 PEs that had at least one link to other PEs identified by human analysts (the blue line); 80% of them were predicted by the top 225 PEs sorted by the LLA scores (the red line). In other words, LLA was used to predict correctly 80% of the links identified by the human analysts.

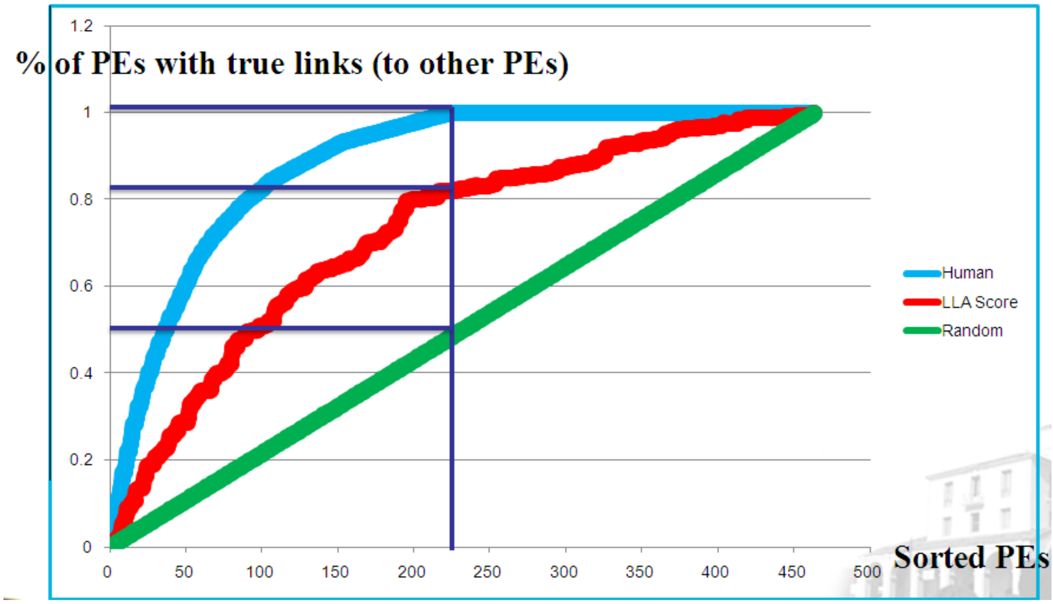


Figure 3. Use of LLA Scores to Predict PE Links: A Gains Chart

As shown in Figure 4, LLA was also used to discover the links that human analysts might not be able to identify—in the example, only the yellow link was identified by human analysts; the remainder were identified by LLA.

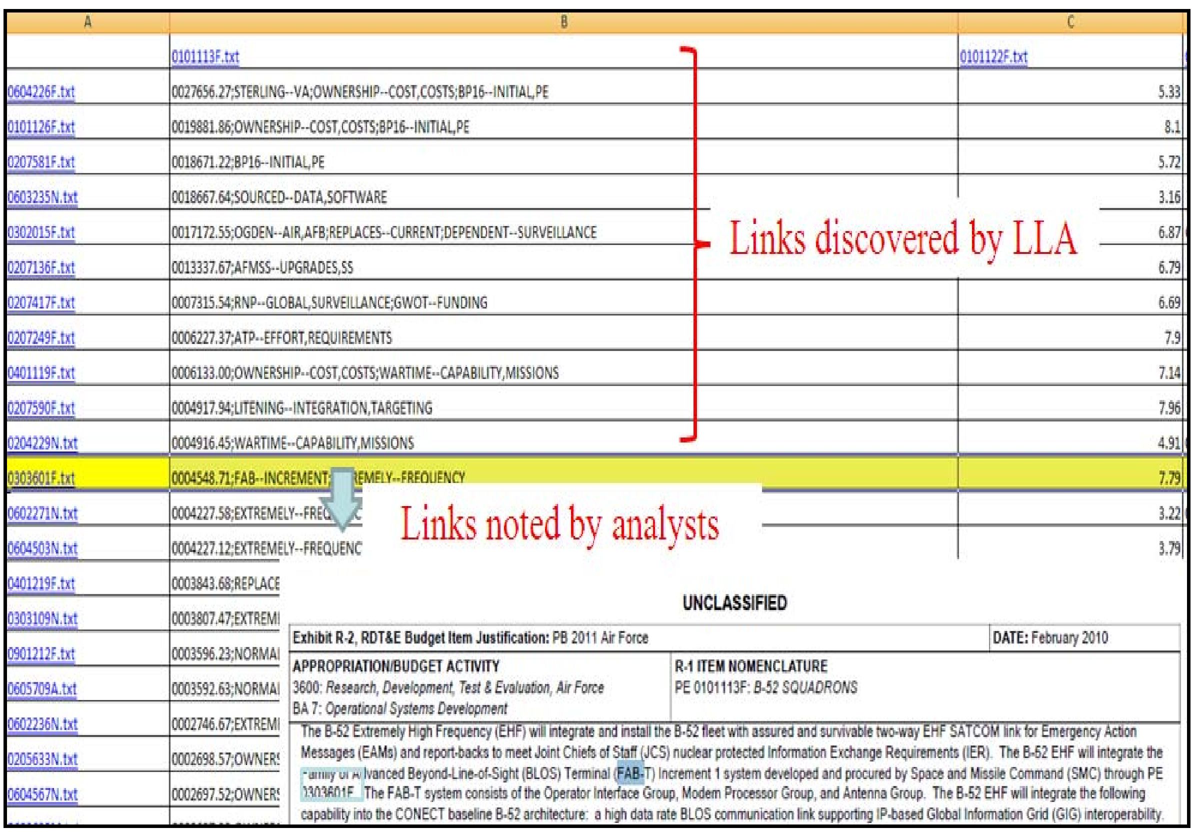


Figure 4. LLA Discovers Links That Are Not Identified by Human Analysts

As an additional demonstration of correlation, Figure 5 shows about 450 PEs sorted according to the numbers of LLA-generated links for each PE with respect to other PEs. Such links were considered as measures for independencies among PEs. We found that LLA-generated links were correlated with the ones identified by human analysts. The correlation between was 0.57, shown in the linear relationship in Figure 5. If outliers (circled in Figure 5) were removed, then the correlation was 0.62. This was a better validation of the LLA method than the one reported in Phase I of the project (Zhao et al., 2010).

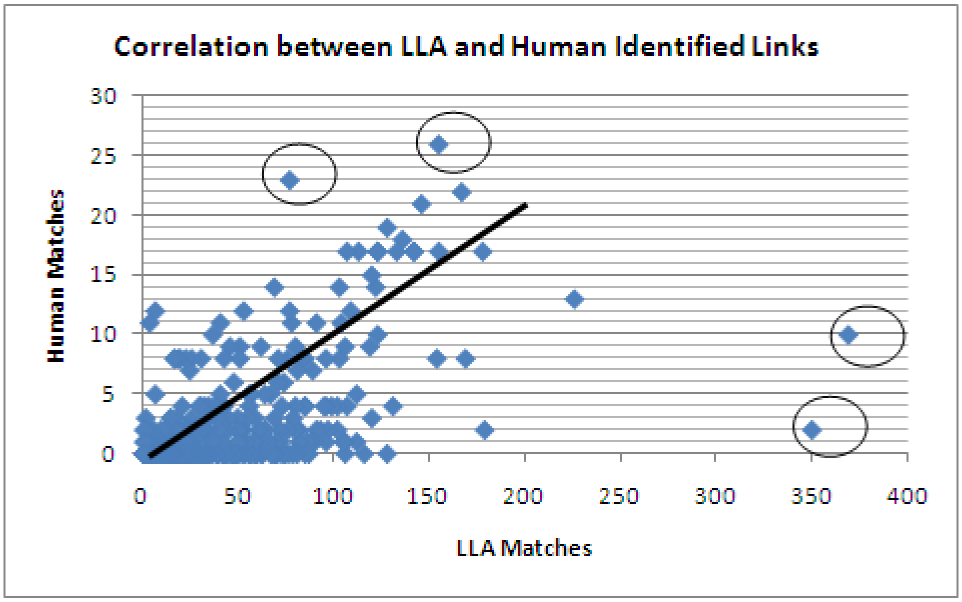


Figure 5. Correlation Between LLA and Human-Identified Matches

Initial Results for Phase II

During the Phase II research period (begun in 2011), we proposed a follow-on research to the NPS Acquisition Research Program using Lexical Link Analysis (LLA). Our goals for Phase II were to:

- Apply LLA to larger-scale data and wider applications, and employ parallel computing and dynamic, 3-D visualizations.
- Apply LLA to become a real-time operational capability of program awareness, the results of which could be periodically updated and presented in a web service.

We started on developing a web service that was designed to integrate the capability we explored in Phase I of the research into an operational capability, which links the budgeting process through PEs to the acquisition process via acquisition programs (MDAPs, ACATIIs) to the warfighters' requirements (UNS, UJTL, etc). We implemented an LLA platform from which to present periodically all the information in a single location so that users can view the trends based on the data in each of the three areas. We gathered the most recent documents in three areas from the following sources:

1. PEs: http://www.dtic.mil/descriptivesum/
2. MDAPs & ACATIIs:
   http://comptroller.defense.gov/defbudget/fy2008/fy2008_weabook.pdf
   http://www.fas.org/man/dod-101/sys/land/wsh2007/index.html

http://www.acq.osd.mil/ara/am/sar/

3. UJTLs: http://www.dtic.mil/doctrine/jel/cjcsd/cjcsm/m350004d.pdf

The web service described below can dramatically speed up efforts to collect the data. For example, each of the 24 sets of PE documents above contained about 200 PDF PEs from http://www.dtic.mil/descriptivesum/, totaling about 5,000 documents. Manually downloading and extracting desired links would be considered extremely difficult and very time intensive. By submitting several parallel jobs to the NPS High Performance Computing (HPC) center, the download took approximately six hours.

Web Service Design

Figure 6 shows the initial web service design detailed as follows:

- Tomcat (http://tomcat.apache.org/index.html) was used as the infrastructure to host multiple learning agents for the web service. A CLA system (QI, 2009) of multiple agents was installed in multiple or a single Tomcat.

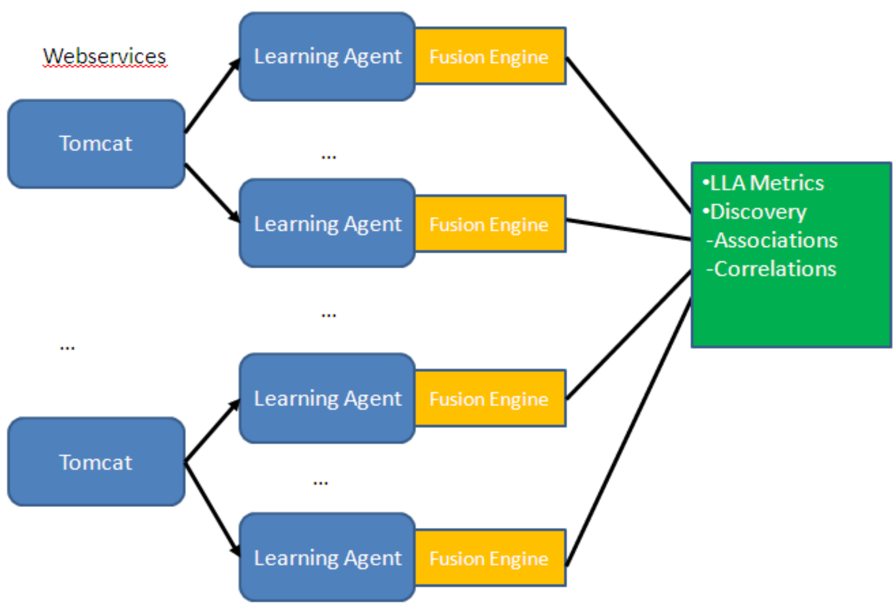


Figure 6. Initial Web Service Design

In Figure 7, the ARP web service is shown hosted via http://disedev4.ern.nps.edu:8080/ARP, which is one of few servers dedicated to the project at the NPS DISE lab. Eventually, we will move the service to the NPS HPC center, where hundreds of learning agents can be hosted in the cloud computing environment Hamming Linux cluster to gather, analyze, and disseminate information in a massive, parallel fashion. The web service administration function includes the following capabilities:

- Peer List: allows the current agent to list the peers with which it shares index and learning models

- One Click Mining: involves "one click" to index and mine the data stored locally
- Properties: parameters used in the "One Click Mining"
- Dashboard Monitor: lexical links discovered from the mining process are displayed continuously through the dashboard
- Back to search: provides the capability to allow a basic search

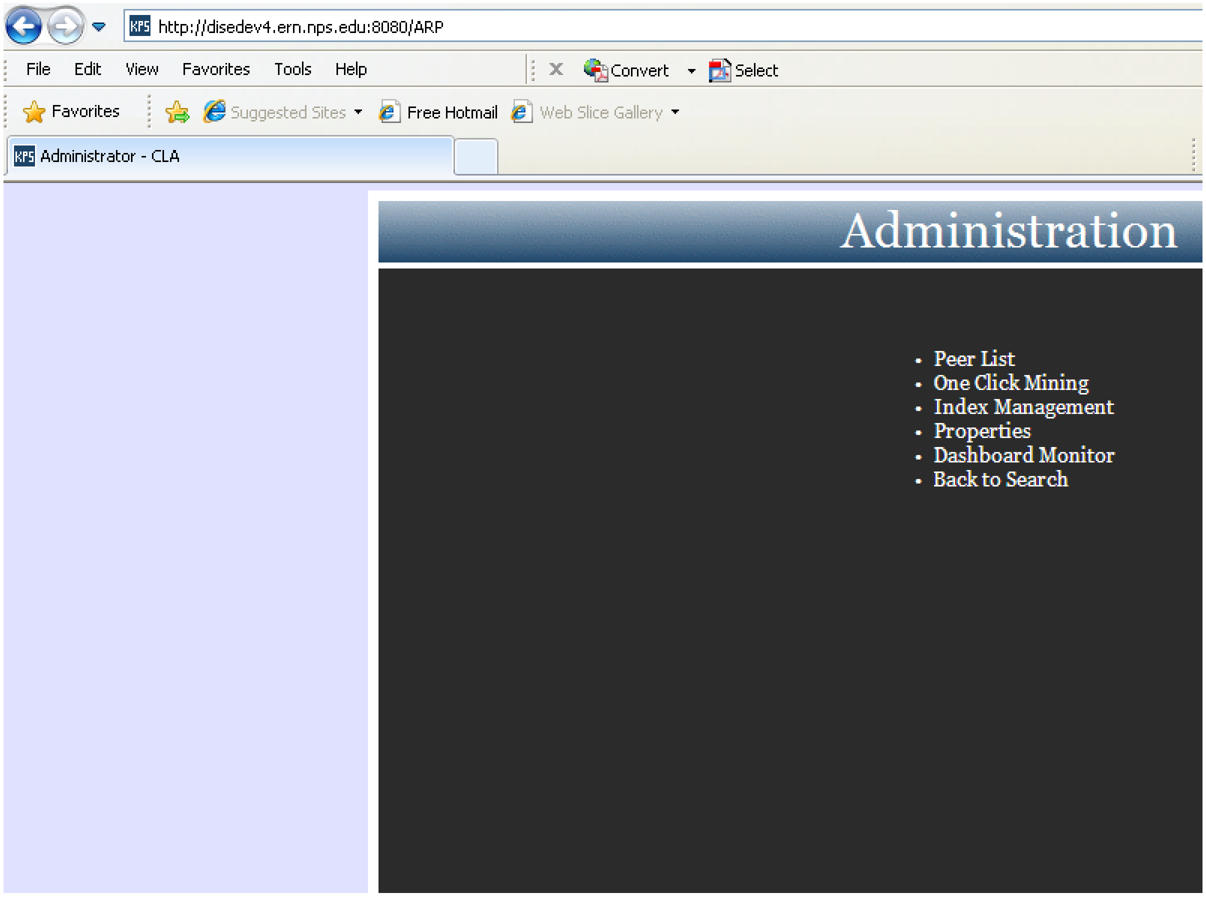


Figure 7. Web Service Hosted Using Tomcat

- A single learning agent was implemented to mine the data that were gathered in each of the categories above (e.g., PEs of Air Force in 2011, as shown in the "One Click Mining" capability in Figure 8). "Path to Data" was used to point to the data stored locally. "Index Name" was used to store the search index and learning model generated from the data.

Figure 8. One Click Mining

- The indexes or learning models generated from above are stored locally in each learning agent, as shown in the “Index Management” in Figure 9. A fusion engine can also be attached to a learning agent. The function of the fusion engine is to combine lexical links discovered from the local index/learning model with the lexical links discovered from its peers in a recursive manner, therefore to form a fused view of all the data from the total learning agent network. As shown in Figure 9, when clicked via “Fuse,” the indexes/learning models selected (e.g., navy_2009, navy_2010, and navy_2011) were combined into one model.

Figure 9. Fusion Engine

An index or learning model contains the following functions:

- Lexical links are highlighted in the search results as shown in the dashboard display in Figure 10. When a lexical link is clicked via “Investigate,” a search is invoked and the source documents containing the link are listed and highlighted.

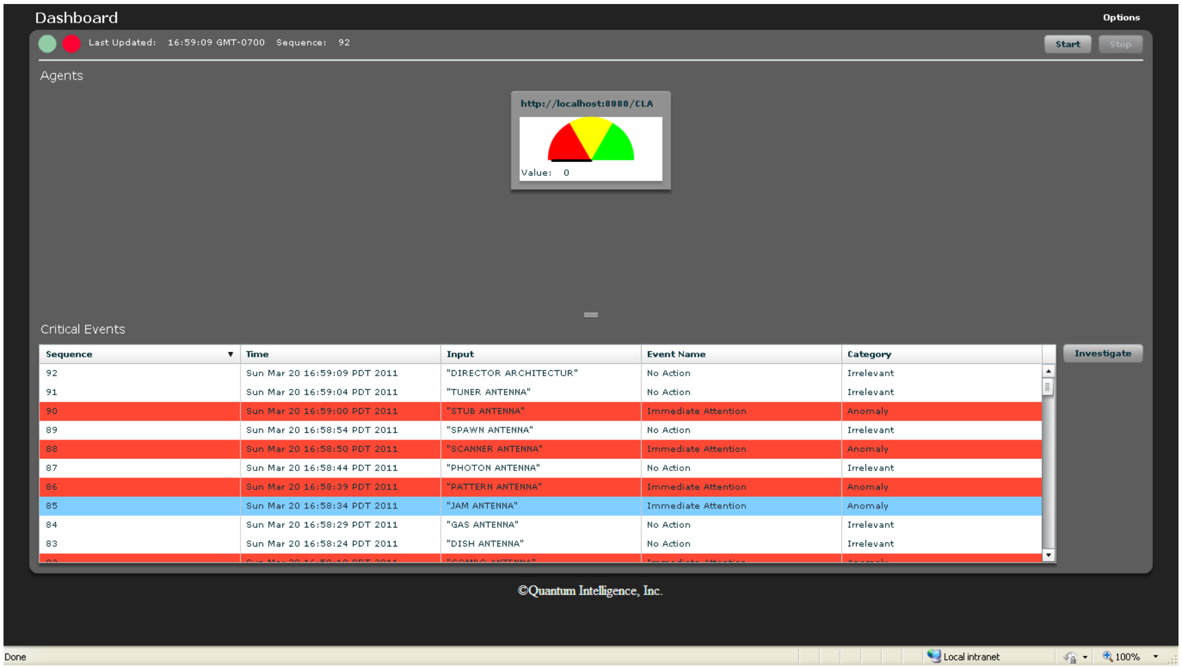


Figure 10. Dashboard to Display Lexical Links Discovered

- The key metrics of lexical link counts are used to measure overlaps and gaps between PEs and other categories of information such as MDAPs, UNS/UJTLs, and changes over time.

The fusion engine described above fuses the learning models and then groups the lexical links into categories to look at the links and overlaps among different services and over years in detail. As shown in Figure 11, a single category (theme) using a triple of word hubs of *Tactic, Combat*, and *Effort* contain lexical links related to the category from different sources of Navy PEs from 2009–2011: red indicates links only in 2011; green indicates links only in 2010; and blue indicates links only in 2009. The purple links are the ones that are in more than two sources.

Figure 11 illustrates a single category (theme) using a triple of word hubs of *Tactic, Combat*, and *Effort* that contain lexical links related to the category from different sources—specifically, Navy PEs from 2009–2011: red indicates links only in 2011; green indicates links only in 2010; and blue indicates links only in 2009. The purple links are those that are found in more than two sources.

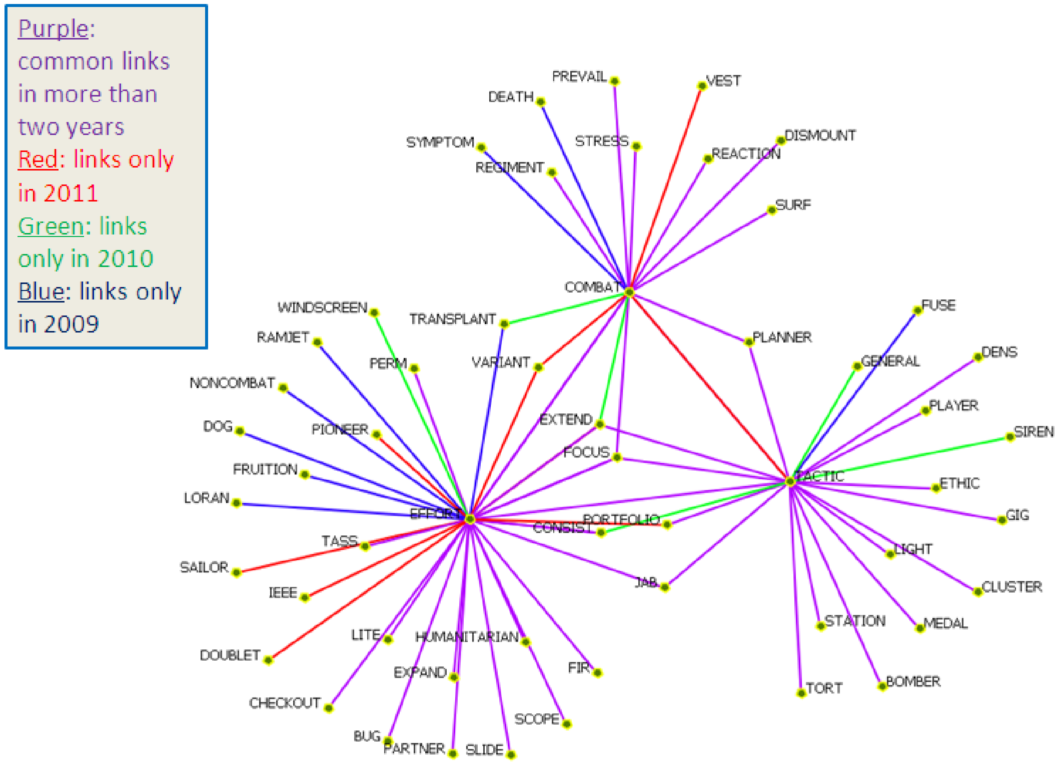


Figure 11. Lexical Links Grouped Into Categories

Figure 12 shows all the groups in one view. Each connected link represents a set of features that belong to a group, such as “Tactic-Combat-Effort” shown in Figure 11. As shown in Figure 12, the total number of features stayed deleted and added over the years (2009–2011) was computed from the lexical links, respectively.

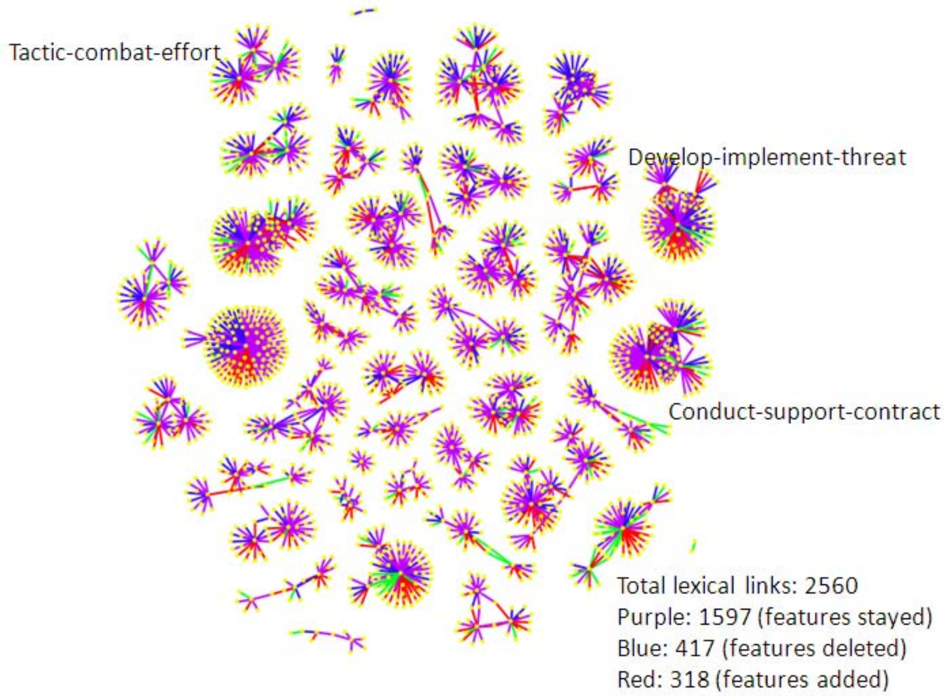


Figure 12. Overall View of Three Years of Navy PEs

LLA networks are visualized using a set of commonly known social network tools such ORA shown above. Another tool we explored is Pajek. Pajek exports a network in an .xsd format, which can be viewed in 3-D, which is a product from the Modeling, Virtual Environments, and Simulation (*MOVES*) Institute at NPS for 3-D visualization and further navigation.

Social Network of PEs

We have been using the initial implementation of the LLA web service in the workflow that benefits acquisition professionals. As an example, the Fusion Engine was used to construct a social network view of PEs. Figures 13 and 14 illustrate the differences between LLA-discovered linkages and those found by human analysts. In Figure 13, PE 0603721N is linked to PEs 0602435N, 0602782N, 0601153N, and 0603235N. Figure 14 indicates PEs identified by human analysts. Titles for the PEs are the following:

- 0602435N: Ocean Warfighting Environment Applied Research
- 0602782N: Mine and Expeditionary Warfare Applied Research
- 0601153N: Defense Research Sciences
- 0603235N: Common Picture Advanced Technology

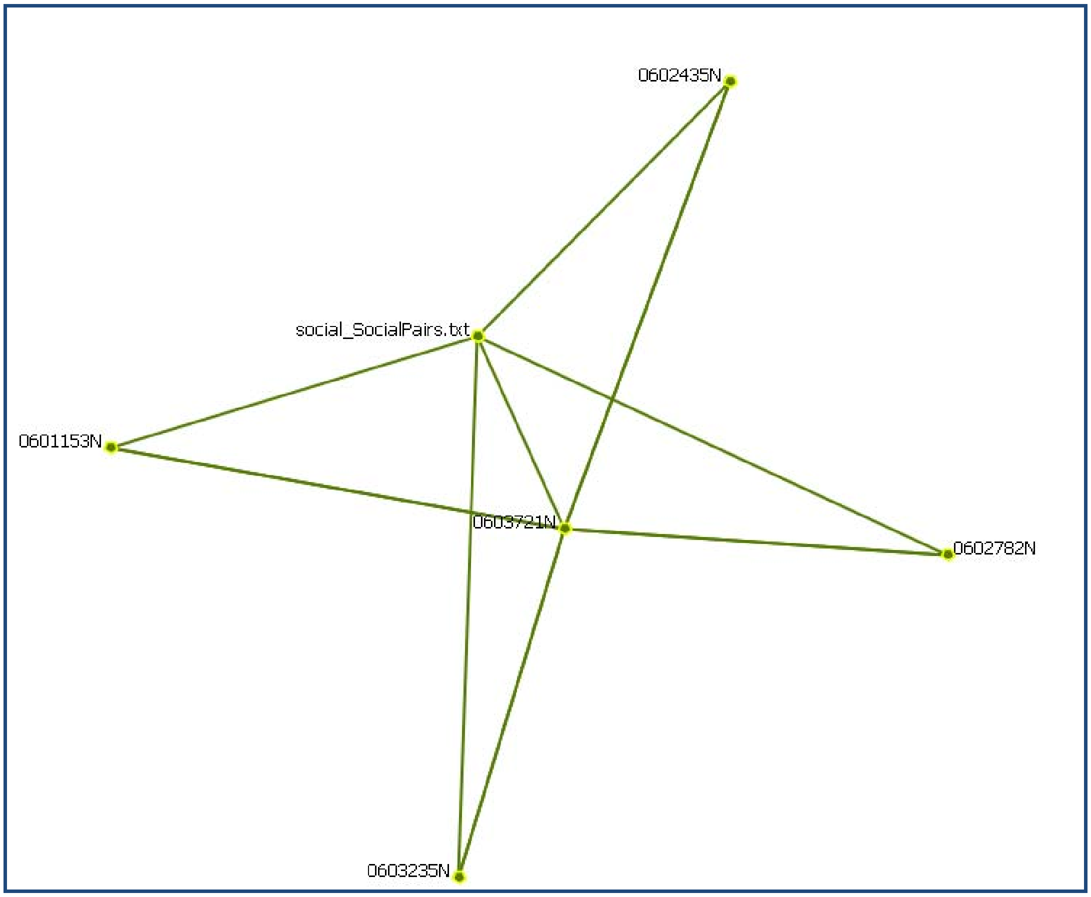


Figure 13. Social Network of PE 0603721N

APPROPRIATION/BUDGET ACTIVITY	PROGRAM ELEMENT NUMBER AND NAME	PROJECT NUMBER AND NAME
RDTEN/BA 4	**0603721N/ENVIRONMENTAL PROTECTION**	**9204/Marine Mammal Research**

FY09: (U) Continue mitigation methodologies for monitoring, new technology and risk assessment through passive acoustic monitoring; active acoustic monitoring; Improved tag development; alternative monitoring; defining risk assessment variables; model risk assessment and determine mitigation effectiveness.

	FY 2007	FY 2008	FY 2009
Acoustic Source Propagation	0.150	0.085	0.113
RDT&E Articles Quantity	0	0	0

FY 07: (U) Continue investigation of acoustic source propagation through 3-D modeling of multiple acoustic sources.

FY08: (U) Continue investigation of acoustic source propagation through 3-D modeling of multiple acoustic sources.

FY09: (U) Continue investigation of acoustic source propagation through 3-D modeling of multiple acoustic sources.

C. OTHER PROGRAM FUNDING SUMMARY:

(U) Related RDT&E: Office of Naval Research (PE 0601153N / PE 0602435N / PE 0602782N / PE 0603235N)

(U) Related RDT&E: Strategic Environmental Research & Development Program (SERDP)

(U) Related RDT&E: National Oceanographic Partnership Program (NOPP)

Figure 14. PE 0603721N Linked to PEs Identified by Human Analysts

Semantic Network of PEs

Compared to the links identified by human analysts, LLA was used to look into the links among PEs from all of the Services as a whole system, and therefore, the links discovered were cross-Service and potential cognitive *blind spots* of human analysts. For example, Figure 16 lists the semantic network for PE 0603721N discovered by LLA. Three of four human-identified links showed up in the top 100 of the LLA links, with 0601153N, 0602435N, and 0603235N ranked 33, 35, and 58, respectively.

Figure 15 shows a total social network view of the PEs using the links identified by human analysts for all the PEs in the year 2008 data and a 3-D view from Pajek. PEs ending with an *A* were Army PEs, those ending with an *F* were Air Force PEs, and those ending with an *N* were Navy PEs. As one can observe, the links in Figure 15 tended to be within the Services; for example, analysts tended to identify Army PEs linked to Army PEs, Air Force to Air Force, and Navy to Navy. The cost of each PE in 2008 is illustrated with the bubble size. As seen in Figure 15, PEs within the Services were more cross-referenced and the cost seemed inversely correlated to the links.

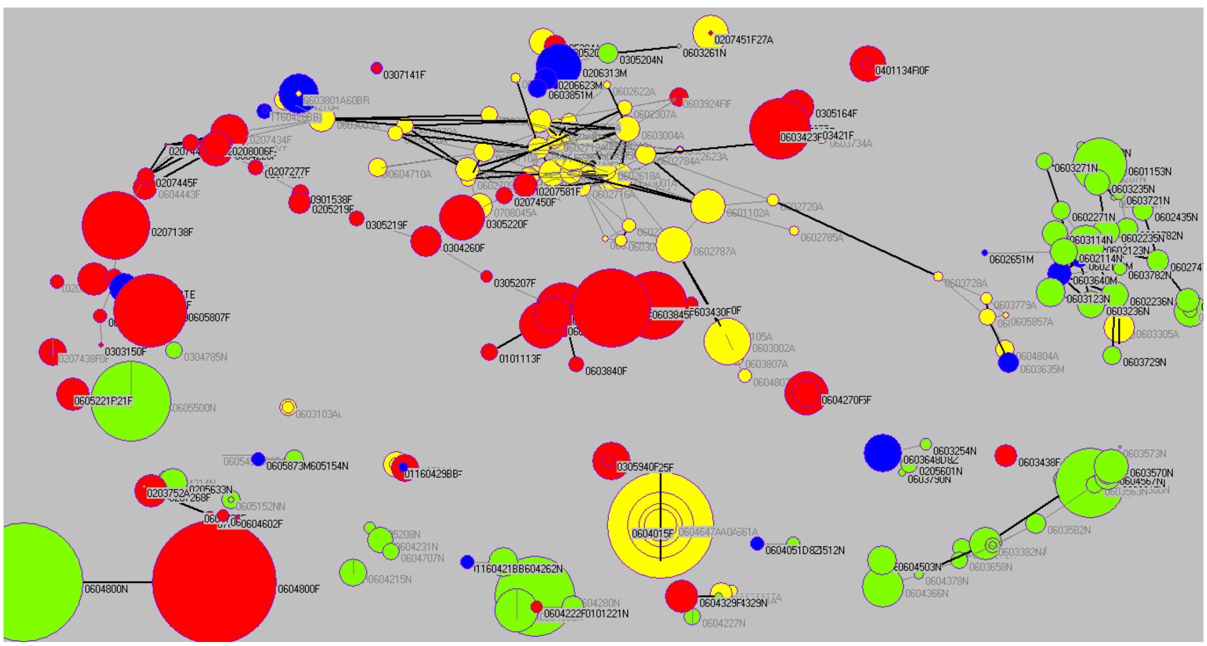


Figure 15. A Social Network View of PEs With the Links Identified by Human Analysts—A 3-D View from Pajek

Table 1. Semantic Network for PE 06043721N

1	0603721N	
2	0000012.74;JET--	0602787A
3	0000012.52;SEDIMENT--	0601102A
4	0000012.43;CLEAN--	0603804A
5	0000011.78;DESTRUCT--	0602203F
6	0000010.77;SHIELD--	0602102F
7	0000010.48;JET--	0601102F
8	0000010.33;CLEAN--	0604804A
9	0000009.67;DESTRUCT--	0604645A
10	0000008.90;SHIELD--	0604231N
11	0000008.49;UNIFORM--	0206313M
12	0000008.15;GAIN--	0603789F
13	0000008.06;ARLINGTON--VA,NJ;RIGOR--	0605013A
14	0000007.94;ARLINGTON--VA,NJ;LABOR--	0603747A
15	0000007.93;CONTENT--	0602202F
16	0000007.72;GAIN--	0603001A
17	0000007.69;JET--	0603790F
18	0000007.68;SHIELD--	0603640M
19	0000007.36;JET--	0603216F
20	0000007.11;SHIELD--	0602601F
21	0000007.11;CLEANUP--	0603728A
22	0000007.05;FINISH--	0605857A
23	0000006.98;UNIFORM--	0601104A
24	0000006.89;SHIELD--	0603005A
25	0000006.73;SEDIMENT--	0602236N
26	0000006.63;ARLINGTON--	0305204A
27	0000006.62;UNIFORM--	0604601A
28	0000006.62;CONTENT--	0602702F
29	0000006.60;DESTRUCT--	0604759F
30	0000006.59;FINISH--	0604661A
31	0000006.55;GOVT--	0604240F
32	0000006.43;GAIN--	0602120A
33	0000006.34;DESTRUCT--	0601153N
34	0000006.22;GAIN--	0604321A
35	0000006.13;SEDIMENT--	0602435N
36	0000005.98;EXPECT--	0602204F
37	0000005.88;STORM--	0207601F
38	0000005.85;GAIN--	0603231F
39	0000005.85;GAIN--	0303140F
52	0000005.01;NORMAL--	0207410F
53	0000004.98;DESTRUCT--	0603004A
54	0000004.93;LABOR--	0605801A
55	0000004.87;CONCERN--	0602747N
56	0000004.86;SHIELD--	0603561N
57	0000004.86;JET--	0603236N
58	0000004.80;SHIELD--	0603235N
59	0000004.76;JET--	0602618A
60	0000004.76;DESTRUCT--	0604660A
61	0000004.75;NORMAL--	0305206F
62	0000004.65;AGREEMENT--	0207418F
63	0000004.61;GAIN--	0604805A
64	0000004.56;JET--	0605805A
65	0000004.54;ARLINGTON--	0203758A
66	0000004.45;BREED--	0207451F
67	0000004.40;GOVT--	0604215N

In addition to the potential to discover human analysts' blind spots to connect the PEs across the Services, we also observed that LLA might discover unique and rare features or research perspectives that two PEs might share. Table 2 shows the examples of these links using highlighted word hubs in Table 1 for the top four PEs linked to PE 06043721N.

Table 2. Unique and Rare Semantic Links

Top 4 PEs linked to PE 06043721N	Titles	Semantic Links
0602787A	Medical Technology	Jet lag, jet fuel exposure
0601102A	Defense Research Sciences	Destruction, containment in water, soil, and sediments resulting from military activities
0603804A	Logistics and Engineer Equipment	The Army fights with clean fuel and drinking water
06032203F	Aerospace Propulsion	Non-destructive test, fuels and lubrication

Observations for the RDT&E Budget Justification Process

We took a detailed look at the RDT&E budget modification practice from 2008 to 2009, in an effort to see if LLA links identified among PEs and to UJTLs are correlated with the changes of the budget allocation from 2008 to 2009. Our observations are summarized in Table 3.

We observed that from 2008 to 2009, as shown in Table 3, the average 2009 budget change in terms of percentage change for each PE, whose number of LLA links to other PEs was larger than 10, was 14%, compared to 40%, whose number of LLA links to other PEs was fewer than 10. The total 2009 cost change was $558 million for the former and $434 million for the latter. This indicated that the current practice tended to reduce the budget for PEs with more links to other PEs and to increase the budget for the ones with less links, allocating resources to avoid overlapping efforts and to fund new and unique projects.

Table 3. Budget Change Sorted Using LLA Links From PEs to PEs

LLA links from PE to PE	Average Budget Change from 2008 to 2009 (in term of percentage change for each PE)	Total Budget Change in Millions
> 10	14%	($558)
< = 10	40%	$434

In contrast, the same 450 PEs sorted according to the numbers of LLA links with respect to UJTLs, as shown in Table 4. Overall, there were fewer numbers of LLA links observed, meaning that there were gaps between the RDT&E resource allocation and the warfighters' requirements. For PEs that had at least one LLA match to UJTLs, the average percentage cost change was 10%, compared to 29% for PEs that had no matches. This indicated a need to consider gaps and the warfighters' requirements as priorities in the RDT&E investment.

We found that the total cost change in dollars for PEs with at least one match to the UJTLs was $735 million, compared to $859 million for PEs with no matches. We found that this was due to the current practice that tended to cut from the budget the more expensive programs, such MDAPs, rather than the less expensive ones.

Table 4. Budget Change Sorted Using LLA Links from PEs to UJTLs

LLA links of PE to UJTL	Average Budget Change from 2008 to 2009 (in term of percentage change for each PE)	Total budget change in millions
> 1	10%	$735
< = 1	29%	($859)

These findings can be useful as validation and guidance for implementing Secretary of Defense Gates' defense cutting plan. For example, Secretary Gates said the Pentagon must get "more bang for its buck and shift its focus to the military's needs for the future" (Hedgpeth, 2010). Top acquisition officials in the nation have been looking for ways to limit spending, identify efficiencies, and eliminate unnecessary cost. Secretary Gates also planned to add 20,000 acquisition workers to implement the cost reduction. The program awareness implemented via the LLA method can link warfighters' requirements to budget and to final weapon products and help all the acquisition workers in their decision-making. The opportunities for the new acquisition workers could be to reduce the overall inefficiency of the 10% versus 29% illustrated in Table 4, instead of just focusing on the MDAPs.

References

Center for Computational Analysis of Social and Organizational Systems (CASOS). (2009). *AutoMap: Extract, analyze and represent relational data from texts*. Retrieved from http://www.casos.cs.cmu.edu

Chairman of the Joint Chiefs of Staff (CJCS). (2009). *Chairman of the Joint Chiefs of Staff instruction for joint capabilities integration and development system (JCIDS)* (J-8 CJCSI 3170.01G). Retrieved from http://www.intelink.sgov.gov/wiki/JCIDS

Dahmann, J., Baldwin, K., Bergin, D., Choudhary, A., Dubon, A., & Eiserman, G. (2005). *Matrix mapping tool (MMT)* (White paper). Washington, DC: OUSD (AT&L) and Defense Systems.

DoD. (2007, February). *Program acquisition costs by weapon system*. Retrieved from http://comptroller.defense.gov/defbudget/fy2008/fy2008_weabook.pdf

Dumais, S. T., Furnas, G. W., Landauer, T. K., Deerwester, S., & Harshman, R. (1988). Using latent semantic analysis to improve information retrieval. In *Proceedings of CHI'88: Conference on Human Factors in Computing* (pp. 281–285). New York, NY: Association for Computing Machinery.

Foltz, P. W. (2002). Quantitative cognitive models of text and discourse processing. In A. C. Graesser, M. A. Gernsbacher, & S. R. Goldman (Eds.), *The handbook of discourse processes*. Mahwah, NJ: Lawrence Erlbaum.

Gallup, S. P., MacKinnon, D. J., Zhao, Y., Robey, J., & Odell, C. (2009, October 6–8). Facilitating decision making, re-use and collaboration: A knowledge management approach for system self-awareness. In *Proceedings of the International Joint Conference on Knowledge Discovery, Knowledge Engineering, and Knowledge Management (IC3K)*. Madeira, Portugal.

Gerber, C. (2005). Smart searching, new technology is helping defense intelligence analysts sort through huge volumes of data. *Military Information Technology*, *9*(9). Retrieved from http://www.mkbergman.com

Hedgpeth, D. (2010, September 14). Gates starts outlining cuts to save $100 billion for defense. *Washington Post*. Retrieved from http://www.washingtonpost.com

Lexical analysis. (2010). In *Wikipedia*. Retrieved September 2010 from http://en.wikipedia.org/wiki/Lexical_analysis

Quantum Intelligence (QI). (2009). *Collaborative learning agents (CLA)*. Retrieved from http://www.quantumii.com/qi/cla.html

Schensul, J. J., Schensul, S. L., & LeCompte, M. D. (1999). *Essential ethnographic methods: Observations, interviews and questionnaires*. Lanham, MD: Rowman Altamira.

Stanford National Language Processing Group. (2009). *Stanford log-linear part-of-speech tagger*. Retrieved from http://nlp.stanford.edu/

Tecuci, G., Boicu, M., Marcu, D., Boicu, C., Barbulescu, M., Ayers, C., & Cammons, D. (2007). Cognitive assistants for analysts. In J. Auger, & W. Wimbish (Eds.), *Proteus futures digest: A compilation of selected works derived from the 2006 Proteus workshop*. Carlisle Barracks: National Intelligence University, Office of the Director of National Intelligence and U.S. Army War College Center for Strategic Leadership.

Zhao, Y., Gallup, S., & MacKinnon, D. (2010). Towards real-time program awareness via lexical link analysis. In *Proceedings of the Seventh Annual Acquisition Research Program*. Monterey, CA: Naval Postgraduate School.

Acknowledgments

We thank Mr. Dave Summer from the Naval Network Warfare Command (NNWC) for helping us understand warfighters' requirements in the Maritime Domain Awareness (MDA) area. We thank Mr. Robert Flowe from the Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics (OUSD[AT&L]), who pointed to large-scale data sets, provided critical acquisition research, asked relevant questions, and participated in insightful discussions with us.