NPS-AM-13-C10P01R010-039

EXCERPT FROM THE

# PROCEEDINGS

OF THE

## TENTH ANNUAL ACQUISITION RESEARCH SYMPOSIUM ACQUISITION MANAGEMENT

# Lexical Link Analysis Application: Improving Web Service to Acquisition Visibility Portal

**Ying Zhao, Shelley Gallup, and Douglas MacKinnon**
**Naval Postgraduate School**

Published April 1, 2013

The research presented in this report was supported by the Acquisition Research Program of the Graduate School of Business & Public Policy at the Naval Postgraduate School.

# Preface & Acknowledgements

Welcome to our Tenth Annual Acquisition Research Symposium! We regret that this year it will be a “paper only” event. The double whammy of sequestration and a continuing resolution, with the attendant restrictions on travel and conferences, created too much uncertainty to properly stage the event. We will miss the dialogue with our acquisition colleagues and the opportunity for all our researchers to present their work. However, we intend to simulate the symposium as best we can, and these *Proceedings* present an opportunity for the papers to be published just as if they had been delivered. In any case, we will have a rich store of papers to draw from for next year’s event scheduled for May 14–15, 2014!

Despite these temporary setbacks, our Acquisition Research Program (ARP) here at the Naval Postgraduate School (NPS) continues at a normal pace. Since the ARP’s founding in 2003, over 1,200 original research reports have been added to the acquisition body of knowledge. We continue to add to that library, located online at www.acquisitionresearch.net, at a rate of roughly 140 reports per year. This activity has engaged researchers at over 70 universities and other institutions, greatly enhancing the diversity of thought brought to bear on the business activities of the DoD.

We generate this level of activity in three ways. First, we solicit research topics from academia and other institutions through an annual Broad Agency Announcement, sponsored by the USD(AT&L). Second, we issue an annual internal call for proposals to seek NPS faculty research supporting the interests of our program sponsors. Finally, we serve as a “broker” to market specific research topics identified by our sponsors to NPS graduate students. This three-pronged approach provides for a rich and broad diversity of scholarly rigor mixed with a good blend of practitioner experience in the field of acquisition. We are grateful to those of you who have contributed to our research program in the past and encourage your future participation.

Unfortunately, what will be missing this year is the active participation and networking that has been the hallmark of previous symposia. By purposely limiting attendance to 350 people, we encourage just that. This forum remains unique in its effort to bring scholars and practitioners together around acquisition research that is both relevant in application and rigorous in method. It provides the opportunity to interact with many top DoD acquisition officials and acquisition researchers. We encourage dialogue both in the formal panel sessions and in the many opportunities we make available at meals, breaks, and the day-ending socials. Many of our researchers use these occasions to establish new teaming arrangements for future research work. Despite the fact that we will not be gathered together to reap the above-listed benefits, the ARP will endeavor to stimulate this dialogue through various means throughout the year as we interact with our researchers and DoD officials.

Affordability remains a major focus in the DoD acquisition world and will no doubt get even more attention as the sequestration outcomes unfold. It is a central tenet of the DoD’s Better Buying Power initiatives, which continue to evolve as the DoD finds which of them work and which do not. This suggests that research with a focus on affordability will be of great interest to the DoD leadership in the year to come. Whether you’re a practitioner or scholar, we invite you to participate in that research.

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

- Office of the Under Secretary of Defense (Acquisition, Technology, & Logistics)
- Director, Acquisition Career Management, ASN (RD&A)
- Program Executive Officer, SHIPS
- Commander, Naval Sea Systems Command
- Program Executive Officer, Integrated Warfare Systems
- Army Contracting Command, U.S. Army Materiel Command
- Office of the Assistant Secretary of the Air Force (Acquisition)
- Office of the Assistant Secretary of the Army (Acquisition, Logistics, & Technology)
- Deputy Director, Acquisition Career Management, U.S. Army
- Office of Procurement and Assistance Management Headquarters, Department of Energy
- Director, Defense Security Cooperation Agency
- Deputy Assistant Secretary of the Navy, Research, Development, Test, & Evaluation
- Program Executive Officer, Tactical Aircraft
- Director, Office of Small Business Programs, Department of the Navy
- Director, Office of Acquisition Resources and Analysis (ARA)
- Deputy Assistant Secretary of the Navy, Acquisition & Procurement
- Director of Open Architecture, DASN (RDT&E)
- Program Executive Officer, Littoral Combat Ships

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

Keith F. Snider, PhD
Associate Professor

# Acquisition Management

**Naval Ship Maintenance: An Analysis of the Dutch Shipbuilding Industry Using the Knowledge Value Added, Systems Dynamics, and Integrated Risk Management Methodologies**

David N. Ford, Thomas J. Housel, and Johnathan C. Mun
*Naval Postgraduate School*

**Time as an Independent Variable: A Tool to Drive Cost Out of and Efficiency Into Major Acquisition Programs**

J. David Patterson
*National Defense Business Institute, University of Tennessee*

**The Impact of Globalization on the U.S. Defense Industry**

Jacques S. Gansler and William Lucyshyn
*University of Maryland*

**Bottleneck Analysis on the DoD Pre-Milestone B Acquisition Processes**

Danielle Worger and Teresa Wu, *Arizona State University*
Eugene Rex Jalao, *Arizona State University and University of the Philippines*
Christopher Auger, Lars Baldus, Brian Yoshimoto, J. Robert Wirthlin, and John Colombi, *The Air Force Institute of Technology*

**Software Acquisition Patterns of Failure and How to Recognize Them**

Lisa Brownsword, Cecilia Albert, Patrick Place, and David Carney
*Carnegie Mellon University*

**Fewer Mistakes on the First Day: Architectural Strategies and Their Impacts on Acquisition Outcomes**

Linda McCabe and Anthony Wicht
*Massachusetts Institute of Technology*

**The Joint Program Dilemma: Analyzing the Pervasive Role That Social Dilemmas Play in Undermining Acquisition Success**

Andrew P. Moore, William E. Novak, Julie B. Cohen, Jay D. Marchetti, and Matthew L. Collins
*Software Engineering Institute, Carnegie Mellon University*

**Acquisition Risks in a World of Joint Capabilities: A Study of Interdependency Complexity**

**Measuring Dynamic Knowledge and Performance at the Tactical Edges of Organizations: Assessing Acquisition Workforce Quality**

Mark E. Nissen
*Naval Postgraduate School*

**Outcome-Focused Market Intelligence: Extracting Better Value and Effectiveness From Strategic Sourcing**

Timothy G. Hawkins, *Naval Postgraduate School*
Michael E. Knipper, *771 Enterprise Sourcing Squadron USAF*
Timothy S. Reed, *Beyond Optimal Strategic Solutions*

# Lexical Link Analysis Application: Improving Web Service to Acquisition Visibility Portal

**Ying Zhao—**Zhao is a research associate professor at the Naval Postgraduate School. 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—**Dr. Gallup is a research associate professor at the Naval Postgraduate School's Department of Information Sciences, 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—**Dr. MacKinnon is a research associate professor at the Naval Postgraduate School (NPS) and the deputy director of the Distributed Information Systems Experimentation (DISE) research group, since 2007. In 2009, he became involved with data mining research and its effect on knowledge management and defense acquisition. He holds a PhD and an engineer's degree from Stanford University, conducting theoretic and field research in knowledge management (KM), theoretically and empirically exploring how individual learning and forgetting affect organizational project performance. He holds Master of Science degrees in information technology management (ITM) and operations research (OR)—each from the Naval Postgraduate School (NPS). [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, 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 comprehensible picture. This can be a daunting task. We have applied a data-driven automation system and methodology, 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, analysis, and decision making. 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 successfully discovered statistically significant correlations, and automatically extracted lexical links, thus improving acquisition professionals' knowledge of their data. This might have otherwise required expensive manpower to perform. We also developed LLA into a web service via several use cases for large-scale LLA applications. In this paper, we show how to apply the LLA web service to the Acquisition Visibility Portal, which is a critical tool to provide the DoD-wide acquisition community with authoritative and accurate data services. The resulting methodology could reduce the workload of decision-makers and achieve improved purchasing decisions, serving to improve the long-term success of acquisition strategies.

## Introduction

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 (MDAPs), Acquisition Category II (ACATII), and so forth. Program Elements (PEs) 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 experienced acquisition professionals.

On a conceptual level, our first question is as follows: How can the information that emerges from the acquisition process be used to produce overall awareness of the fit between programs, projects, systems, and of the needs for which they were intended?

In precise terms, we observed that there were three important processes that seemed fundamentally disconnected. Specifically, 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), as shown in Figure 1. Yet, these were not analyzed and compared together in a dynamic, holistic methodology that could keep pace with changes and reflect patterns of relationships. In the past three years, we employed the lexical link analysis (LLA) automation methodology to analyze the data in three areas, illustrated in Figure 1.

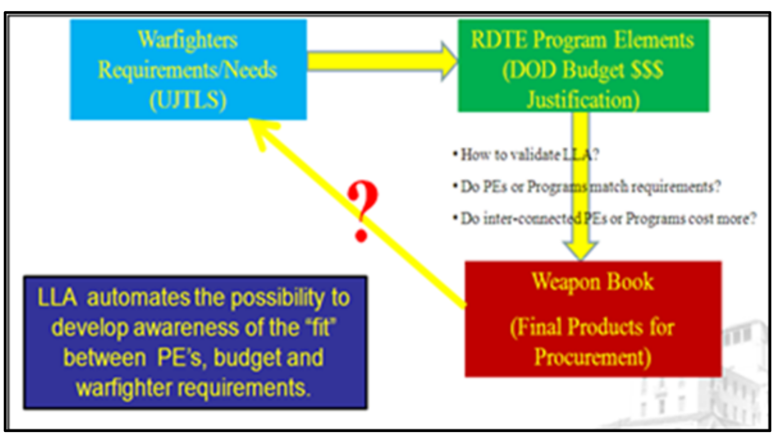


**Figure 1. Determining Business Processes Links From Requirements to DoD Budget Justification to Final Products**

In the past, we have explored how analytic and visualization tools such as LLA can help detect data inconsistency and gaps (bad data; see Research Status section). We have further systematically improved our understanding of the quality of the data by automatically discovering new patterns that were previously unknown, and identified data dependencies that might be indicators for program or investment performances. However, much more work is needed in this area as well as continued in-depth analysis performed at the different levels of the Acquisition Visibility Portal (AVP). AVP is a critical tool that provides the DoD-wide acquisition community with authoritative and accurate data services via interfaces to DTIC and DAMIR for programs (e.g., MDAPs, ACATIIs) with milestones, costs, schedules and performance data, Selected Acquisition Reports (SARs), and Acquisition Strategy Reports (ASRs), among others.

We seek to show how LLA can be adapted to the AVP's ongoing requirements and continuous improvement of DoD data quality and decision-making.

## Methodology

### *Overview of Lexical Link Analysis*

As in military operations, where the term *situational awareness* was coined, we note that our efforts can inform awareness of analyzed data in a unique way that helps improve a decision-maker's understanding or awareness of its content. We therefore define *awareness* as the cognitive interface between decision-makers and a complex system, expressed in a range of terms or "features," or specific vocabulary or "lexicon," to describe the attributes and surrounding environment of the system. Specifically, LLA is a form of text mining in which word meanings represented in lexical terms (e.g., word pairs) can be represented as if they are in a community of a word network.

Link analysis "discovers" and displays a network of word pairs. These word pair networks are characterized by one-, two-, or three-word themes. The *weight* of each theme is determined by its frequency of occurrence. Figure 2 shows a visualization of common lexical links shared between Systems 1 and 2, shown in the red box. Unlinked, outer vectors (outside the red box) indicate unique system features. For example, Figure 3 shows the information from three categories that can be compared, and Figure 4 shows the information from two time periods that can be compared.

Each node, or word hub, represents a system *feature*, and each color refers to the collection of lexicon links (features) that describes a concept or theme. The overlapping area nodes are *lexical links*. What is unique here is that LLA constructs these linkages via intelligent agent technology using social network grouping methods.

The closeness of the systems in comparison can be visually examined or examined using the Quadratic Assignment Procedure (QAP; Hubert & Schultz, 1976; e.g., in UCINET; Borgatti, Everett, & Freeman, 2002) to compute the correlation and analyze the structural differences in the two systems, as shown in Figure 5.

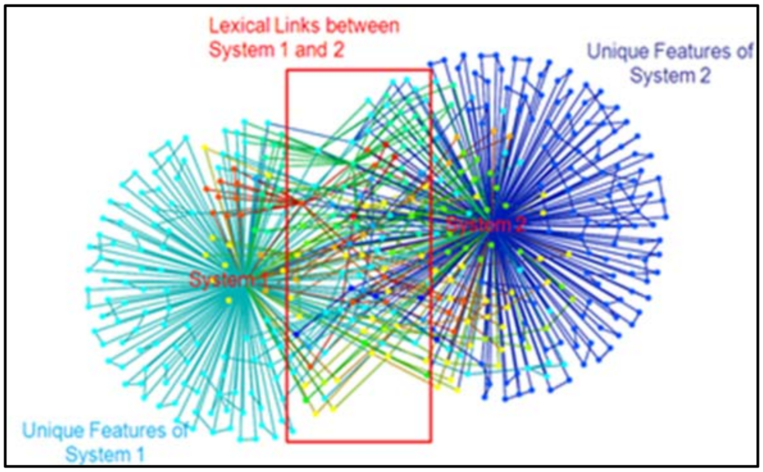


**Figure 2. Comparing Two Systems Using LLA**

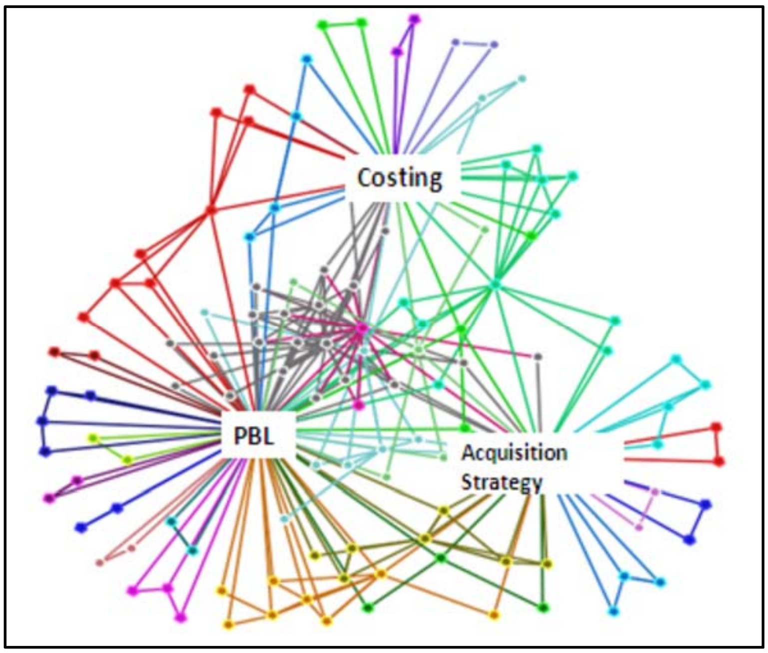


**Figure 3. Comparing Three Categories Using LLA**

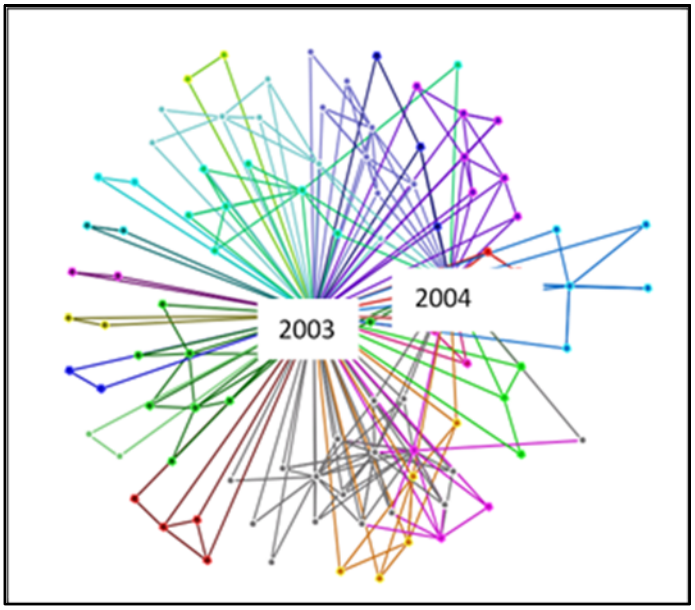


**Figure 4. Comparing Two Time Periods**

QAP Correlations

| | | 1<br>lla_n | 2<br>lla_n | 3<br>lla_n | 4<br>lla_n | 5<br>lla_n | 6<br>lla_n | 7<br>lla_n | 8<br>lla_n |
|---|---|---|---|---|---|---|---|---|---|
| 1 | lla_network_1_2010-AcquisitionStrategy | 1.000 | 0.174 | 0.156 | 0.155 | 0.036 | 0.111 | 0.020 | 0.062 |
| 2 | lla_network_1_2003-AcquisitionStrategy | 0.174 | 1.000 | 0.447 | 0.149 | 0.052 | 0.119 | 0.043 | 0.089 |
| 3 | lla_network_1_2004-AcquisitionStrategy | 0.156 | 0.447 | 1.000 | 0.111 | 0.047 | 0.119 | 0.051 | 0.080 |
| 4 | lla_network_1_2005-AcquisitionStrategy | 0.155 | 0.149 | 0.111 | 1.000 | 0.156 | 0.084 | 0.034 | 0.088 |
| 5 | lla_network_1_2006-AcquisitionStrategy | 0.036 | 0.052 | 0.047 | 0.156 | 1.000 | 0.067 | 0.036 | 0.056 |
| 6 | lla_network_1_2007-AcquisitionStrategy | 0.111 | 0.119 | 0.119 | 0.084 | 0.067 | 1.000 | 0.097 | 0.123 |
| 7 | lla_network_1_2008-AcquisitionStrategy | 0.020 | 0.043 | 0.051 | 0.034 | 0.036 | 0.097 | 1.000 | 0.286 |
| 8 | lla_network_1_2009-AcquisitionStrategy | 0.062 | 0.089 | 0.080 | 0.088 | 0.056 | 0.123 | 0.286 | 1.000 |

QAP P-Values

| | | 1<br>lla_n | 2<br>lla_n | 3<br>lla_n | 4<br>lla_n | 5<br>lla_n | 6<br>lla_n | 7<br>lla_n | 8<br>lla_n |
|---|---|---|---|---|---|---|---|---|---|
| 1 | lla_network_1_2010-AcquisitionStrategy | 0.000 | 0.020 | 0.020 | 0.020 | 0.020 | 0.020 | 0.020 | 0.020 |
| 2 | lla_network_1_2003-AcquisitionStrategy | 0.020 | 0.000 | 0.020 | 0.020 | 0.020 | 0.020 | 0.020 | 0.020 |
| 3 | lla_network_1_2004-AcquisitionStrategy | 0.020 | 0.020 | 0.000 | 0.020 | 0.020 | 0.020 | 0.020 | 0.020 |
| 4 | lla_network_1_2005-AcquisitionStrategy | 0.020 | 0.020 | 0.020 | 0.000 | 0.020 | 0.020 | 0.020 | 0.020 |
| 5 | lla_network_1_2006-AcquisitionStrategy | 0.020 | 0.020 | 0.020 | 0.020 | 0.000 | 0.020 | 0.020 | 0.020 |
| 6 | lla_network_1_2007-AcquisitionStrategy | 0.020 | 0.020 | 0.020 | 0.020 | 0.020 | 0.000 | 0.020 | 0.020 |
| 7 | lla_network_1_2008-AcquisitionStrategy | 0.020 | 0.020 | 0.020 | 0.020 | 0.020 | 0.020 | 0.000 | 0.020 |
| 8 | lla_network_1_2009-AcquisitionStrategy | 0.020 | 0.020 | 0.020 | 0.020 | 0.020 | 0.020 | 0.020 | 0.000 |

QAP statistics saved as datafile QAP Correlation Results

**Figure 5. QAP Correlation via UCINET**

Figure 6 shows a visualization of LLA with connected keywords or concepts as groups or themes. Words are linked as word pairs that appear next to each other in the original documents. Different colors indicate different clusters of word groups. They were produced using a link analysis method—a social network grouping method (Girvan et al., 2001) where words are connected, as shown in a single color, as if they are in a social community. A "hub" is formed around a word centered or connected with a list of other words ("fan-out" words) centered on other hub words. For instance, Figure 7 shows a detailed view of a theme or word group in Figure 6: the words "analysis, research, approach" are connected and centered around other related words. We use three words such as "analysis, research, approach" to label a group.

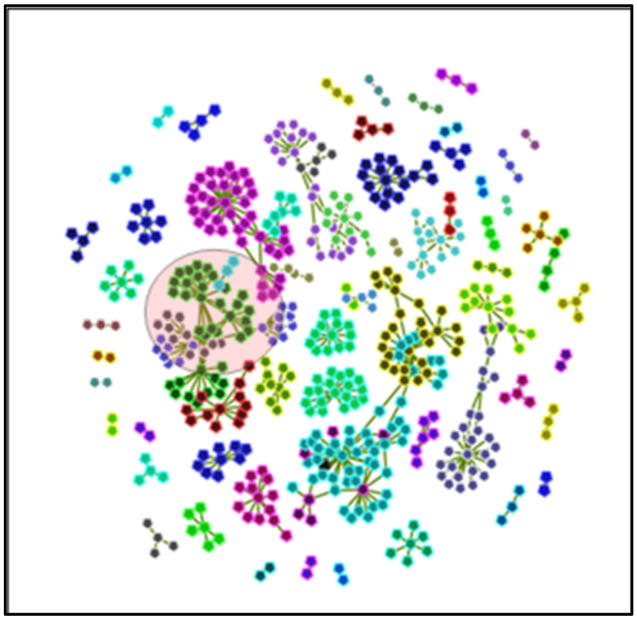

**Figure 6. Word and Term of Themes Discovered and Shown in Colored Groups**

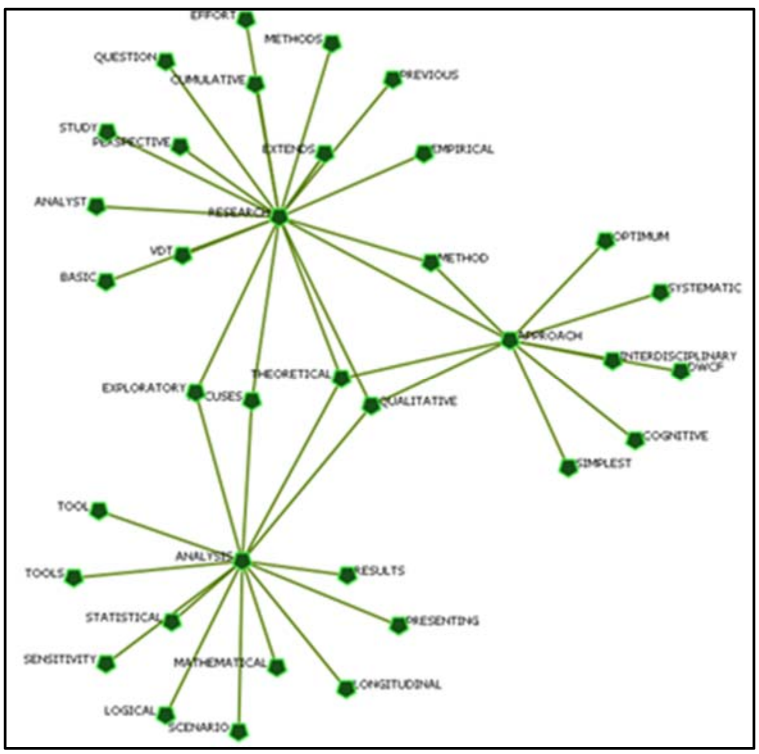


**Figure 7. A Detailed View of a Theme or Word Group From Figure 6**

The detailed steps of LLA processing include applying collaborative learning agents (CLAs) and generating visualizations, including a lexical network visualization via AutoMap (2009), radar visualization, and matrix visualization (Zhao, Gallup, & MacKinnon, 2010). The following are the steps for performing an LLA:

- Read each set of documents.

- Select feature-like word pairs.
- Apply a social network community finding algorithm (e.g., Newman grouping method; Girvan et al., 2001) to group the word pairs into themes. A theme includes a collection of lexical word pairs connected to each other.
- Compute a “weight” for a theme for the information of a time period, that is, how many word pairs belong to a theme for that time period and for all time periods.
- Sort theme weights by time, and study the distributions of these themes by time.

### *Business Problems That LLA Addresses*

General areas that LLA usually informs are the following:

- Discovering themes and topics in the unstructured documents and sorting the importance of the themes
- Discovering social and semantic networks of organizations that were involved, comparing the two networks to obtain insights to answer the following questions:
    - Demonstrating what were the organizations involved in the *important* themes
    - Illustrating how semantic networks might suggest improved potential collaboration when compared to social networks

### *Social and Semantic Networks Analysis*

Current research of social network analysis mostly focuses on people or organizations of direct associations, regardless of the contents linked. The so-called study of centrality (Girvan, 2002; Feldman, 2007) has been a focal point for the social network structure study. Finding the *centrality* of a network lends insight into the various roles and groupings such as the connectors (e.g., mavens, leaders, bridges, isolated nodes), the clusters (and who is in them), the network core, and its periphery. We have been working toward two areas of innovations in the network analysis:

- Extracting social networks based on the entity extraction
- Extracting semantic networks based on the contents and word pairs using LLA
- Applying characteristics and centrality measures from the semantic networks and social networks to predict latent properties such as emerging leadership, for example, emerging techniques that might dominate, in the social networks. These characteristics are further categorized into themes and time-lined trends for informed prediction of future events.

### *Implementation Details*

In the past year, we continued our efforts at the Naval Postgraduate School (NPS) by using CLAs (Quantum Intelligence [QI], 2009) and expanded to other tools, including AutoMap (Center for Computational Analysis of Social and Organizational Systems [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.

We have been generating visualizations including a lexical network visualization using various open source tools. We began by using the Organizational Risk Assessment (ORA; CASOS, 2009) tool and expanded to other tools. For example, in the past year, we developed 3-D network views using Pajek (2011) and X3D (2011). We also developed our visualizations Radar view and Match view (Zhao et al., 2010).

LLA uses a computer-based learning agent called Collaborative Learning Agents (CLA; QI, 2009) to employ an unsupervised learning process that separates patterns and anomalies. CLA is a computer-based learning agent, or agent collaboration, capable of ingesting and processing data sources, leveraged via an educational license with Quantum Intelligence, Inc. The unsupervised agent learning is implemented by indexing each set of documents separately and in parallel using multiple learning agents. Multiple agents can work collaboratively and in parallel. We set up a cluster utilizing Linux servers in the NPS High Performance Computing Center (HPC) to handle the large-scale data and secure environment in the NPS Secure Technology Battle Laboratory (STBL).

### *Relations to Other Methods*

The LLA approach is more properly related to Latent Semantic Analysis (LSA; Dumais, Furnas, Landauer, Deerwester, & Harshman, 1988) and Probabilistic Latent Semantic Analysis (PLSA). In the LSA approach, a term-document matrix is the starting point for analysis. The elements of the term-document or feature-object (term as feature and document as object) matrix are the occurrences of each word in a particular document, that is, $A = [a_{ij}]$, where $a_{ij}$ denotes the frequency in which term *j* occurs in document *i.* The term-document matrix is usually sparse. LSA uses singular value decomposition (SVD) to reduce the dimensionality of the term-document matrix. SVD cannot be applied to the cases where the vocabulary (the unique number of terms) in the document collection is large. LSA has been widely used to improve information indexing, search/retrieval, and text categorization.

A recent development related to this method is called latent Dirichlet allocation (LDA; Blei, Ng, & Jordan, 2003), which is a generative probabilistic model of a corpus. In LDA, a document is considered to be composed of a collection of words—a “bag of words,” where word order and grammar are not considered important. The basic idea is that documents are represented as random mixtures over latent topics, where each topic is characterized by a statistical distribution (Dirichlet distribution) over the corpus. Our theme generation from LLA is different than LDA, in which a collection of lexical terms are connected to each other semantically, as if they are in a social community, and social network grouping methods are used to group the words, and unlike LSA, our method is easily scaled to analyze a large vocabulary and is generalizable to any sequential data.

### *Anticipated Benefits*

Our LLA method provides the solutions to meet the critical needs of the acquisition research. The key advantage is to provide an innovative near real-time self-awareness system to transfer diversified data services into strategic decision-making knowledge, specifically through the following:

- Automation: High correlation of LLA results—with the link analysis done by human analysts—makes it possible to save human power and improve responsiveness. Automation is achieved via computer program or software *agents* to perform LLA frequently, and in near real-time.

- Discovery: LLA “discovers” and displays a network of word pairs. These word pair networks are characterized by one, two, or three word themes. The weight of each theme is determined based on its frequency of occurrence. It may also discover blind spots of human analysis that are caused by the overwhelming data for human analysts to consider.
- Validation: LLA may provide different perspectives of links. In the acquisition context, links discovered by human analysts may emphasize component and part connections that do not necessarily reflect content overlaps. LLA looks for the overlapping of contents to help identify improved affordability and improved response to meeting warfighter requirements, and achieve better acquisition decisions. Consequently, it can provide improved results in terms of trust and quality of association discovery and can help break through the taxonomy of ignorance (Denby & Gammack, 1999) and organizational boundaries, and help improve organizational reach.

## Research Status

### *Acquisition Visibility Portal Background*

Our goal is to demonstrate the LLA web service for assisting the DoD-wide effort of integrating and maintaining authoritative and accurate acquisition data services in both legacy and new platforms. Specifically, we wanted to analyze the data sources from the Acquisition Visibility Portal (https://portal.acq.osd.mil) by examining consistency, correlation, and gaps among categories of information for each individual program listed in the portal.

One of the biggest risk factors in defense acquisition is the unanticipated effects of program interactions. For example, ASD(SE) and Dahmann worked toward identifying interdependence among programs within a system of systems (SoS). Yet, more broadly, and as a result of required joint capabilities, portfolios often include program interdependencies and system-of-systems effects. Ultimately, the current “program-centric” acquisition paradigm is increasingly ill-suited to identify and address program risks that arise outside of program boundaries. LLA can help isolate these issues from the body of information collected, which have yet to be effectively identified.

To begin to address this risk, we observed that very little of the information generated for program oversight is amenable to effective analysis. Every major acquisition program’s milestone review generates volumes of information, which the OSD staff is supposed to review to determine if the program is properly prepared for the next milestone. Although they are beginning to compile these artifacts centrally to facilitate review and analysis, at present, the only way to analyze the information in these artifacts is to read them. With limitations on staffing, little time is available to thoroughly review these artifacts. Moreover, each functional community is required to review only the particular document for which it is responsible. For example, the systems engineering community typically only examines the Systems Engineering Plan (SEP), the test and evaluation community looks only at the Test & Evaluation Master Plan (TEMP), and the acquisition community looks at the Acquisition Strategy Report (ASR). Rarely do any of these stakeholders review multiple reports or jointly discuss them to determine if they are mutually consistent and to consider inconsistencies that might indicate programmatic risk. There is even less incentive and opportunity to look for external factors that would potentially invalidate the assumptions that underpin the basic cost, schedule, and performance targets of each program execution.

Overlaying the concept maps for each of the major categories of artifacts to conduct a pair-wise comparison might expose significant disconnects between them. We are motivated by a situation in which the SEP identifies a critical dependency between the program and an external system, but the TEMP doesn't have a corresponding reference to testing that interdependency. Therefore, it may be productive to compare the acquisition strategy to the SEP or TEMP.

***Results***

LLA maps of these artifacts from one category to another, for example, the SEP at Milestone B, are significantly different from the SEP at Milestone C that might indicate a reduction in system functionality resulting from cost increases elsewhere. These maps, reported as themes, concepts, and word pairs, may help cue a decision-maker's attention to the potential issues and help the decision-maker consider specific and productive directions for further scrutiny.

To develop comprehensive LLA maps, we first extracted a sample from a representative MDAP from the Acquisition Visibility Portal (AVP) with categories of information to demonstrate the method, as follows:

- SEP: 2 documents, 222 pages
- TEMP: 5 documents, 62 pages
- ASR: 11 documents including metrics, 634 pages
- SARs: 9 documents, 313 pages
- DAES: 19 documents, 447 pages
- Milestone B 2366b Certification Acquisition Decision Memorandum (ADM) 12 documents, 105 pages
- APB: 3 documents, 39 pages
- TRA: 1 document, 1 page

Figure 8 lists the top 20 themes discovered for comparing data for ASR and SEP with the highest correlations. In Row 2, there are 299 word pairs for the two sources together classified in Theme 117(E); 47 of them appear in both sources, indicating potential feature overlaps. The *correlation* is the ratio=47/299=0.157 which indicates 15.7% of the features represented as word pairs shared in both artifacts. As a detail shown in Figure 9, part of 299 word pairs in Theme 117(E) are visualized in red, yellow, and green links, representing the shared word pairs, unique ones to ASR and SEP, respectively.

| 1 | Theme Id | All Sources | _ASR | _SEP | Overlap | Correlation |
|---|---|---|---|---|---|---|
| 2 | 117(E) | 299 | 201 | 51 | 47 | 0.157 |
| 3 | 347(P) | 481 | 346 | 67 | 68 | 0.141 |
| 4 | 395(P) | 500 | 330 | 102 | 68 | 0.136 |
| 5 | 130(P) | 590 | 428 | 89 | 73 | 0.124 |
| 6 | 281(P) | 469 | 372 | 42 | 55 | 0.117 |
| 7 | 210(P) | 570 | 400 | 105 | 65 | 0.114 |
| 8 | 298(P) | 599 | 348 | 184 | 67 | 0.112 |
| 9 | 388(P) | 508 | 381 | 73 | 54 | 0.106 |
| 10 | 263(P) | 666 | 517 | 79 | 70 | 0.105 |
| 11 | 368(P) | 669 | 472 | 127 | 70 | 0.105 |
| 12 | 330(P) | 546 | 391 | 99 | 56 | 0.103 |
| 13 | 147(E) | 234 | 181 | 29 | 24 | 0.103 |
| 14 | 224(E) | 331 | 236 | 62 | 33 | 0.100 |
| 15 | 144(P) | 490 | 350 | 92 | 48 | 0.098 |
| 16 | 270(P) | 502 | 371 | 82 | 49 | 0.098 |
| 17 | 235(E) | 431 | 329 | 60 | 42 | 0.097 |
| 18 | 245(E) | 281 | 215 | 39 | 27 | 0.096 |
| 19 | 113(E) | 334 | 245 | 57 | 32 | 0.096 |
| 20 | 310(P) | 586 | 441 | 90 | 55 | 0.094 |
| 21 | 182(A) | 197 | 157 | 22 | 18 | 0.091 |

**Figure 8. Themes for Comparing SEP and ASR, Sorted According to Correlation Ascending**

Figure 9 shows that there are concepts related to these word nodes that appear uniquely to the ASR or SEP.

Since the SEP document is supposed to support the ASR, the illustrations and visualizations of it might inform acquisition professionals about why concepts in the SEP were missing from the ASR, and vice versa.

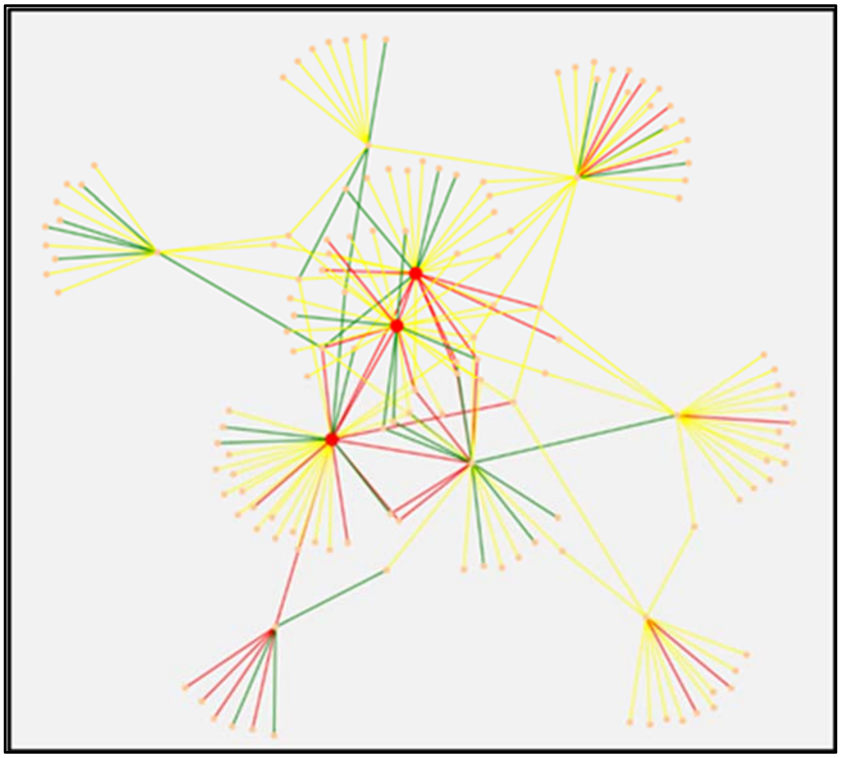

**Figure 9. Detail of Word Pairs for Theme 117(E): Red Links for Shared Word Pairs for SEP and ASR (Yellow Links for Unique Word Pairs Unique to ASR, and Green links for Unique Word Pairs Unique to SEP)**

Figure 10 lists the least correlated themes discovered for comparing data for ASR and SEP. In Row 2, there are 149 word pairs for the two sources together, classified in Theme 359(E)(A); four of them appear in both sources (overlap). The correlation is the ratio=4/149=0.027. A detail shown in Figure 9, part of 149 word pairs in Theme 359(A) are visualized in red, yellow, and green links, representing the shared word pairs, unique ones to the ASR and SEP, respectively.

| 1 | Theme Id | All Sources | ASR | SEP | Overlap | Correlation |
|---|---|---|---|---|---|---|
| 2 | 359(A) | 149 | 127 | 18 | 4 | 0.027 |
| 3 | 390(A) | 173 | 150 | 18 | 5 | 0.029 |
| 4 | 419(A) | 95 | 73 | 18 | 4 | 0.042 |
| 5 | 267(A) | 149 | 123 | 19 | 7 | 0.047 |
| 6 | 238(A) | 170 | 121 | 41 | 8 | 0.047 |
| 7 | 293(A) | 231 | 184 | 36 | 11 | 0.048 |
| 8 | 76(E) | 249 | 208 | 28 | 13 | 0.052 |
| 9 | 408(E) | 419 | 376 | 21 | 22 | 0.053 |
| 10 | 287(A) | 223 | 187 | 24 | 12 | 0.054 |
| 11 | 203(E) | 259 | 170 | 75 | 14 | 0.054 |
| 12 | 334(E) | 276 | 218 | 43 | 15 | 0.054 |
| 13 | 135(E) | 271 | 218 | 38 | 15 | 0.055 |
| 14 | 104(A) | 196 | 163 | 22 | 11 | 0.056 |
| 15 | 63(E) | 314 | 253 | 43 | 18 | 0.057 |
| 16 | 373(P) | 480 | 403 | 49 | 28 | 0.058 |
| 17 | 372(P) | 608 | 509 | 62 | 37 | 0.061 |
| 18 | 389(A) | 155 | 137 | 8 | 10 | 0.065 |
| 19 | 331(E) | 383 | 246 | 112 | 25 | 0.065 |
| 20 | 205(P) | 561 | 420 | 104 | 37 | 0.066 |
| 21 | 127(P) | 490 | 414 | 43 | 33 | 0.067 |

**Figure 10. Themes for Comparing SEP and ASR, Sorted According to Descending Correlation**

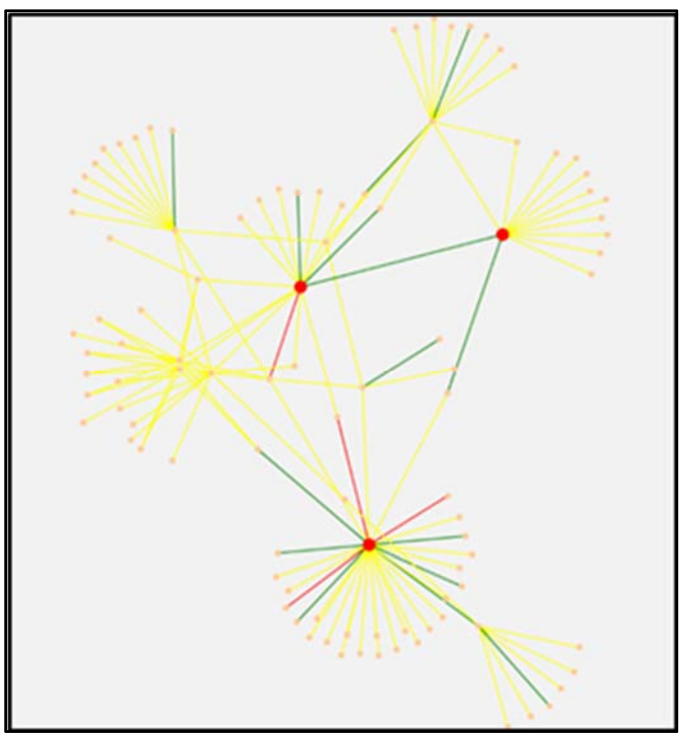

**Figure 11. Detail of Word Pairs for Theme 359(A): Red Links for Shared Word Pairs for SEP and ASR (Yellow Links for Unique Word Pairs Unique to ASR, and Green Links for Unique Word Pairs Unique to SEP)**

In Figure 11, there are also concepts that are more prevalent in the ASR than in the SEP. The ASR includes other concepts that are not in the SEP that might be important.

LLA also categorizes themes into popular (P), emerging (E), and anomalous (A). Comparing Figure 8 and Figure 10, one can see that popular themes tend to have higher correlations among data sources (ASR and SEP), while anomalous themes tend to have lower correlations among data sources.

For each pair of comparisons for two categories of information, we use the ratio of the number of word pairs that appear in both categories and the total number of word pairs as an overall correlation for each pair.

In Table 1, the highlighted cells are the ones with correlation > 0.06. The categories DAES, SARs, and SEP have higher overall correlations with other ones. The most correlated two categories are SARs and DAES (correlation = 0.117). The category TEMP has the lowest overall correlations with other categories. Although TEMP and SEP were both produced in the test and evaluation community, the correlation between the two is low (0.027).

**Table 1. LLA Correlations Between Categories of Information**

| | APB | ASR | 2366B_Cert | DAES | SARs | SEP | TEMP | TRA |
|---|---|---|---|---|---|---|---|---|
| **APB** | 1.000 | 0.007 | 0.027 | 0.022 | 0.080 | 0.014 | 0.010 | 0.005 |
| **ASR** | 0.007 | 1.000 | 0.015 | 0.048 | 0.025 | 0.075 | 0.028 | 0.001 |
| **2366B_Cert** | 0.027 | 0.015 | 1.000 | 0.026 | 0.038 | 0.026 | 0.018 | 0.068 |
| **DAES** | 0.022 | 0.048 | 0.026 | 1.000 | 0.117 | 0.073 | 0.023 | 0.003 |
| **SARs** | 0.080 | 0.025 | 0.038 | 0.117 | 1.000 | 0.044 | 0.020 | 0.004 |
| **SEP** | 0.014 | 0.075 | 0.026 | 0.073 | 0.044 | 1.000 | 0.027 | 0.003 |
| **TEMP** | 0.010 | 0.028 | 0.018 | 0.023 | 0.020 | 0.027 | 1.000 | 0.002 |
| **TRA** | 0.005 | 0.001 | 0.068 | 0.003 | 0.004 | 0.003 | 0.002 | 1.000 |

When discussing the findings with the domain expert, it seems the correlation is surprisingly low for DAES and SARs. DAES and SARs reports are similar in context and content (both relate to acquisition performance); they would be expected to have higher correlation. Further investigations, such as the following: are needed to see what might be the causes for the low correlation:

- To investigate if significantly different content appears in the two types of reports; for example, DAES reports may include more details than SARs reports.
- To differentiate the SAR and DAES reports by year and compute the correlations over time, to see when the significant discrepancies, that is, the drop in the correlation, came into the picture.
- To correlate the DAES or SAR reports over time separately to see if the correlation increases and decreases might have to do with the new features being introduced into the program, and therefore correlate to the significance of low or high changes found in LLA with the numeric metrics such as cost, schedule, funding, and performance.

## Future Work

Since this is the first program to have undergone a relatively comprehensive LLA analysis using multiple types of acquisition documents, the findings cannot be evaluated in terms of being “good” or “bad,” “normal” or “unusual,” and so forth. Therefore, future investigation should consider the following additional studies:

- Analyze additional programs in the AVP, compute the correlation matrices like Table 1, and compare the results to determine if the correlation patterns are similar or different.
- Discuss the findings in detail with the domain experts and personnel associated with the programs to see if the correlation patterns have significance, as follows:
    - if the correlation are the indications for data quality issues and
    - if the correlation patterns have impacts for the costs, schedules, funding, and performance of the programs.

## Conclusion

In this paper, we demonstrated how to apply LLA to generate maps of the acquisition artifacts among multiple categories of data. These maps, reported as themes, concepts, and word pairs, may help identify the issues and offer specific and productive directions for further examination as to why there are gaps among the categories of information.

## References


Blei, D., Ng, A., & Jordan, M. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research, 3*, 993–1022. Retrieved from http://jmlr.csail.mit.edu/papers/volume3/blei03a/blei03a.pdf

Borgatti, S. P., Everett, M. G., & Freeman, L. C. (2002). *Ucinet for Windows: Software for social network analysis.* Harvard, MA: Analytic Technologies.

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

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/Defense Systems.

Denby, E., & Gammack, J. (1999). *Modelling ignorance levels in knowledge-based decision support*. Retrieved from http://wawisr01.uwa.edu.au/1999/DenbyGammack.pdf

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).

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.

Girvan, M., & Newman, M. E. J. (2002, June). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, *USA, 99*(12), 7821–7826.

Hubert, L., & Schultz, J. (1976). Quadratic assignment as a general data-analysis strategy. *British Journal of Mathematical and Statistical Psychology, 29*, 190–241.

Pajek. (2011). Retrieved from http://vlado.fmf.uni-lj.si/pub/networks/pajek

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

X3D. (2011). Retrieved from http://www.web3d.org

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

Zhao, Y., Gallup, S. P., & MacKinnon, D. J. (2011a). A web service implementation for large-scale automation, visualization, and real-time program-awareness via lexical link analysis. In *Proceedings of the Eighth Annual Acquisition Research Program.* Monterey, CA: Naval Postgraduate School.

Zhao, Y., Gallup, S. P., & MacKinnon, D. J. (2011b). Lexical link analysis for the Haiti earthquake relief operation using open data source. In *Proceedings of the Sixth ICCRTS, International Command and Control, Research and Technology Symposium*. Québec City, Canada.

Zhao, Y., Gallup, S. P., & MacKinnon, D. J. (2011c). *Towards real-time program awareness via lexical link analysis* (Acquisition Research Sponsored Report Series, NPS-AM-10-174). Monterey, CA: Naval Postgraduate School.

Zhao, Y., Gallup, S. P., & MacKinnon, D. J. (2011d, September). System self-awareness and related methods for improving the use and understanding of data within DoD. *Software Quality Professional, 13*(4), 19–31. Retrieved from http://asq.org/pub/sqp

Zhao, Y., Gallup, S. P., & MacKinnon, D. J. (2012, May). *Applications of lexical link analysis web service for large-scale automation, validation, discovery, visualization, and real-time program-awareness.* Presentation at the Ninth Annual Acquisition Research Symposium, Monterey, CA.

## Acknowledgements

We thank Mr. Robert Flowe from AT&L/ARA/FP&OS, OSD, who provided sponsorship to access the Acquisition Visibility Portal, and who offered relevant questions along with insightful discussions.

ACQUISITION RESEARCH PROGRAM
GRADUATE SCHOOL OF BUSINESS & PUBLIC POLICY
NAVAL POSTGRADUATE SCHOOL
555 DYER ROAD, INGERSOLL HALL
MONTEREY, CA 93943

www.acquisitionresearch.net