NPS-IS-22-009

NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

STRUCTURED AND UNSTRUCTURED DATA SCIENCES AND BUSINESS INTELLIGENCE FOR ANALYZING REQUIREMENTS POST MORTEM

by

Ying Zhao

December 2022

Distribution Statement A: Approved for public release. Distribution unlimited.

Prepared for: N8 - Integration of Capabilities & Resources. This research is supported by funding from the Naval Postgraduate School, Naval Research Program (PE 0605853N/2098). NRP Project ID: NPS-22-N332-A

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED	
		START DATE	**END DATE**
02 Dec 2022	Technical Report	02 Jan 2022	31 Dec 2022

4. TITLE AND SUBTITLE
Structured and Unstructured Data Sciences and Business Intelligence for Analyzing Requirements Post Mortem

5a. CONTRACT NUMBER	5b. GRANT NUMBER	5c. PROGRAM ELEMENT NUMBER
		0605853N/2098
5d. PROJECT NUMBER	**5e. TASK NUMBER**	**5f. WORK UNIT NUMBER**
NPS-22-N332-A; W2223		

6. AUTHOR(S)
Dr. Ying Zhao

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
Naval Postgraduate School 1 University Circle Monterey, CA 93943	NPS-IS-22-009

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)	11. SPONSOR/MONITOR'S REPORT NUMBER(S)
Naval Postgraduate School, Naval Research Program, Monterey, CA Integration of Capabilities & Resources (N8)	NRP, N8	NPS-IS-22-009; NPS-22-N332-A

12. DISTRIBUTION/AVAILABILITY STATEMENT
Distribution Statement A: Approved for public release. Distribution unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The US Navy systems may have unexpected significant cost growth for many reasons. The Office of the Chief of Naval Operations (OPNAV) manually and periodically reviews big data (structured and unstructured data) that were created within the Department of Defense requirements process to identify the programs that create excessive cost or cost growth. This research explores two questions:

1. What are the common elements of requirements that create excessive cost growth in Navy systems?
2. Assuming the elements are identified, what is the risk (likelihood and magnitude) of cost growth from common elements for both procurement and sustainment costs?

We applied classic data sciences and business intelligence tools towards a more advanced artificial general intelligence framework to analyze structured and unstructured data and identify elements and factors that create excessive cost growth. We found patterns and deep causes for high cost or cost growth programs using lexical link analysis, natural language processing (NLP) tools, a semantic network analyzer, anomaly detection, and causal learning concepts. Programs with anomalous characteristics can lead to high costs or high growth. These tools provide counterfactual and drill-down discovery of the key words that explain the deep causes of cost growth. The recommendations are to apply these tools for the total benefits of analyzing Navy programs and requirements of post mortem data, towards modernizing the OPNAV's Program Budget Information System (PBIS) to become a knowledge system that can effectively learn from historical data to make better risk predictions and decisions for the future Program Objectives Memorandum (POM).

15. SUBJECT TERMS
lexical link analysis, LLA, named entity extraction, NEE, parts of speech tagging, POS, spaCy, semantic network analysis, SNA, centrality measures, unsupervised machine learning, transformers, Program Objectives Memorandum, POM, Program Budget Information System, PBIS

16. SECURITY CLASSIFICATION OF: U			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES
a. REPORT	**b. ABSTRACT**	**C. THIS PAGE**	UU	47
U	U	U		

19a. NAME OF RESPONSIBLE PERSON	19b. PHONE NUMBER *(Include area code)*
Ying Zhao	831.656.3789

STANDARD FORM 298 (REV. 5/2020)
Prescribed by ANSI Std. Z39.18

THIS PAGE INTENTIONALLY LEFT BLANK

NAVAL POSTGRADUATE SCHOOL
Monterey, California 93943-5000

Ann E. Rondeau President	Scott Gartner Provost

The report entitled "**Structured and Unstructured Data Sciences and Business Intelligence for Analyzing Requirements Post Mortem**" was prepared for Integration of Capabilities & Resources (N8) and funded by Naval Postgraduate School, Naval Research Program (NRP), (PE 0605853N/2098).

Distribution Statement A: Approved for public release. Distribution unlimited.

This report was prepared by:

ZHAO.YING.1384570476 Digitally signed by ZHAO.YING.1384570476 Date: 2022.12.14 10:24:10 -08'00'

Ying Zhao
Research Professor, Information Sciences

Reviewed by:	**Released by:**
COOK.GLENN.ROBERT.1040837619 Digitally signed by COOK.GLENN.ROBERT.1040837619 Date: 2022.12.14 10:16:02 -08'00'	SMITH.KEVIN.B.1230026366 Digitally signed by SMITH.KEVIN.B.1230026366 Date: 2022.12.16 17:43:32 -08'00'
Alex Bordetsky, Chairman Information Sciences Department	Kevin B. Smith Vice Provost for Research

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

The US Navy systems may have unexpected significant cost growth for many reasons. The Office of the Chief of Naval Operations (OPNAV) manually and periodically reviews big data (structured and unstructured data) that were created within the Department of Defense requirements process to identify the programs that create excessive cost or cost growth. This research explores two questions:

1: What are the common elements of requirements that create excessive cost growth in Navy systems?

2: Assuming the elements are identified, what is the risk (likelihood and magnitude) of cost growth from common elements for both procurement and sustainment costs?

We applied classic data sciences and business intelligence tools towards a more advanced artificial general intelligence framework to analyze structured and unstructured data and identify elements and factors that create excessive cost growth. We found patterns and deep causes for high cost or cost growth programs using lexical link analysis, natural language processing (NLP) tools, a semantic network analyzer, anomaly detection, and causal learning concepts. Programs with anomalous characteristics can lead to high costs or high growth. These tools provide counterfactual and drill-down discovery of the key words that explain the deep causes of cost growth. The recommendations are to apply these tools for the total benefits of analyzing Navy programs and requirements of post mortem data, towards modernizing the OPNAV's Program Budget Information System (PBIS) to become a knowledge system that can effectively learn from historical data to make better risk predictions and decisions for the future Program Objectives Memorandum (POM).

I. INTRODUCTION

The US Navy's Office of the Chief of Naval Operations (OPNAV) is charged, among other responsibilities, with executing the Planning, Programming, Budgeting, and Execution (PPBE) process through a series of concurrent annual planning cycles guided by a Program Objectives Memorandum (POM), collectively referred to as POM-Year X (C. Marsh, email to author, November 4, 2022).

Navy systems may have unexpected significant cost growth for many reasons. The US Navy's OPNAV is charged, among other responsibilities, with executing the planning, programming, budgeting, and execution (PPBE) process through a series of concurrent annual planning cycles guided by a Program Objectives Memorandum (POM), collectively referred to as POM-Year X (C. Marsh, email to author, November 4, 2022).

The objective is to leverage advanced analytics to help the OPNAV understand the common elements and causes of existing Navy systems that have significant cost growth from historical data, requirements documents, and open-source media.

The research questions are:

1: What are common elements of requirements that create excessive cost growth in Navy systems?

2: Assuming the elements are identified, determine the risk (likelihood and magnitude) of cost growth from common elements for both procurement and sustainment costs?

The PBIS has been modernized as an authoritative knowledge system including historical data of planned and executed POM information and spending each year. Data relevant to PBIS include structured data and unstructured data. For example, structured data include number of platforms procured and procurement and sustainment costs for Navy systems. Budget Exhibits (BE) contain PPBE information as well as unstructured data of unclassified high-level program descriptions and their elements. Initial capability documents (ICDs), key performance parameters (KPPs), or key-systems attributes (KSAs) from capability development documents (CDDs) and operational requirements documents (ORDs) are classified data sources from previous requirements processes that may have contributed to excessive cost growth. These data can be structured, such as KPPs and KSAs, and unstructured, such as BEs, ICDs, and CDDs.

We applied two categories of methods: 1. classic data sciences and business intelligence tools and 2. an artificial general intelligence framework to address the needs and research questions to analyze structured and unstructured data together and correlate them with excessive cost or cost growth of Navy systems. Specifically, we applied LLA, a semantic network analyzer, anomaly detection, and causal learning to discover patterns and deep causes that can lead to high cost or cost growth.

We analyzed two unclassified data sets provided by the topic sponsors. The first data set included seven PE documents that are processed using the LLA, artificial general intelligence NLP named entity extraction (NEE) and parts of speech (POS) tagging tools. POS features include extracted noun and verb word features. NEE features include extracted person, organization, location, product, money, event, law, language, date, time, percent, ordinal, cardinal, quantity, nationality or religious group, infrastructure, and work of art.

To discover the anomalous characteristics, we first applied LLA to compute the similarity of every two pairs of programs, then applied community finding and centrality calculation algorithms to discover the programs that are far away from community centers or on the edges of the semantic networks, which are indicators of anomalies. We used a semantic network analyzer to visualize that these Navy systems located in the center or edge of the semantic networks. The number of links are also indicators of system independences represented in the word feature networks discovered by LLA. Less linked BEs are anomalous via the unsupervised learning because they may have more unique features or innovations. We also used LLA's drill-down search capability and counterfactual reasoning of causal inferences to narrow down the key words as potential causes for the anomalous characteristics.

Some data and meta-data for the project are in the secret level. We documented the methodology and demonstrated the approaches using a subset of unclassified data downloaded from public domains, i.e., Budget Exhibits (BE), in this report. The deliverables are also based on the unclassified data.

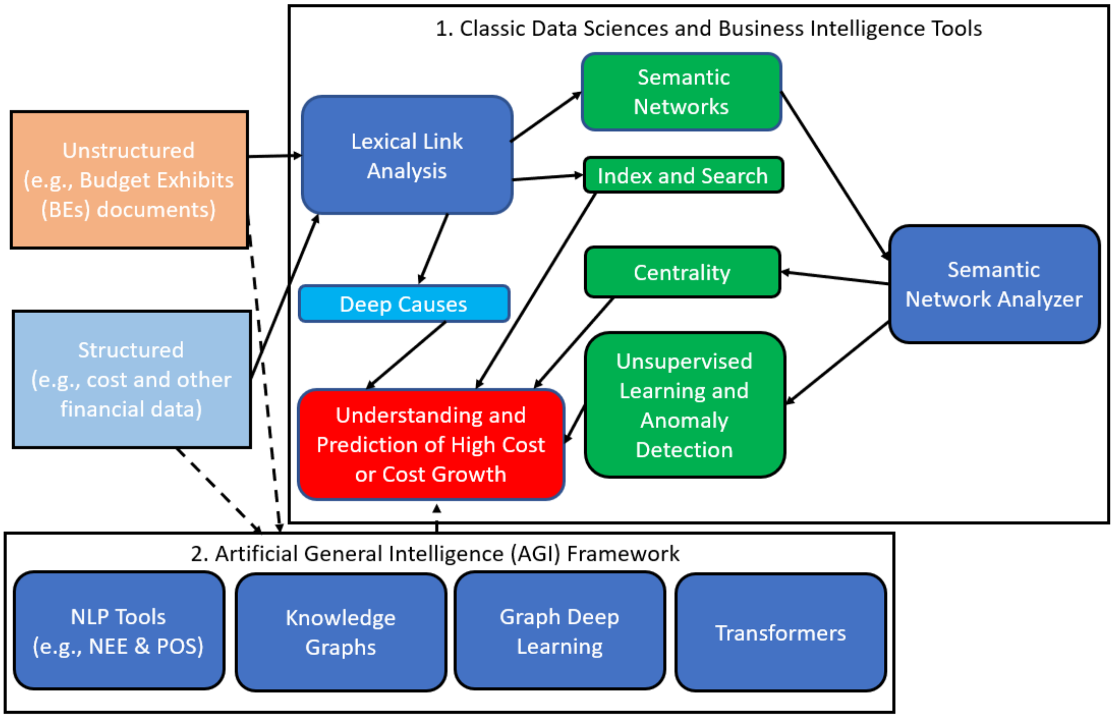


Figure 1. Two categories of methods, i.e., 1. classic data sciences and business intelligence tools, and 2. artificial general intelligence (AGI) framework, are the technical concepts for the project.

II. APPROACHES

Figure 1 shows that centered around understanding and prediction high cost or cost growth for Navy systems, the author considers two categories of methods, i.e., 1. classic data sciences and business intelligence tools, and 2. artificial general intelligence (AGI) framework to address the needs. The classic data sciences and business intelligence tools are the focus of the paper. The author first reviews each method and element in the following sections.

A. STRUCTURED AND UNSTRUCTURED DATA

Program Budget Information System (PBIS) has been modernized as an authoritative knowledge system including historical data of planned and executed POM information and spending each year. Data relevant to PBIS include structured data and unstructured data. For example, structured data include number of platforms procured, procurement and sustainment costs for Navy systems. Program elements or Budget Exhibits (BEs) contain PPBE information as well as unstructured data of unclassified high-level descriptions of the programs and their elements. Data from Initial Capabilities Documents (ICDs) and CDDs structured data attributes of Key Performance Parameters (KPP), or Key-Systems Attributes (KSA) from CDDs, which are mostly classified, may have contributed to cost growth. Some requirements documents (ICDs) are unclassified, although none of the pilot programs.

B. LEXICAL LINK ANALYSIS (LLA)

LLA is a data-driven text and data mining method. In an LLA, a complex system can be expressed in a list of attributes or features with specific vocabularies or lexicon terms to describe its characteristics. LLA is data-driven text analysis. For example, word pairs or bi-grams as lexical terms can be extracted and learned from a document repository. LLA automatically discovers word features, links, and groups and displays them as networks. Nodes are words and bi-grams are the links between words. Bi-gram also allows LLA to be extended to numerical or categorical data. This allows the study of the numeric metrics and structured data attributes such as Key Performance Parameters

(KPP), or Key-Systems Attributes (KSA) integrated with the word features and characteristics of capability requirements linked to the cost growth.

LLA is related to but significantly different from bag-of-words (BOW) methods, Latent Semantic Analysis (LSA, Dumais, Furnas, Landauer, & Deerwester, 1988; Probabilistic Latent Semantic Analysis (PLSA, Hofmann, 1999), WordNet (Miller, 1995), Automap (CASOS, 2009), and Latent Dirichlet Allocation (LDA) (Blei, Ng, & Jordan, 2003).

C. SEMANTIC NETWORKS, SEMANTIC NETWORK ANALYZER, CENTRALITY, UNSUPERVISED LEARNING, AND ANOMALY DETECTION

LLA outputs semantic networks. It divides word node features into three categories by applying network community finding algorithms:

- Authoritative or popular (P) themes: These themes resemble eigenvalue centrality measures in network sciences. These represent the main topics in a data set.
- Emerging (E) themes: These themes tend to become popular or authoritative over time.
- Anomalous (A) themes: These themes may not seem to belong to the data domain as compared to others. They are interesting and could be high-value for further investigation.

Community detection algorithms have been illustrated by Newman in terms of a quality function as the "modularity" measure for a community (cluster) and optimized using a dendrogram-like greedy algorithm (Newman, 2003) as if word features or objects (e.g., programs) in a social community. In a network theory, the most connected nodes, i.e., nodes with higher measures of centrality, are typically considered the most important nodes (Newman, 2006). However, the uniqueness of LLA is that it extracts emerging and anomalous information (word features) which might be more interesting for anomaly detection such as detecting programs with excessive cost growth, and then rank programs with significant cost growth. For example, in the context of the proposed research, emerging and anomalous word features in the capability requirement data might correspond to the innovativeness and uniqueness of a capability requirement. This relates to unsupervised learning algorithms such as K-means, Principal Component Analysis (PCA), and spectral clustering (Ng, Jordan, & Weiss, 2002) for anomaly detection in classic data sciences. Bi-gram also allows LLA to be extended to structured data (Zhao & Zhou, 2014), where a word is an attribute combined with its possible values. LLA automatically discovers word feature networks of social and semantic for extremely large number of word features, scalable to the data attributes and their possible values, similar to the Generative Pre-trained Transformer 3 (GPT-3) model (Brown, 2020), which can handle about 175 billion word features.

Related research questions are listed as follows:

- How does the cost growth correlate with the popular, emerging, and anomalous categories and common elements of the requirement data?
- Does the cost growth correlate with the innovativeness of the requirements?

LLA can be jointly used with NEE and PoS methods (Section 1.6) to address the following questions:

- Do the numbers of people and organizations detected in the requirements correlate with the cost growth?
- Do the number of verbs (actions) and nouns (concepts) detected in the requirement data correlate with the cost growth?
- Do the subsystem independences represented in the word feature networks correlate with the cost growth?

D. CAUSAL LEARNING AND DEEP CAUSES

Anomaly detection often needs to understand causes behind any anomalous behaviors such as excessive cost and/or cost growth (observable effects). This calls a systematic approach of causal machine learning. The key factors for causal learning include the three layers of a causal hierarchy - association, intervention, and counterfactuals (Pearl, 2018; Pearl, & Mackenzie, 2018). A typical causal machine learning method needs to select a cause (C) that maximizes the counterfactual difference $P(E|C) - P(E|Not\ C)$, where the effect E is observable data and cause C is actionable and controllable variable, which might be hidden inside big data (structured or unstructured). If causal learning can reason and detect the causes for good or bad effects, decision makers might be able to fix the causes, avoid bad effects, and achieve desired effects. Interventions are often tested as causes since they are actionable and their effects can be measured. LLA allows a causality analysis. LLA uses causal learning and computes counterfactual proportion difference, *i.e.,*

$$\boldsymbol{cf} = [P(E|C) - P(E|Not\ C)] \times (pooled\ sample\ size) \qquad (1)$$

as the strength of the link of two word feature nodes, where $P(E|C)$ is the probability of event E if event C occurs. The pooled sample size is an average number of historical event E and C occur together normalized by the priors. ***cf*** is a z-score (PSU, 2021) and we use ***cf*** *>1.96* for *p-value < 0.05* as the statistical significance for the link strength of the nodes. With the computation, the network nodes are linked causally.

E. INDEX AND SEARCH

LLA is used to index and search for structured and unstructured data sources implemented in a set of collaborative learning agents (CLAs). For a single CLA, it first indexes and data-mines the data and allows search and retrieve data based on causal knowledge patterns discovered from data. The key difference is that LLA search and a typical search engine is that it can address the question of sorting and ranking important and interesting information based on the different needs. Traditionally in knowledge graph analysis (e.g., semantic networks), the importance of a network node is a form of high-value information. Among various centrality measures, sorting and ranking information based on authority is compared with page ranking of a typical search engine. Current automated methods such as graph-based ranking used in PageRank, require established hyperlinks, citation networks, social networks (e.g., Facebook), or other forms of crowd-sourced collective intelligence. However, these methods are not applicable to situations where there exist no pre-established relationships among network nodes such as intelligence analysis. This makes the traditional centrality measures or PageRank-like methods difficult to apply. Furthermore, current methods mainly score popular information that are important for marketing applications, however, emerging and anomalous information are important for discovering anomalies, e.g., for intelligence data analysis. Patterned, emerging, and anomalous themes in the LLA search is used to sort and rank important information based on the needs of different applications.

F. ARTIFICIAL GENERAL INTELLIGENCE (AGI) FRAMEWORK - NATURAL LANGUAGE PROCESSING (NLP)

An AGI framework typically contains large-scale machine learning models with billions of parameters to learn and recognize patterns from multimodality of data such as imagery, text, geospatial information, video, acoustics, radio frequencies, and time series.

In an AGI framework, natural language processing (NLP) of text analysis include indexing/search, topics and theme extraction, summarization, categorization, sentiment analysis, entity extraction (e.g., people and locations), and sorting/ranking importance of topics and themes. The tool spaCy and prodigy (Explosion, 2016, 2021) are used for many of these analyses. For example, Air Force uses the combination for monitoring AI and Autonomy research: they are using spaCy to track public AI/autonomy papers, patents, compare them with the internal air force project descriptions. The system is called Landscapes for Autonomy using the tool prodigy. Orange (UOL, 2021) has a text mining package including sentiment analysis. Some text analysis tools are supervised machine learning, some are unsupervised machine learning methods. However, if one wants specific automation to extract keywords related to "fundamental understanding" and "utility," it may be difficult to categorize automatically for the semantic categories and need at least some data with manual labels.

Named Entity Extraction (NEE) (Explosion, 2016; NIST, 2022; Stanford NLP, 2019) and Parts of Speech (PoS) tagger (Toutanova K. & Manning, C., 2000; Explosion, 2016; NIST, 2022) are the techniques used as pre-processing tools. An entity can be a person, organization, location, money, and dates, etc. The tool can also extract PoS such as nouns and verbs which are important to the application in this paper.

G. TRANSFORMERS

An AGI framework typically includes a category of algorithms so-called Transformers. AGI Transformers include deep neural network models and contain large number of parameters (e.g., billions, Generative Pre-trained Transformer (GPT) Neo (Eleuther.ai, 2022; OpenAI, 2022) or Bidirectional Encoder Representations from Transformers (BERT, Devlin, Chang, Lee, & Toutanova, 2018), pre-trained from big data (e.g., the entire internet), can use much less data (few-shots) and better understand and make sense unstructured data. Fine-tuning GPT Neo or BERT can adapt the models to the domain specific data such as exercise logs, intelligence analysis and reports, and Navy systems and programs data.

H. KNOWLEDGE GRAPHS AND GRAPH DEEP LEARNING

In recent years, knowledge graphs (Turing Institute, 2022) revive as knowledge databases that use graph-structured data models or topologies to integrate data can store interlinked descriptions of entities – objects, events, situations or abstract concepts (Wiki, 2022). The generalization of AGI Transformers to knowledge and graph domain is termed Geometric neural network (GNN) or Graph deep learning (GDL) (Bronstein, 2021). Learning from knowledge graphs can model the broad class of data that has objects (treated as nodes) with some known relationships (treated as edges). Knowledge graphs represented as knowledge networks and combine structured, unstructured, and multi-modality data via embedding and encoder techniques for nodes and edges (Barp et al. 2022).

III. DATA SETS AND RESULTS

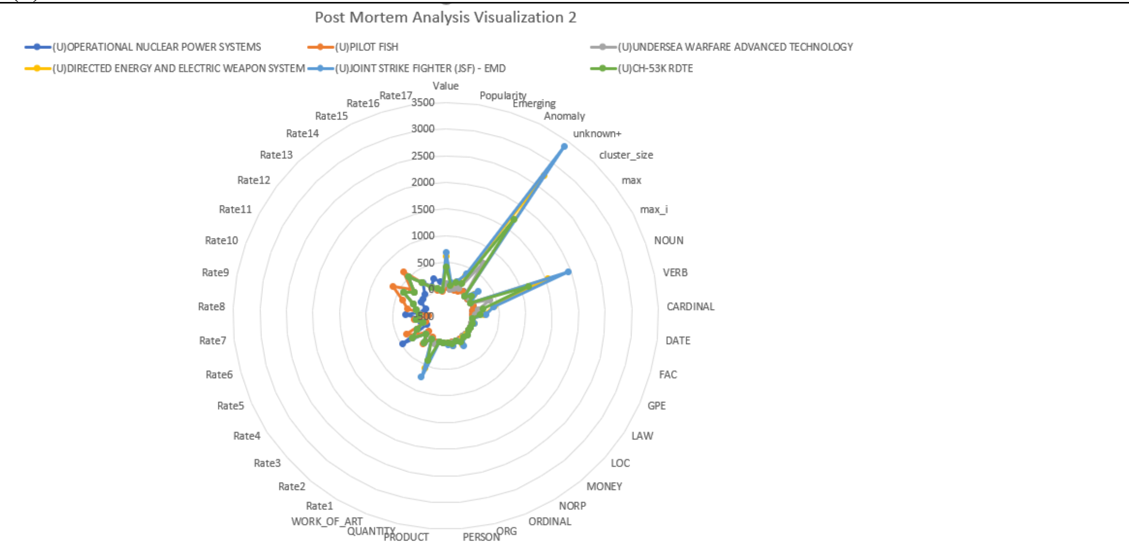
In this section, the author shows two data sets the methods described in Section 1 applied.

A. DATA SET 1

The first data set includes seven budget exhibits documents that are processed using the LLA, AGI NLP NEE & POS (spaCy) tools as shown in Figure 2. Figure 2 (a) show numbers of popularity, emerging, anomaly word features, unknown, and total (value) extracted for the PE documents using LLA. Unknown features are word features do not exist in other programs but uniquely exist in a specific program. POS features include extracted noun and verb word features. NEE features include extracted person, organization (ORG), location (LOC), product, money, event, law, language, date, time, percent, ordinal, cardinal, quantity, nationality or religious group (norp), infrastructure (FAC), work of art. Cost rates are projected for 17 years (Rate1 to Rate17). Figure 2 (b) show a radial graph for the data dimensions from Figure 1 (a). Note that the features extracted do not show deep causes (e.g., key words) for potential high cost growth. In this example, the cost growth does not correlate with the popular, emerging, and anomalous categories of PE documents. Cost growth may correlate with the innovativeness of the programs, there is an example that the number of unknown (e.g. unique) features are correlated with high cost growth, i.e., (U)CH-53K RDTE. The numbers of people and organizations detected in the PE documents do not seem to correlate with the cost growth. The number of nouns (concepts) detected in the data may correlate with the cost growth.

PE_LABEL	Value	Popularity	Emerging	Anomaly	unknown+	cluster_size	max	max_i	NOUN	VERB	CARDINAL	DATE	FAC	GPE	LAW	LOC	MONEY	NORP	ORDINAL	ORG	PERSON	PRODUCT	QUANTITY	WORK_OF_ART	Rate1
(U)OPERATIONAL NUCLEAR POWER SYSTEMS	42	4	12	26	68	7	14	21	49	7	5	0	1	0	0	2	0	7	0	1	1	0	0	19	29.56
(U)PILOT FISH	37	4	10	23	62	7	14	21	47	7	7	0	2	1	0	0	0	5	0	1	1	0	0	29	-29.25
(U)UNDERSEA WARFARE ADVANCED TECHNOLOGY	107	7	32	68	704	7	40	21	371	112	52	0	3	0	0	4	0	33	2	11	1	1	0	62	11.31
(U)DIRECTED ENERGY AND ELECTRIC WEAPON SYSTEM	618	102	182	334	2714	7	173	21	1547	442	214	0	51	0	4	14	2	124	8	75	5	4	0	557	11.31
(U)JOINT STRIKE FIGHTER (JSF) - EMD	682	124	181	377	3380	7	268	21	1955	418	250	0	48	0	2	53	13	147	2	68	36	3	3	723	11.31
(U)CH-53K RDTE	412	82	146	184	1714	7	114	21	1148	212	137	3	14	0	6	50	18	59	5	46	14	0	1	397	11.31
(U)ASW SYSTEMS DEVELOPMENT - MIP	166	9	34	123	878	7	70	21	474	102	54	0	7	0	1	5	3	26	4	10	1	1	2	61	2E+0

(a)



(b)

Figure 2. (a) Seven PE documents that are processed using the LLA, AGI NLP NEE & POS (spaCy) tools: (a) (b) A radial graph for the data dimensions from (a). There is an example that number of nouns and unknown/unique features are correlated with high cost growth, e.g., (U)CH-53K RDTE.

B. DATA SET 2

The second data set includes 14 budget exhibit documents. Figure 3 (a) shows an example where the maximum total program cost, e.g., 5223 million and cost increase 136 percent are used as measures of cost growth for this program and attached to the program folder shown in Figure 3 (b).

Exhibit P-40, Budget Line Item Justification: PB 2024 Navy — **Date:** April 2023

Appropriation / Budget Activity / Budget Sub Activity: 1611N: Shipbuilding and Conversion, Navy / BA 02: Other Warships / BSA 01: Other Warships

P-1 Line Item Number / Title: 2122 / DDG-51

ID Code (A=Service Ready, B=Not Service Ready)**:** A — **Program Elements for Code B Items:** N/A — **Other Related Program Elements:** N/A

Line Item MDAP/MAIS Code: N/A

Resource Summary	Prior Years	FY 2022	FY 2023	FY 2024 Base	FY 2024 OCO	FY 2024 Total	FY 2025	FY 2026	FY 2027	FY 2028	To Complete	Total
Procurement Quantity *(Units in Each)*	87	2	2	2	-	2	2	2	1	1	2	101
Gross/Weapon System Cost *($ in Millions)*	99,459.611	3,930.919	4,417.537	4,364.003	0.000	4,364.003	4,328.523	4,447.255	2,714.061	2,259.750	4,927.728	130,849.387
Less PY Advance Procurement *($ in Millions)*	2,910.850	-	-	-	-	-	-	-	-	-	-	2,910.850
Less Cost To Complete *($ in Millions)*	2,203.070	-	-	-	-	-	-	-	-	-	-	2,203.070
Less Subsequent Year Full Funding *($ in Millions)*	433.000	-	-	-	-	-	-	-	-	-	-	433.000
Less Hurricane *($ in Millions)*	227.100	-	-	-	-	-	-	-	-	-	-	227.100
Less EOQ *($ in Millions)*	1,621.241	254.932	41.000	233.588	-	233.588	232.995	232.990	193.786	-	-	2,810.532
Less Escalation *($ in Millions)*	48.200	-	-	-	-	-	-	-	-	-	-	48.200
Less Transfer *($ in Millions)*	218.500	-	-	-	-	-	-	-	-	-	-	218.500
Net Procurement (P-1) *($ in Millions)*	91,797.650	3,675.987	4,376.537	4,130.415	0.000	4,130.415	4,095.528	4,214.265	2,520.275	2,259.750	4,927.728	121,998.135
Plus Subsequent Year Full Funding *($ in Millions)*	433.000	-	-	-	-	-	-	-	-	-	-	433.000
Full Funding TOA *($ in Millions)*	92,230.650	3,675.987	4,376.537	4,130.415	-	4,130.415	4,095.528	4,214.265	2,520.275	2,259.750	4,927.728	122,431.135
Plus CY Advance Procurement *($ in Millions)*	3,332.434	-	-	-	-	-	-	-	-	-	-	3,332.434
Plus Cost To Complete *($ in Millions)*	1,149.086	45.753	228.577	225.917	-	225.917	114.695	149.446	130.912	158.684	-	2,203.070
Plus EOQ *($ in Millions)*	1,454.589	120.000	618.352	196.007	-	196.007	-	-	-	-	-	2,388.948
Plus Escalation *($ in Millions)*	48.200	-	-	-	-	-	-	-	-	-	-	48.200
Plus Transfer *($ in Millions)*	218.500	-	-	-	-	-	-	-	-	-	-	218.500
Plus Hurricane *($ in Millions)*	227.100	-	-	-	-	-	-	-	-	-	-	227.100
Total Obligation Authority *($ in Millions)*	**98,660.559**	**3,841.740**	**5,223.466**	**4,552.339**	**0.000**	**4,552.339**	**4,210.223**	**4,363.711**	**2,651.187**	**2,418.434**	**4,927.728**	**130,849.387**

(a)

P40_OPN_0946_2024PB_155107_2027_242m_131pct
P40_SCN_2122_2024PB_154748_1_2023_5223m_136pct
P40_WPN_2327_2024PB_154946_2027_272m_54pct
R2_0205601N_2024PB_153316_4_2022_133m_53pct
0204152n_7_pb_2014_1_2m_0pct
U_1507N_PB_2024_1_122m_23pct
U_0204228N_7_PB_2020_1_36m_300pct
U_0603564N_4_PB_2022_2_76m_400pct
U_0604234N_5_PB_2023_1_421m_38pct
U_0604269N_5_PB_2019_2_243m_77pct
U_0604274N_5_PB_2018_1_584m_16pct
U_0604282N_5_PB_2024_1_241m_92pct
U_0604307N_5_PB_2020_1_416m_9.5pct
U_0604454N_4_PB_2020_1_13m_120pct

(b)

Figure 3. (a) An example where the maximum total program cost, e.g., 5223 million and cost increase 136 percent are used as measures of cost growth for a program and attached to the program folder shown in (b).

LLA outputs a match matrix as shown in Figure 4, which include the numbers of word features matched for any two BEs in the data set.

Match Matrix From Lexical Link Analysis

		Match Score	PE2_U_0604234N_5_PB_2023_1_421m_38pct	new_R2_0205601N_2024PB_153316_4_2022_133m_53pct	new_P40_OPN_0946_2024PB_155107_2027_242m_131pct	PE2_U_0604454N_4_PB_2020_1_13m_120pct	Un Sco
1	PE2_U_0604234N_5_PB_2023_1_421m_38pct	246.00	_	133.00	23.00	58.00	617
2	new_R2_0205601N_2024PB_153316_4_2022_133m_53pct	212.00	133.00	_	25.00	58.00	291
3	PE2_U_0604307N_5_PB_2020_1_416m_9.5pct	197.00	127.00	109.00	24.00	59.00	678
4	PE2_U_0604269N_5_PB_2019_2_243m_77pct	195.00	135.00	133.00	22.00	57.00	264
5	PE2_U_0604274N_5_PB_2018_1_584m_16pct	185.00	113.00	113.00	18.00	60.00	117
6	PE2_U_0603564N_4_PB_2022_2_76m_400pct	157.00	108.00	98.00	19.00	63.00	354
7	PE2_U_0604282N_5_PB_2024_1_241m_92pct	156.00	94.00	93.00	18.00	56.00	58.
8	PE2_U_0204228N_7_PB_2020_1_36m_300pct	152.00	101.00	94.00	17.00	56.00	431
9	new_P40_WPN_2327_2024PB_154946_2027_272m_54pct	130.00	28.00	49.00	62.00	19.00	160
10	PE2_0204152n_7_pb_2014_1_2m_0pct	101.00	91.00	63.00	15.00	53.00	50.
11	PE2_U_1507N_PB_2024_1_122m_23pct	99.00	22.00	23.00	57.00	18.00	66.
12	new_P40_SCN_2122_2024PB_154748_1_2023_5223m_136pct	90.00	32.00	31.00	37.00	18.00	267
13	new_P40_OPN_0946_2024PB_155107_2027_242m_131pct	82.00	23.00	25.00	_	16.00	492
14	PE2_U_0604454N_4_PB_2020_1_13m_120pct	78.00	58.00	58.00	16.00	_	68.

Figure 4. A match matrix output from LLA showing the numbers of matched word features every two programs.

Figure 5 shows a semantic network visualization for the data in Figure 4. The nodes represent BEs and edges are the links in Figure 4. More linked BEs which have higher degree centrality locate in the center. Less linked BEs locate outside, which are indicators of anomalies. The number of links are indicators of system independences represented in the word feature networks may correlate with excessive cost or cost growth because less linked BEs locate outside of the network centrality layout are the anomalies via the unsupervised learning.

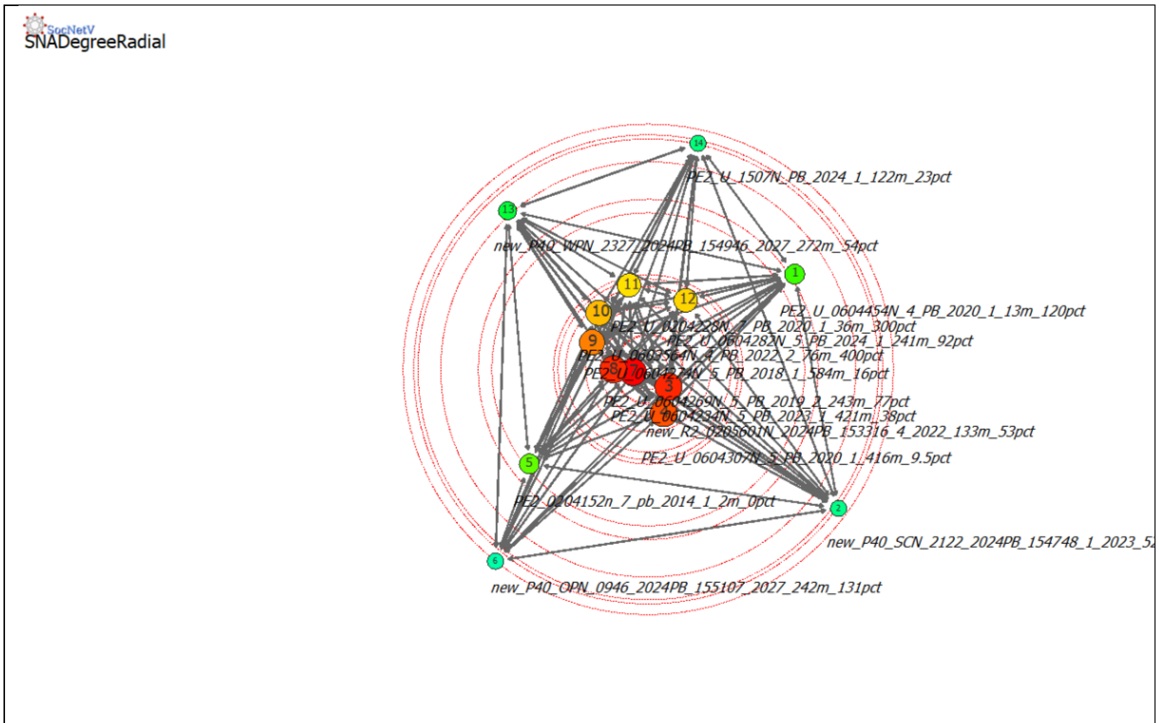


Figure 5. A semantic network visualization linked BEs. More linked BEs which have higher degree centrality locate in the centers. Less linked BEs locate outside, which are indicators of anomalies (Node 2, 6, 13, 14, and 1). The visualization was created using Socnetv (2022).

Figure 6, 7, and 8 show the LLA drill-down searches that are performed for the anomalous BEs in Figure 5. Figure 6 shows that "sole source" only show in "P40_SCN_2122_2024PB_154748_1_2023_5223m_136pct," "U_1507N_PB_2024_1_122m_23pct," and "U_0604307N_5_PB_2020_1_416m_9.5pct," which might be causes for the excessive cost or cost growth.

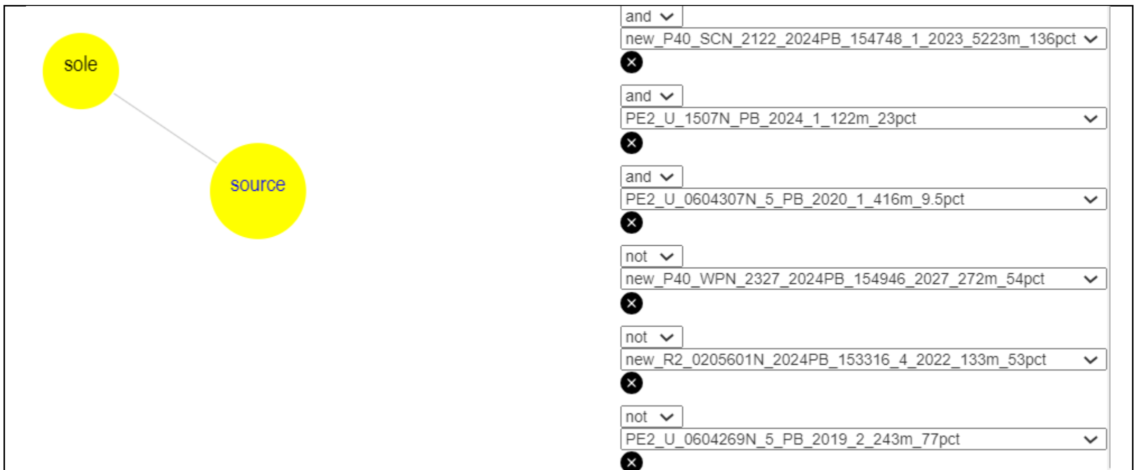


Figure 6. LLA drill-down search. "Sole source" only show in "P40_SCN_2122_2024PB_154748_1_2023_5223m_136pct," "U_1507N_PB_2024_1_122m_23pct," and "U_0604307N_5_PB_2020_1_416m_9.5pct," which might be the causes for the excessive cost or cost growth.

Figure 7 show using LLA search to drill down to word features around "recurring cost," "recurring engineering," "recurring equipment," "recurring procurement," and "recurring swan," which might be causes for the excessive cost or cost growth for anomalous BEs "P40_WPN_2327_2024PB_154946_2027_272m_54pct," "U_1507N_PB_2024_1_122m_23pct," "P40_OPN_0946_2024PB_155107_2027_242m_131pct."

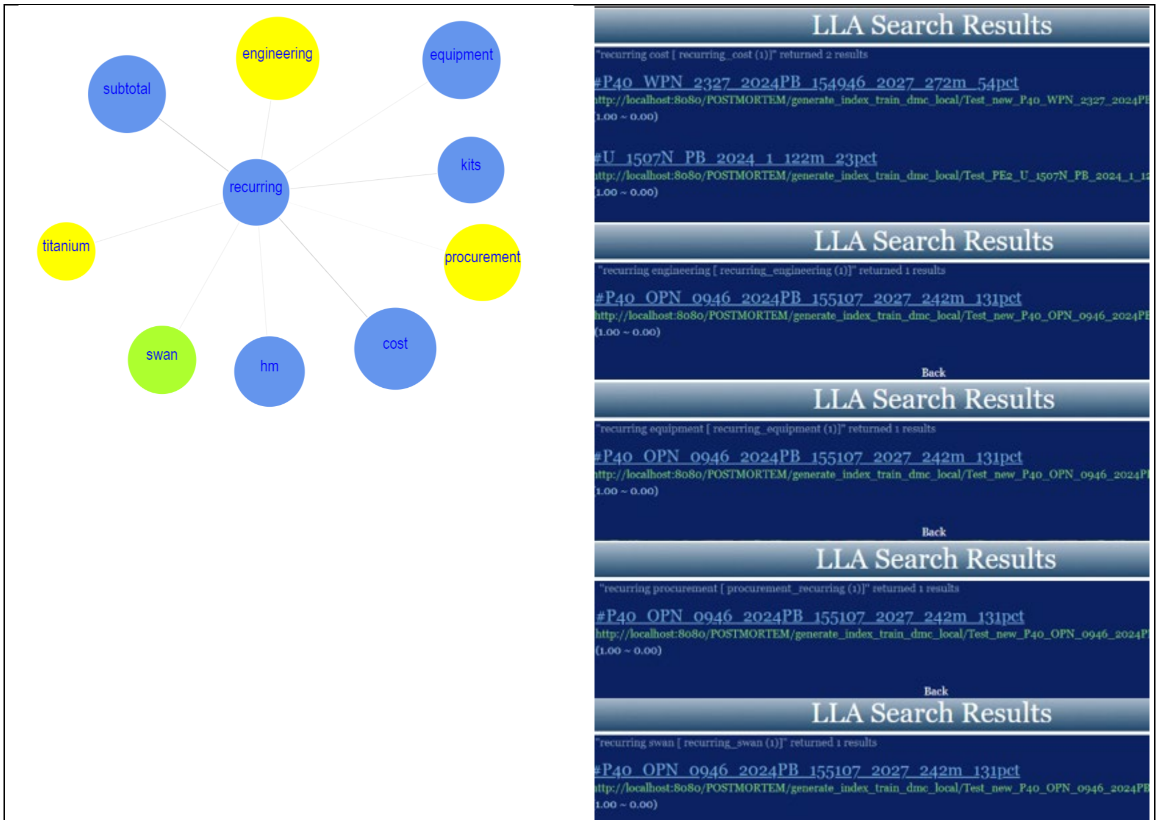


Figure 7. LLA search to drill down to word features around "recurring cost," "recurring engineering," "recurring equipment," "recurring procurement," and "recurring swan," which might be causes for the excessive cost or cost growth for anomalous BEs "P40_WPN_2327_2024PB_154946_2027_272m_54pct," "U_1507N_PB_2024_1_122m_23pct," "P40_OPN_0946_2024PB_155107_2027_242m_131pct."

Figure 8 shows a LLA search to drill down to word features around "unclassified" might be also for some high cost or cost growth programs such as "U_0604269N_5_PB_2019_2_243m_77pct," "U_0603564N_4_PB_2022_2_76m_400pct," "U_0604234N_5_PB_2023_1_421m_38pct," "U_0204228N_7_PB_2020_1_36m_300pct."

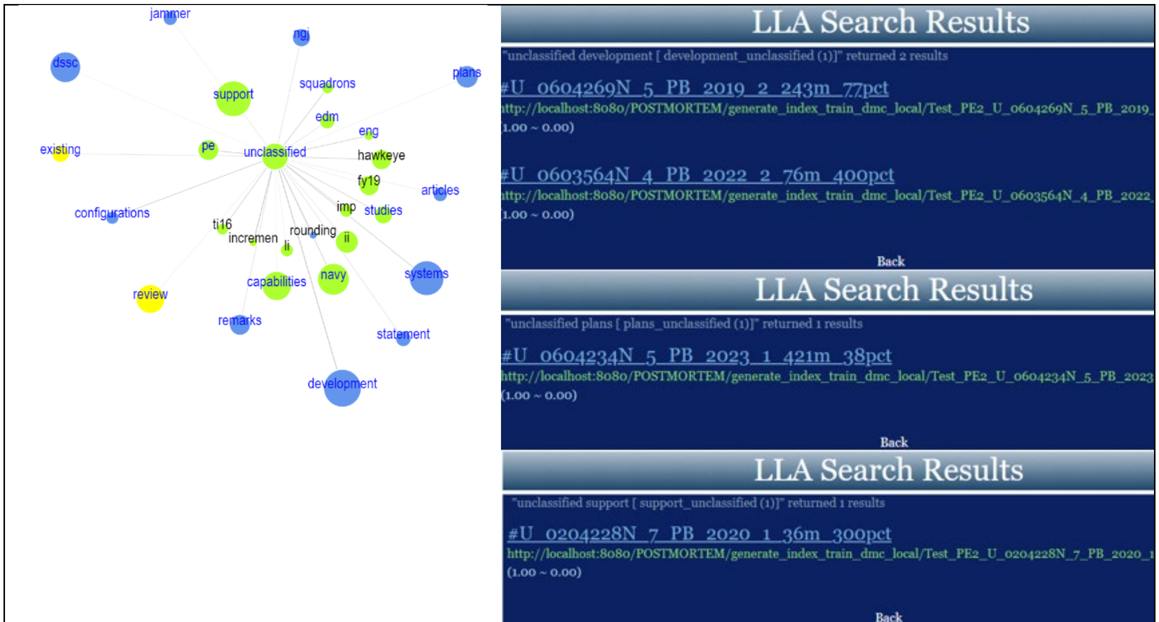


Figure 8. LLA search to drill down (by clicking on the word pair features) to word features around “unclassified” might be also for some high cost or cost growth programs such as “U_0604269N_5_PB_2019_2_243m_77pct,” “U_0603564N_4_PB_2022_2_76m_400pct,” “U_0604234N_5_PB_2023_1_421m_38pct,” “U_0204228N_7_PB_2020_1_36m_300pct.”

In summary, in order to understand and eventually predict high cost or cost growth for evaluating new programs, classic data sciences and business intelligence provide immediate tools to drill down and discover deep causes. The future work to scale the concepts up via the tools in the AGI framework for more accurate prediction, however, deep causes may remain hidden. It is vital to combine the classic data sciences and business intelligence and AGI.

IV. CONCLUSIONS AND RECOMMENDATIONS

In this project, we showed the feasibility to apply the classic data sciences and business intelligence tools and artificial general intelligence (AGI) framework to address the common elements and deep causes of Navy programs and systems that create excessive cost growth. We demonstrated the potential to enable a knowledge system of unstructured and structured data that can effectively learn from historical data and environment and make discovery and prediction. The deliverables include the presentation, demonstration shown to the topic sponsors on November 4, 2022 (Appendix A) and submission a paper proposal/abstract to the 20th Annual Acquisition Research Symposium, May, 2023, Monterey (Appendix B).

- Apply the combined analytic tools explored in this project to the other classified or unclassified, structured and unstructured data sets scale up the combined analytic tools from the OPNAV's Program Budget Information System (PBIS) towards to accurately predict the risk (likelihood and magnitude) of cost growth for future Navy systems.
- Enable the PBIS to become a knowledge system that can effectively learn from human, data, and its surrounding environment to make good assessments and decisions for the future Program Objectives Memorandum (POM).

Appendix A: The presentation and demonstration shown to the topic sponsors on November 4, 2022

Appendix B: The paper proposal/abstract to the 20th Annual Acquisition Research Symposium, May, 2023, Monterey

LIST OF REFERENCES

ARServices (2017). OPNAV POM Process Improvement Observations and Recommendations. In the Operational Effectiveness Initiative.

Barp, A., et al. (2022). Geometric methods for sampling, optimization, inference and adaptive agents. https://arxiv.org/abs/2203.10592

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

Bronstein, M., et al. (2021). Geometric deep learning, grids, groups, graphs, geodesics, and gauges. https://arxiv.org/pdf/2104.13478.pdf

Brown, T., et al. (2020). Language Models are Few-Shot Learners. https://arxiv.org/abs/2005.14165

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

Devlin, J., Chang, M., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. https://arxiv.org/abs/1810.04805

Dumais, S. T., Furnas, G. W., Landauer, T. K. & Deerwester, S. (1988). Using latent semantic analysis to improve information retrieval. In: Proceedings of CHI88: Conference on Human Factors in Computing, 281-285.

Eleuther.ai (2022) https://github.com/EleutherAI/gpt-neo

Explosion (2021). https://prodi.gy/

Explosion (2016). spaCy. https://spacy.io/, https://explosion.ai/

Hofmann, T. (1999). Probabilistic latent semantic analysis. In: Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence, Stockholm, Sweden (1999).

Miller, G. A. (1995). WordNet: a lexical database for English. Communications of the ACM, 38(11), page 39-41.

National Institute of Standards and Technology (NIST) (2022). MUC-7: Named Entity Tasks http://www-nlpir.nist.gov/related_projects/muc/proceedings/muc_7_toc.html#named

Newman, M. (2003). Fast algorithm for detecting community structure in networks http://arxiv.org/pdf/cond-mat/0309508.pdf

Newman, M. (2006). Finding community structure in networks using the eigenvectors of matrices. Phys. Rev. E 74, 036104.

Ng, A., Jordan, M., & Weiss, Y. (2002). On spectral clustering: analysis and an algorithm. In: T. Dietterich, S. Becker, and Z. Ghahramani (Eds.), Advances in Neural Information Processing Systems 14 (pp. 849-856), (2002). MIT Press. http://ai.stanford.edu/~ang/papers/nips01-spectral.pdf

OpenAI (2022). https://openai.com/blog/openai-api/

Pearl, J. (2018). The Seven Pillars of Causal Reasoning with Reflections on Machine Learning. http://ftp.cs.ucla.edu/pub/stat_ser/r481.pdf

Pearl, J. & Mackenzie, D. (2018). The Book of Why: The New Science of Cause and Effect. Basic books.

Penn State University (PSU), (2021). Online Statistics: Normal Approximation Method Formulas. https://online.stat.psu.edu/stat200/lesson/9/9.1/9.1.2/9.1.2.1

Schmidt, E. (2022). Interview. https://www.youtube.com/watch?v=AGNImy8E02w

Stanford NLP (2019). https://nlp.stanford.edu/

Socnetv (2022). Social Network Analysis and Visualization Software. https://socnetv.org/

The National Security Commission on Artificial Intelligence (NSCAI), (2021). The final report. https://www.nscai.gov/2021-final-report/.

Toutanova, K. & Manning, C. (2000). Enriching the knowledge sources used in a maximum entropy part-of-speech tagger. In EMNLP/VLC 1999, pages 63–71.

Turing Institute (2022). Knowledge graphs: How do we encode knowledge to use at scale in open, evolving, decentralized systems? https://www.turing.ac.uk/research/interest-groups/knowledge-graphs

University of Ljubljana (UOL) (2018). Orange: Data Mining Fruitful and Fun. https://orangedatamining.com/

Wiki (2022). Knowledge graph. https://en.wikipedia.org/wiki/Knowledge graph

Zhao, Y., & Zhou, C. (2014). System and method for knowledge pattern search from networked agents (U.S. Patent No. 8,903,756). U.S. Patent and Trademark Office. https://www.google.com/patents/US8903756

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California

3. Research Sponsored Programs Office, Code 41
   Naval Postgraduate School
   Monterey, CA 93943

4. CAPT Christopher Gilmore
   N8 - Integration of Capabilities & Resources
   2000 Navy Pentagon Rm 4D445
   Washington DC 20350

5. Christopher Marsh
   N8 - Integration of Capabilities & Resources
   2000 Navy Pentagon Rm 4D445
   Washington DC 20350

6. Dr. Ying Zhao
   1 University Circle, Room Root 201F
   Monterey, CA 93945

Appendix A

Structured and Unstructured Data Sciences and Business Intelligence for Analyzing Requirements Post Mortem

NPS-22-N332-A

Researcher: Dr. Ying Zhao, Naval Postgraduate School, yzhao@nps.edu

Sponsor: N8 - Integration of Capabilities & Resources

Topic Sponsor POC: Mr. Christopher Marsh , christopher.d.marsh4.ctr@us.navy.mil

11/4/2022

Data Sources

- Program elements
  - PBIS_LI_5 NPS Export.xls

Able to Locate

- 0204152n_7_pb_2014_1_1.90
- U_0204228N_7_PB_2020_1_36.389
- U_0603564N_4_PB_2022_2_75.544
- U_0604234N_5_PB_2023_1_421.001
- U_0604269N_5_PB_2019_2_242.719
- U_0604274N_5_PB_2018_1_584.538
- U_0604282N_5_PB_2024_1_241.472
- U_0604307N_5_PB_2020_1_415.625
- U_0604454N_4_PB_2020_1_12.500
- U_P40_2238_BSA-2_BA-2_APP-1507N_PB_2024_1_121.840

U_0604274N_5_PB_2018_1_584.538

UNCLASSIFIED

Exhibit R-2, RDT&E Budget Item Justification: PB 2020 Navy	Date: March 2019
Appropriation/Budget Activity 1319: *Research, Development, Test & Evaluation, Navy* / BA 5: *System Development & Demonstration (SDD)*	**R-1 Program Element (Number/Name)** PE 0604274N / *Next Generation Jammer (NGJ)*

COST ($ in Millions)	Prior Years	FY 2018	FY 2019	FY 2020 Base	FY 2020 OCO	FY 2020 Total	FY 2021	FY 2022	FY 2023	FY 2024	Cost To Complete	Total Cost
Total Program Element	1,814.232	584.538	449.429	524.261	-	524.261	434.223	178.364	0.000	0.000	0.000	3,985.047
0557: *Next Generation Jammer*	1,814.232	584.538	449.429	524.261	-	524.261	434.223	178.364	0.000	0.000	0.000	3,985.047

Program MDAP/MAIS Code:
Project MDAP/MAIS Code(s): P445

A. Mission Description and Budget Item Justification

The Next Generation Jammer (NGJ) is the next step in the evolution of Airborne Electronic Attack (AEA) and is a critical capability necessary to address current, emerging, and evolving Electronic Warfare gaps, ensure kill chain wholeness against growing threat capabilities and capacity, keep pace with enemy threat weapon systems' advancements, and support the continuous expansion of the AEA mission areas that exceed the capability of currently fielded systems. NGJ will utilize enhanced techniques and tactics to deliver significantly improved radar and communications jamming effectiveness as well as other classified capabilities. Utilizing an Open Systems Architecture that supports software and hardware updates to rapidly counter emergent and evolving threats, NGJ is a key enabler and force multiplier for operations across the spectrum of missions defined in the Defense Strategic Guidance, including strike warfare, projecting power in highly contested environments, and counterinsurgency/irregular warfare. NGJ will also address the shortfalls in scalability, flexibility, supportability, interoperability, availability, and capability of the existing AN/ALQ-99 Tactical Jamming System.

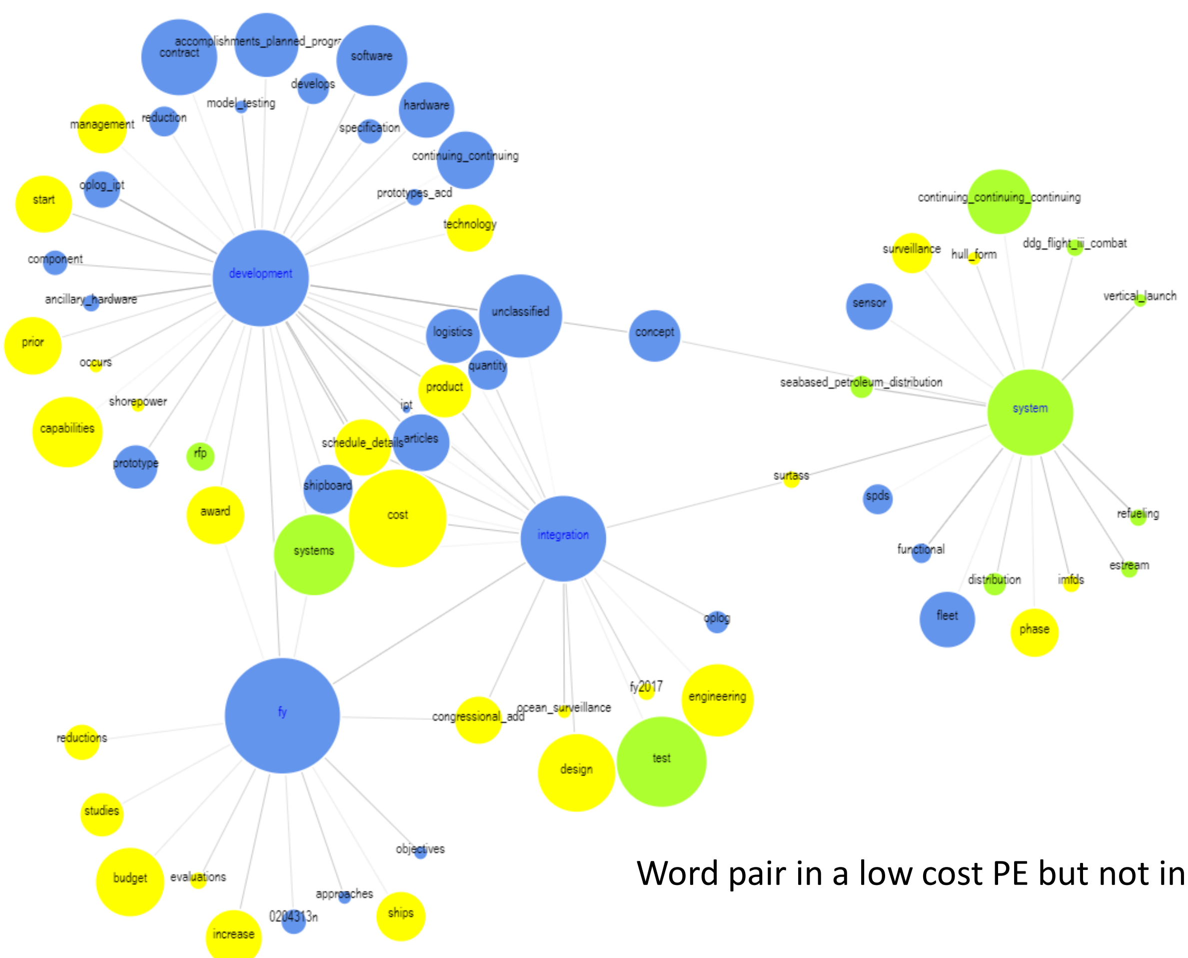
Missing ones

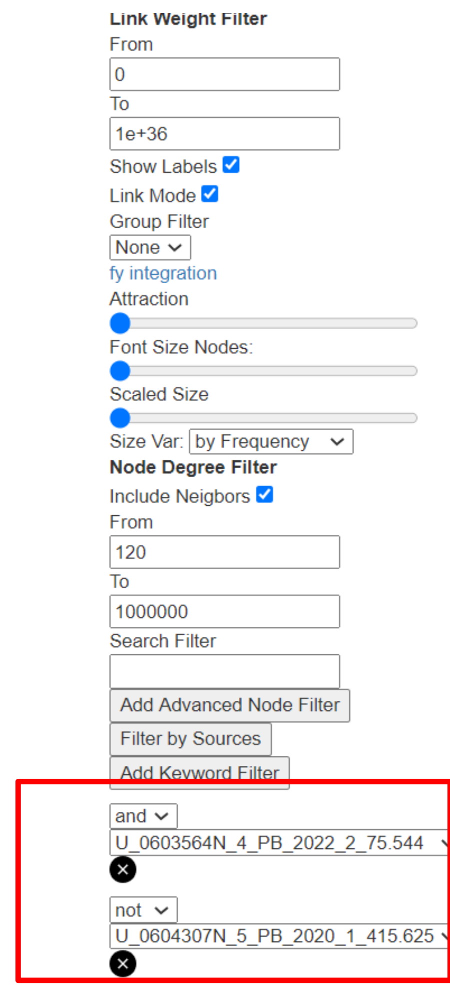
- 0204154N
- 0204162N
- 0204222N
- 0204223N
- 0204269N
- 0204411N
- 0205601N
- 0206138M
- 0502326N
- 0712876N

Methods

- POS and Entity extraction
  – spACY does not seem to reveal the correlations
- Lexical link analysis
  – Drill-down to key words in PEs to correlate with their costs
- Deep learning and knowledge graph to predict risk

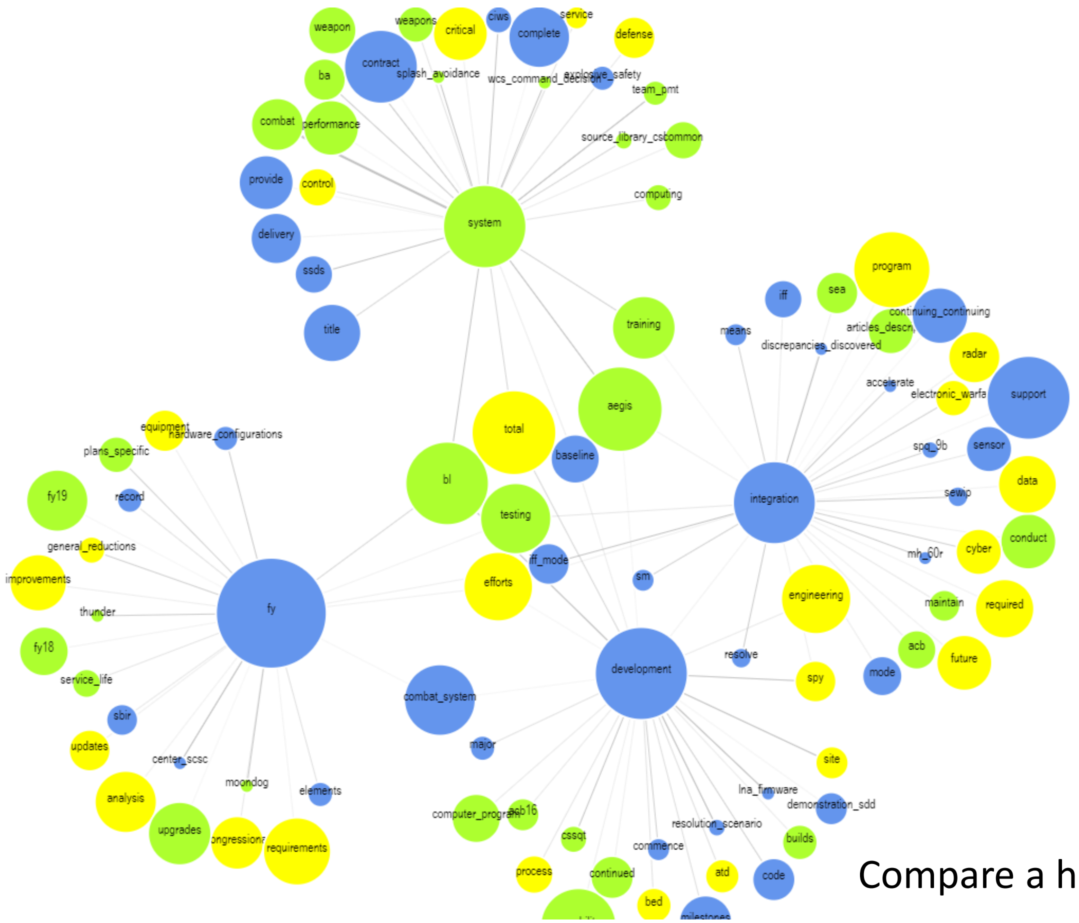
Lexical Link Analysis



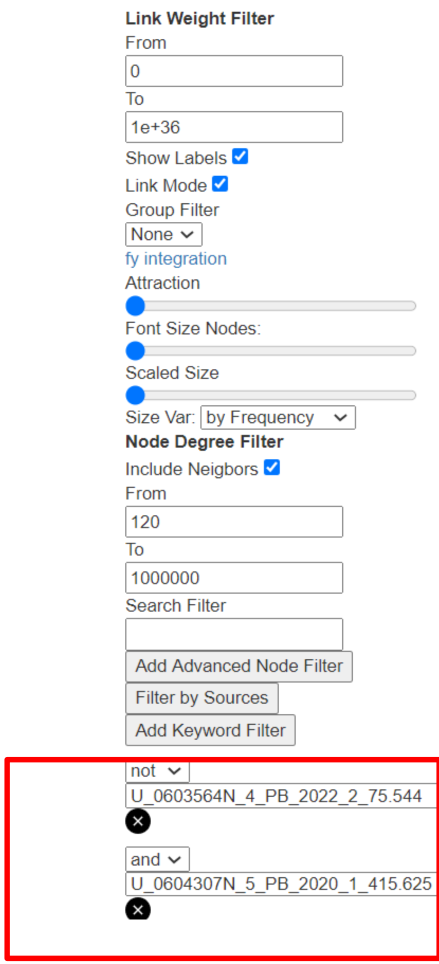


Word pair in a low cost PE but not in a high one

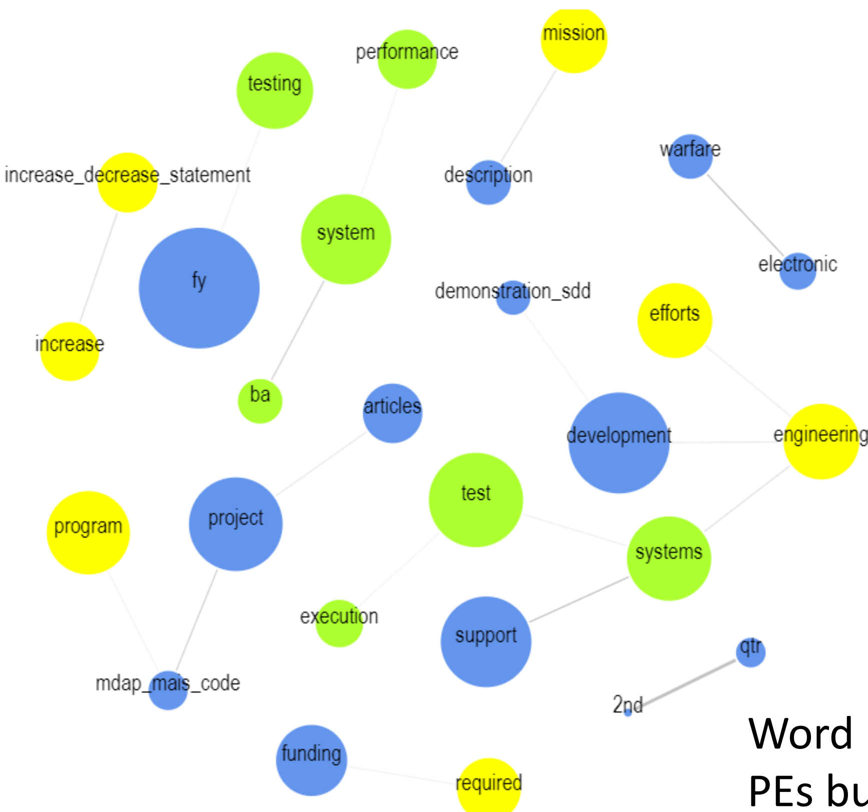
Lexical Link Analysis



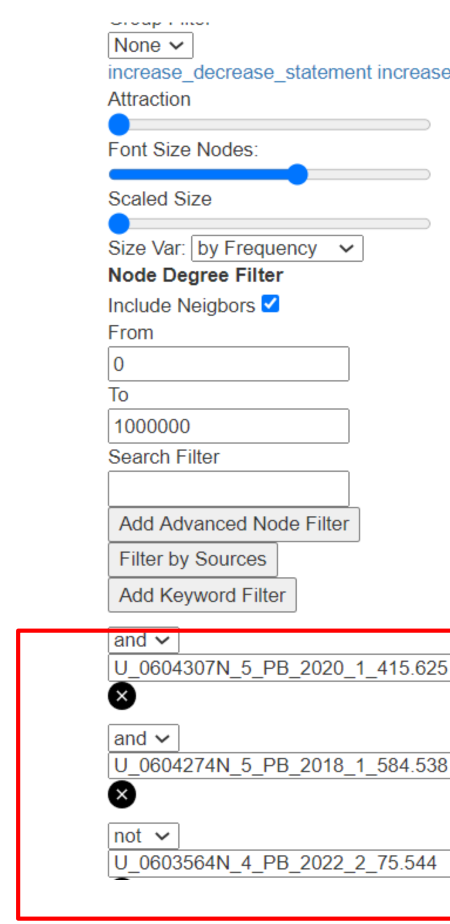
Compare a high cost and low one

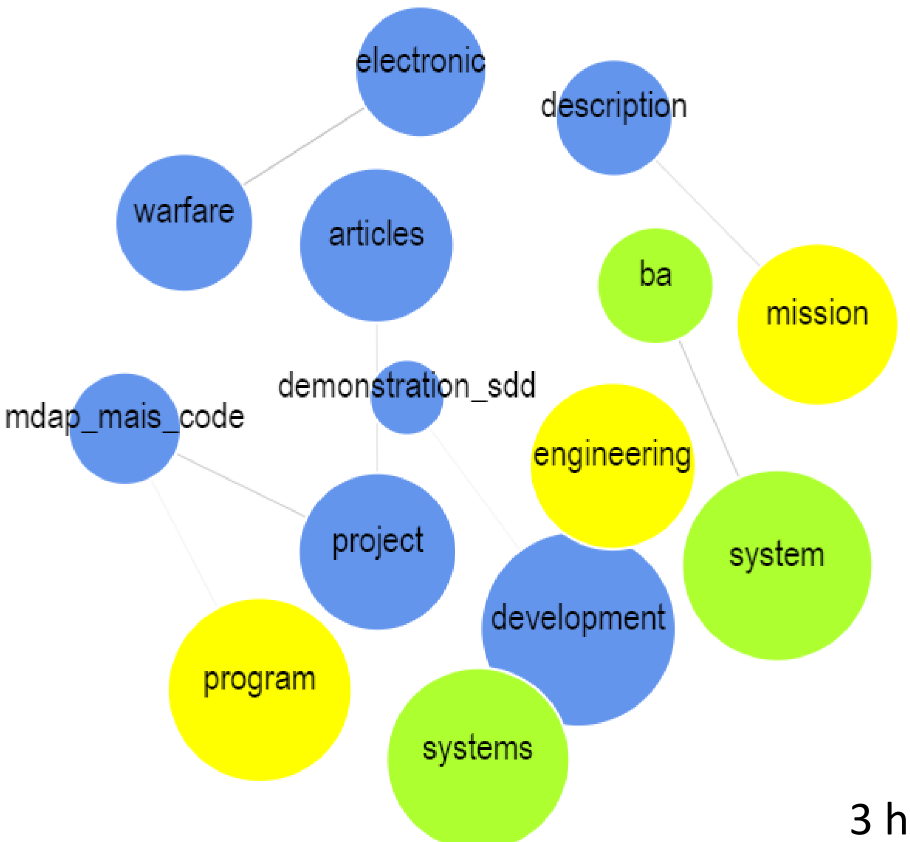


Two High Programs and One Low

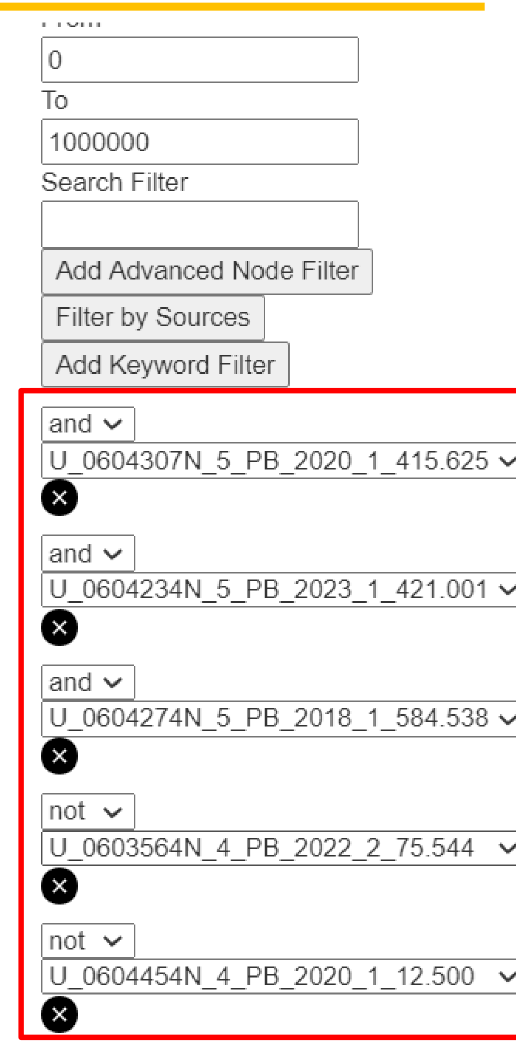


Word pair in two high cost PEs but not in a low one





3 highs and 2 lows

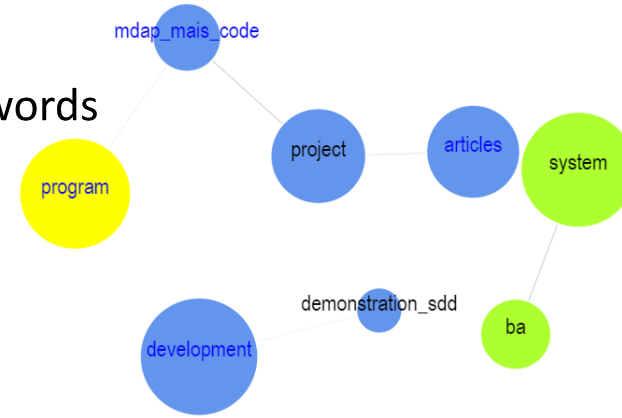


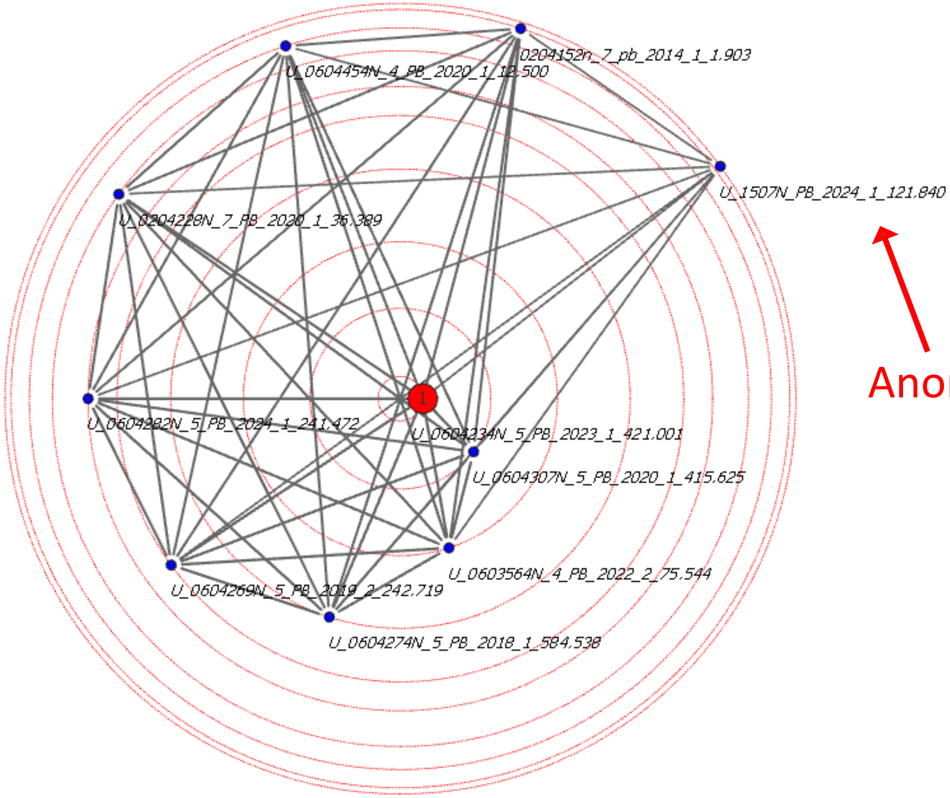
Lexical Link Analysis: Match Matrix

Match Matrix From Lexical Link Analysis: Updated on Using 'Combined' Word Pairs

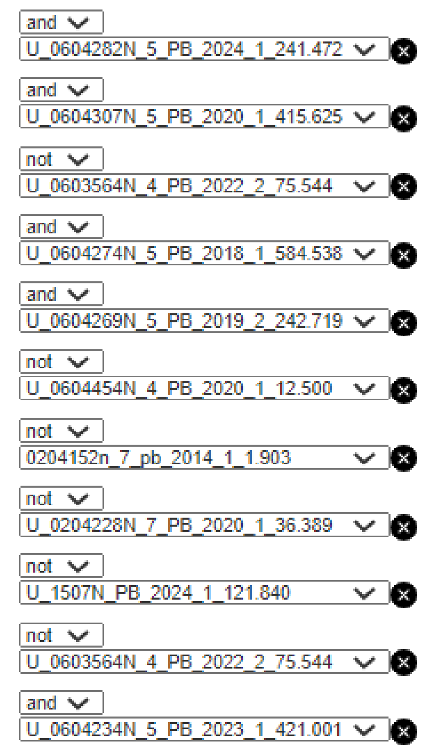
		Match Score	U_0604234N_5_PB_2023_1_421.001	U_0604307N_5_PB_2020_1_415.625	U_0603564N_4_PB_2022_2_75.544	Uniqueness Score
1	U_0604234N_5_PB_2023_1_421.001	257.00	_	138.00	132.00	1010.00
2	U_0604307N_5_PB_2020_1_415.625	213.00	138.00	_	117.00	1209.00
3	U_0603564N_4_PB_2022_2_75.544	204.00	132.00	117.00	_	532.00
4	U_0604274N_5_PB_2018_1_584.538	193.00	119.00	85.00	90.00	221.00
5	U_0604269N_5_PB_2019_2_242.719	189.00	134.00	123.00	111.00	407.00
6	U_0604282N_5_PB_2024_1_241.472	153.00	90.00	72.00	82.00	183.00
7	U_0204228N_7_PB_2020_1_36.389	153.00	87.00	102.00	81.00	591.00
8	U_0604454N_4_PB_2020_1_12.500	89.00	51.00	58.00	68.00	55.00
9	0204152n_7_pb_2014_1_1.903	67.00	53.00	43.00	49.00	64.00
10	U_1507N_PB_2024_1_121.840	12.00	9.00	11.00	10.00	48.00

Keywords

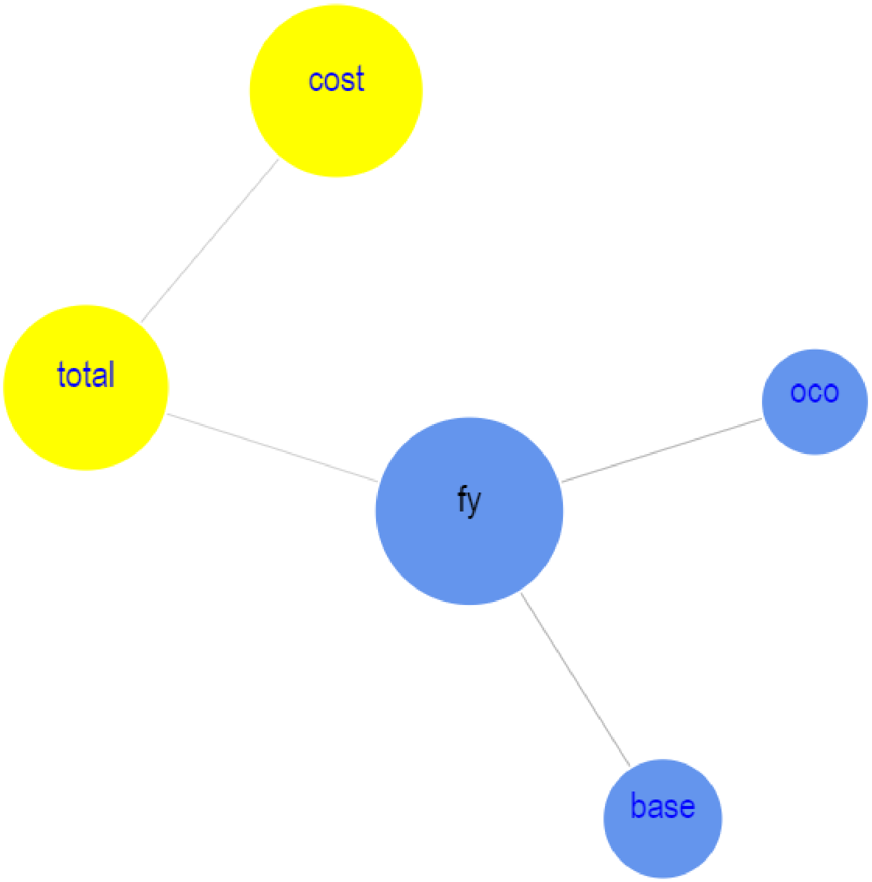




Anomaly?



The centered ones are more overlapped with others



Add Advanced Node Filter

Filter by Sources

Add Keyword Filter

and
U_0604454N_4_PB_2020_1_12.500

and
0204152n_7_pb_2014_1_1.903

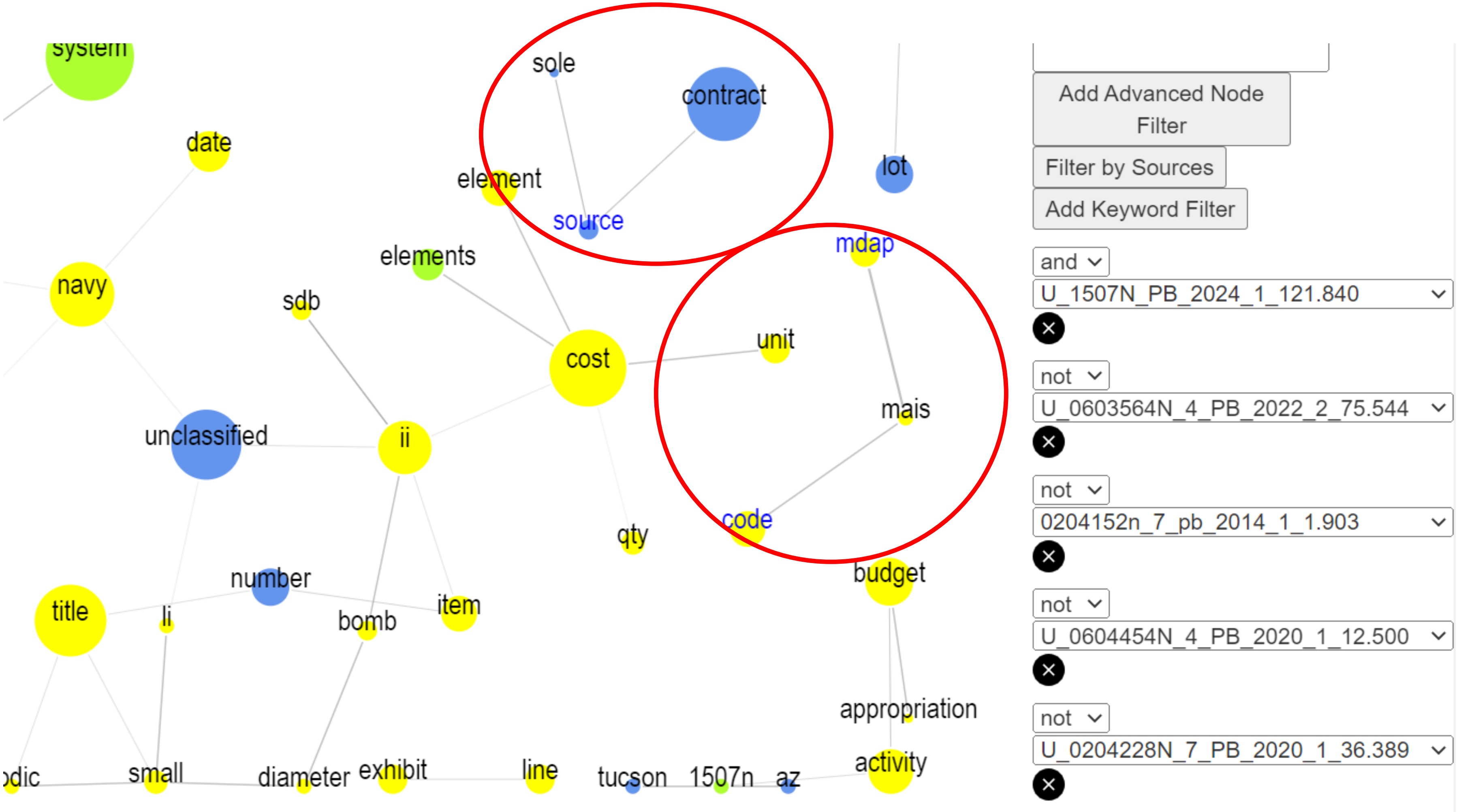
and
U_0204228N_7_PB_2020_1_36.389

and
U_1507N_PB_2024_1_121.840

and
U_0603564N_4_PB_2022_2_75.544

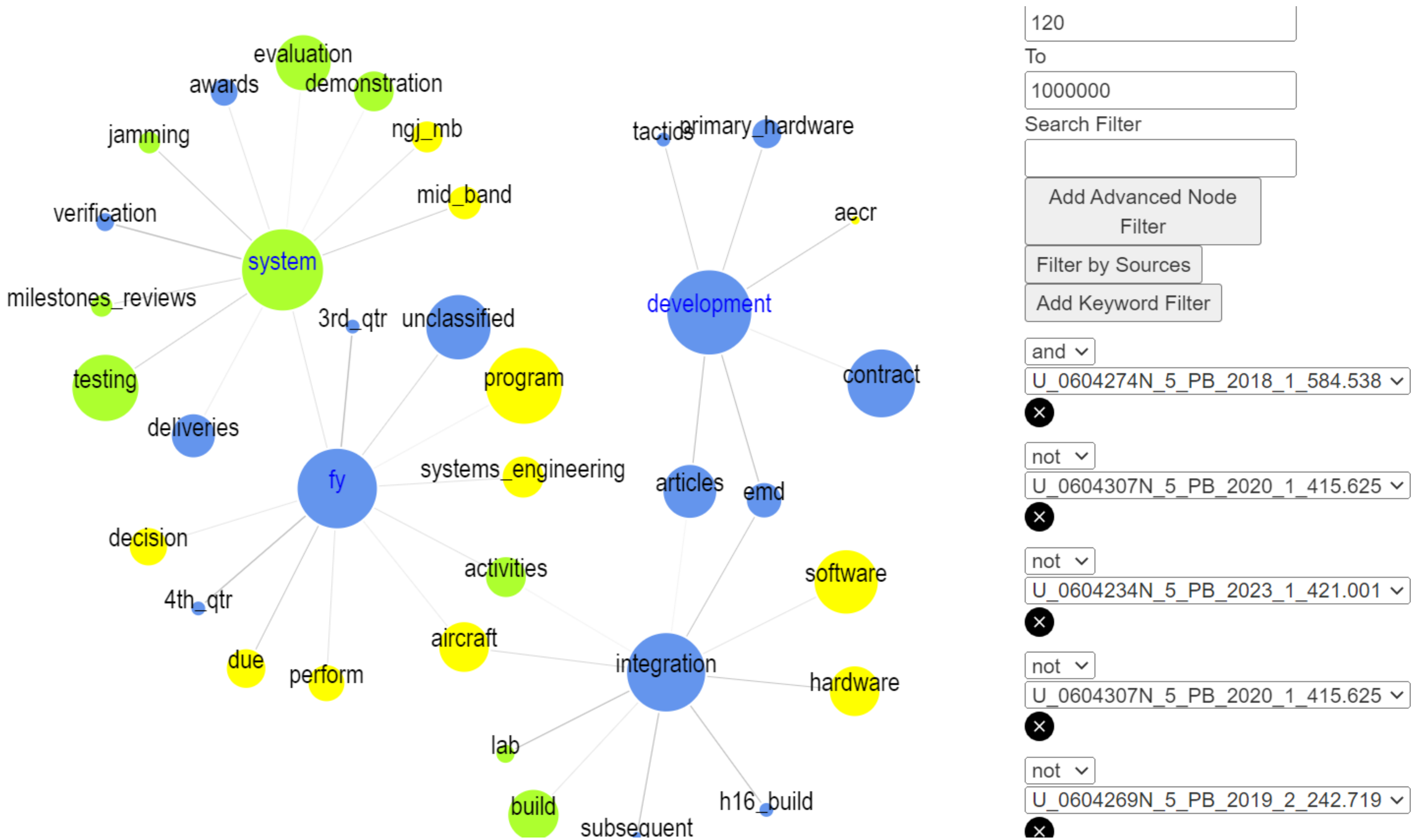
and
U_0604234N_5_PB_2023_1_421.001

and
U_1507N_PB_2024_1_121.840



Discover the key words as causes for higher prices

THANK YOU


evaluation
awards
demonstration
jamming
ngj_mb
verification
mid_band
system
milestones_reviews
3rd_qtr
unclassified
testing
program
deliveries
fy
systems_engineering
decision
activities
4th_qtr
aircraft
due
perform
tactics
primary_hardware
aecr
development
contract
articles
emd
software
integration
hardware
lab
build
subsequent
h16_build
120
To
1000000
Search Filter
Add Advanced Node Filter
Filter by Sources
Add Keyword Filter
and
U_0604274N_5_PB_2018_1_584.538
not
U_0604307N_5_PB_2020_1_415.625
not
U_0604234N_5_PB_2023_1_421.001
not
U_0604307N_5_PB_2020_1_415.625
not
U_0604269N_5_PB_2019_2_242.719

Appendix B

Proposal Details

20th Annual Acquisition Research Symposium

ID: P23-0019
Created On: November 10 2022
Title: Structured and Unstructured Data Sciences and Business Intelligence for Analyzing Requirements Post Mortem
Type: Paper/Presentation
Status: Received
Keywords: lexical link analysis,named entity extraction,NEE,parts of speech tagging,PoS,spaCy,network analysis,centrality measures,supervised machine learning,predictive and scoring models

Paper/Panel Paper

Name of Presenter Ying Zhao

Presenter Organization Naval Postgraduate School

Presenter Email Address yzhao@nps.edu

Presenter Phone Number 408-218-8484

Abstract Navy systems may have unexpected significant cost growth for many reasons. There is an urgent need to leverage advanced analytics to understand the common elements and causes of significant cost growth from existing requirements documents and open-source media. The need includes identifying the characteristics of capability requirements from Initial Capability Documents (ICD), Key Performance Parameters (KPP), or Key-Systems Attributes (KSA) from Capability Development Documents (CDD) and Operational Requirements Documents from previous requirements processes that may have contributed to cost growth.

The author applied various text analyses, link analysis, network analysis, and causality analysis to the DoD programs requirements data from the operational requirements documents and previous processes. The automatic discovery of the correlations and causations using deep analytics will greatly facilitate the prediction and prevention of the financial risks for building Navy systems in the future.

Research Issue The research issues are listed as follows
1. What are common elements of requirements that create excessive cost growth in Navy systems?
2. Assuming the elements are identified, determine the risk (likelihood and magnitude) of cost growth from common elements for both procurement and sustainment costs.

Research Results Statement The author located the cost growth risks (likelihood and magnitude) in terms of characteristics including capability requirements (unstructured), key performance parameters (structured data), key systems attributes (structured data), keywords, themes, and entities. Tools also included lexical link analysis, spaCy for entity extraction.

The author also applied apply network/graph tools to visualize the risks and capabilities in terms of relations and centralities of the networks of keywords and measures. The author also applied causal sciences and counterfactual calculation in junction with lexical link analysis to discover the key words that are associated with higher cost increase rates for Navy systems.

No Files have been uploaded

Authors

Name	Title	Organization	Phone	Email	Primary	Attending
Ying Zhao	Research Professor	Naval Postgraduate School	408-218-8484	yzhao@nps.edu	Yes	Yes

Biographies

Ying Zhao

Dr. Ying Zhao is a research professor at the Naval Postgraduate School (NPS). Her research focused on data sciences, machine learning, artificial intelligence, artificial general intelligence methods, including lexical link analysis (LLA), collaborative learning agents (CLA), and reinforcement learning for search, visualization, and analysis, for defense military applications in the areas of semantic and social networks, common tactical air pictures, combat identification, logistics, wargaming, and mission planning. Since joining NPS, Dr. Zhao has been a principal investigator (PI) of many awarded DoD research projects. Dr. Zhao is a co-author of four U.S. patents in knowledge pattern search from networked agents, data fusion, and visualization for multiple anomaly detection systems. She received her PhD in mathematics from MIT and is the co-founder of Quantum Intelligence, Inc.