



## Acquisition Research Program: Creating Synergy for Informed Change

# LEXICAL LINK ANALYSIS (LLA) APPLICATION: IMPROVING WEB SERVICE TO DEFENSE ACQUISITION VISIBILITY ENVIRONMENT(DAVE) May 13-14, 2015

Dr. Ying Zhao, Dr. Douglas J. MacKinnon, Dr. Shelley P. Gallup,  
Research Associate Professors

Distributed Information Systems Experimentation, Naval Postgraduate School

# Report Documentation Page

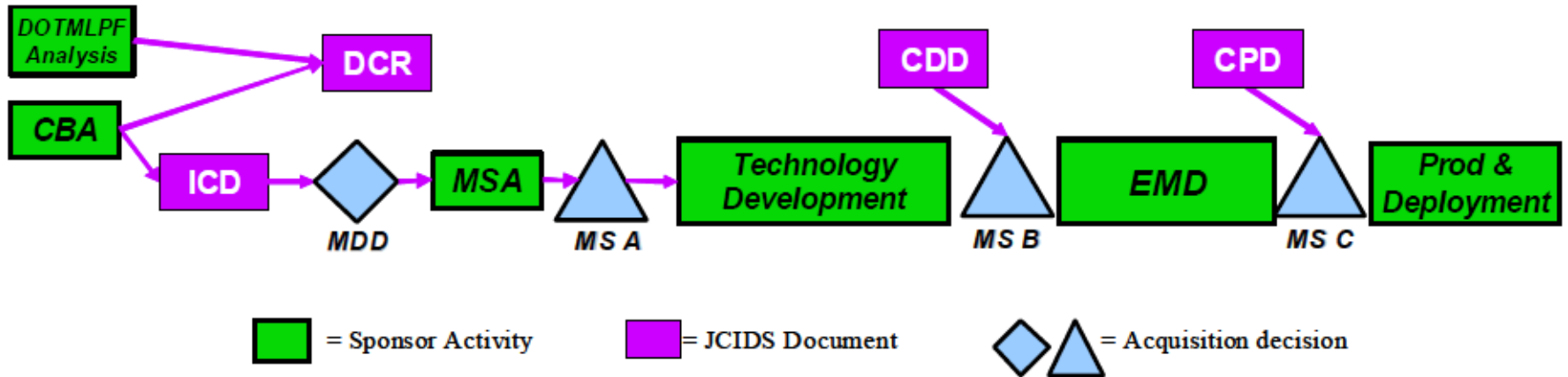
Form Approved  
OMB No. 0704-0188

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

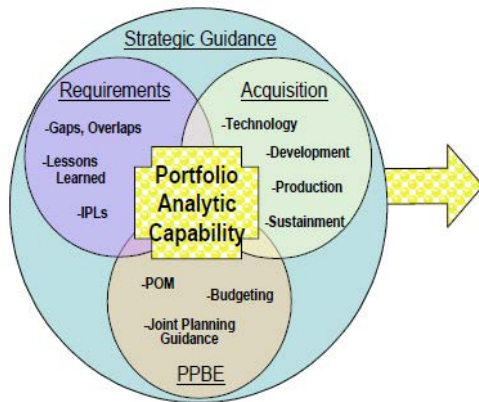
1. REPORT DATE <b>MAY 2015</b>		2. REPORT TYPE		3. DATES COVERED <b>00-00-2015 to 00-00-2015</b>	
4. TITLE AND SUBTITLE <b>Lexical Link Analysis (LLA) Application: Improving Web Service to Defense Acquisition Visibility Environment (DAVE)</b>				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) <b>Naval Postgraduate School, Monterey, CA, 93943</b>				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT <b>Approved for public release; distribution unlimited</b>					
13. SUPPLEMENTARY NOTES <b>Presented at the 12th Annual Acquisition Research Symposium held May 13-14, 2015 in Monterey, CA.</b>					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT <b>Same as Report (SAR)</b>	18. NUMBER OF PAGES <b>16</b>	19a. NAME OF RESPONSIBLE PERSON
a. REPORT <b>unclassified</b>	b. ABSTRACT <b>unclassified</b>	c. THIS PAGE <b>unclassified</b>			



# Background: Critical Needs of Automation, Validation, and Discovery



## JCIDS Process and Acquisition Decisions (J-8 CJCSI 3170.01G)(JCIDS, 2009)



### Multiple Portfolio Views:

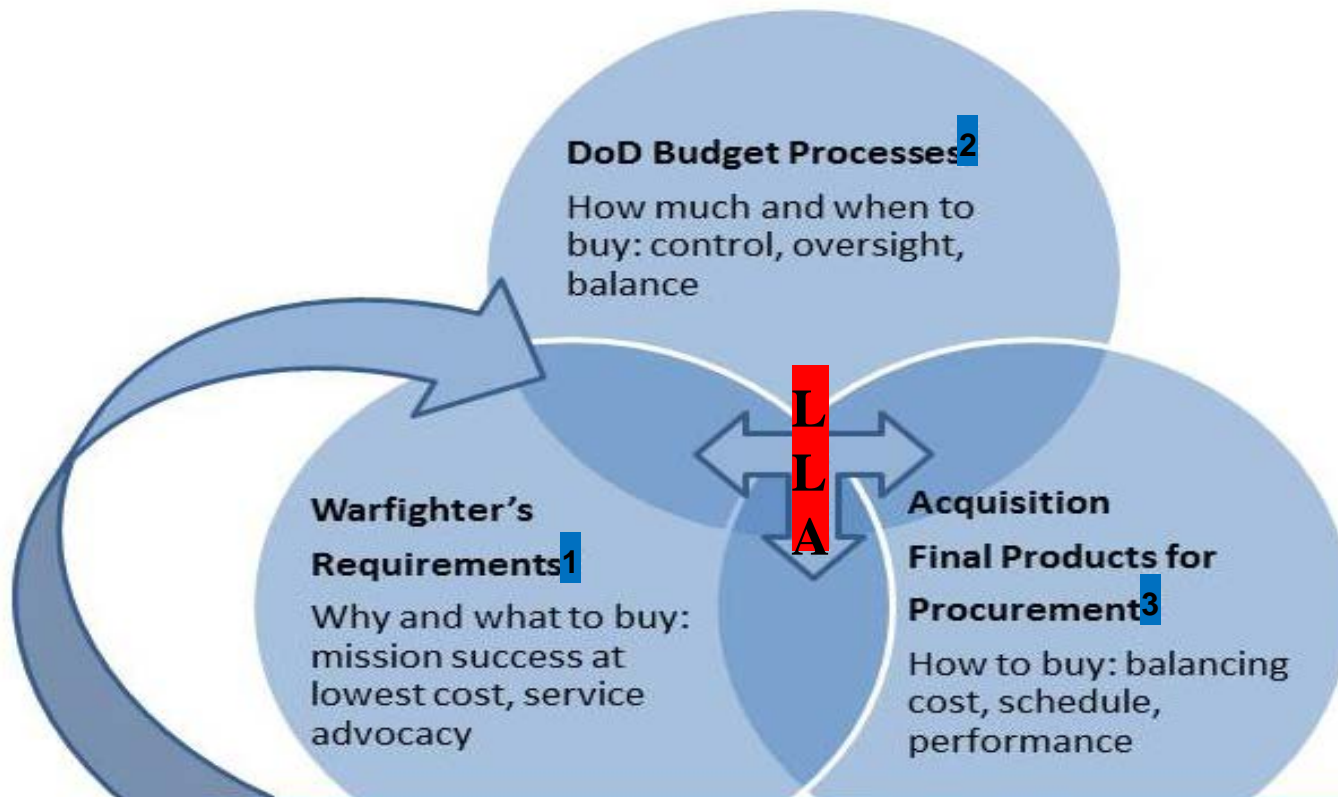
- Systems vs. Capabilities
- Investment vs. Capabilities
- System Context
- Highly dependent programs (Joint Enablers)
- Procurement Optimization
- S&T vs. future needs
- Sustainment Efficiency
- Market Value

- Data are too voluminous, unformatted, and unstructured!
- Need to leverage automation
  - Extract relations among PE, MDAP, and ACATII
  - Extract costs





# DoD Acquisition Decision Making



LLA automates the possibility to develop awareness of the “fit” between the budget, final products, and requirements.

<sup>1</sup>e.g. Universal Joint Tasks Lists, Urgent Needs Statements, <sup>2</sup>e.g. Program Elements, <sup>3</sup>e.g. Weapon Books





# Methods

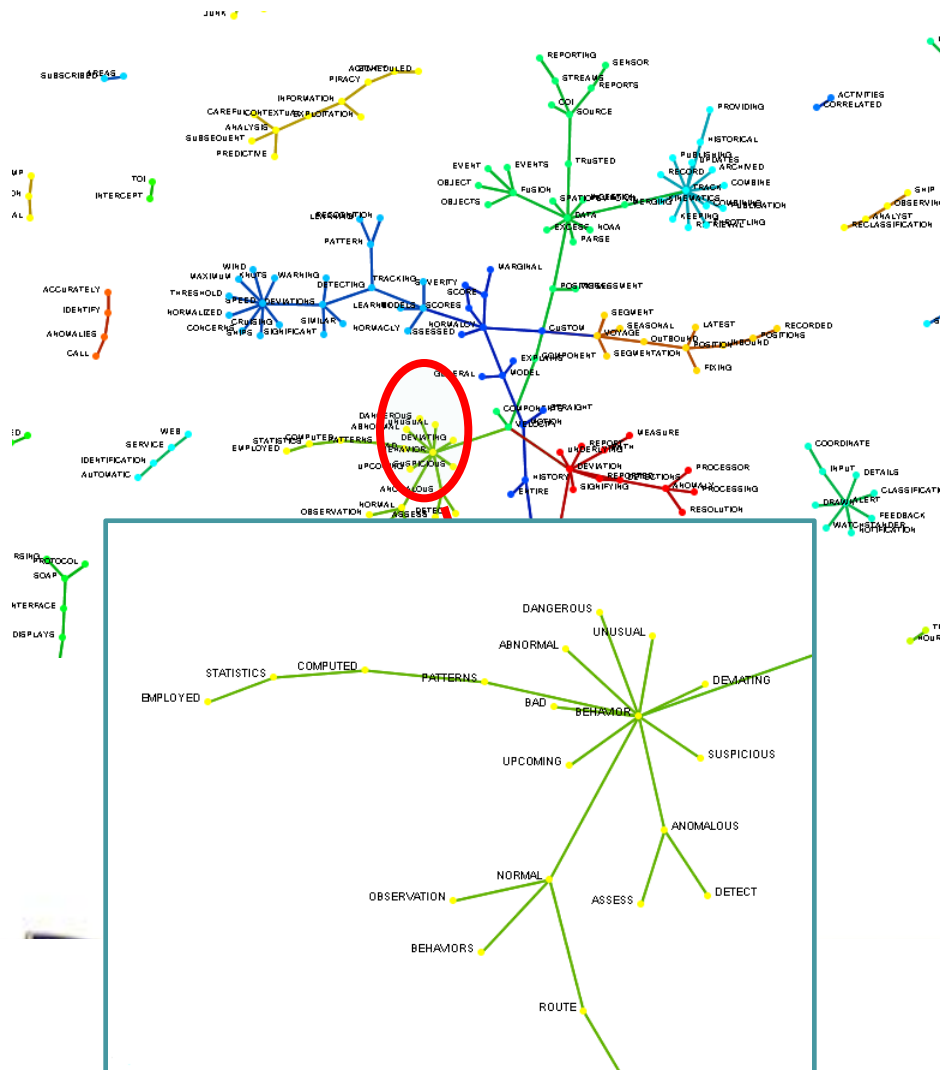
---

- Lexical Link Analysis (LLA) Core
  - LLA Reports and Visualizations
- Collaborative Learning Agents (CLA) for pattern recognition that scales up to Big Data
- System Self-Awareness (SSA)
- Big data and Deep Learning (BDDL) / Big Data Architecture and Analytics (BDDA)





# Lexical Link Analysis (LLA) Core



## Language-Independent Processing Steps

1. Extract word pairs (bigram) in context, connect them as word networks (text-as-networks)
2. Group word networks into clusters / themes (by color)

## Optional – Language-dependent Pre-processing Steps\*

1. Stop-word Removal: Leave out stop words (e.g., a, of, the)
2. Named Entity Extraction: Leave in/out people, places and organizations
  - Focus on semantics
  - Use NE in social networks
3. Parts of Speech Tagging (Manning & Schütze, 1999): Use only nouns and verbs
4. Stemming: Use only word roots

\* Source: Stanford NLP (SNLP, 2012), Apache openNLP

# Applying LLA as a Data-Driven Automation Technology and Methodology across the DoD Acquisition Process



- Surface themes and their relationships across multiple data sources
- Compare/correlate data from multiple data sources
- Sort/rank important and interesting information
- Discover high value areas for investment





# Research Progress 2015

---

- **Task 1:** Work with the OUSD(AT&L) to install the LLA/CLA/SSA system as a web service in the Defense Acquisition Visibility Environment (DAVE) test bed via the AT&L eBusiness Center Service.
  - Approved and in progress.





# Defense Acquisition Visibility Environment (DAVE)

- DAVE interfaces with authoritative data sources to support the OUSD(AT&L) and his staff in decision-making, oversight and analysis for major programs
- Sample data sources
  - **Defense Technical Information Center (DTIC)**
    - Program Elements (PEs):  
<http://www.dtic.mil/descriptivesum/>
  - **Defense Acquisition Management Information Retrieval (DAMIR)**  
(<http://www.acq.osd.mil/damir/>)
    - Selected Acquisition Report (SAR)
    - Defense Acquisition Executive Summary (DAES)
    - Acquisition Program Baseline (APB)
  - **Acquisition Information Repository (AIR)**
    - Test & Evaluation Master Plan (TEMP)
    - Systems Engineering Plan (SEP)
    - Life Cycle Sustainment Plan (LCSP)
    - Acquisition Strategy Reports (ASR)
    - Acquisition Decision Memorandum (ADM, for Milestone B 2366b Certification)
    - Technology Readiness Assessment (TRA), and LCSP



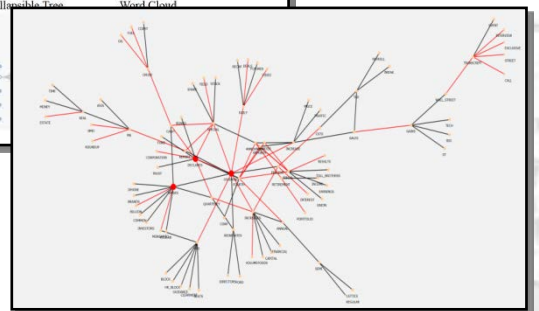
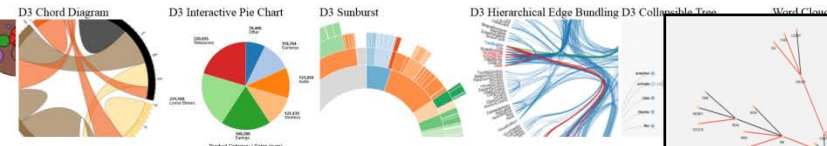
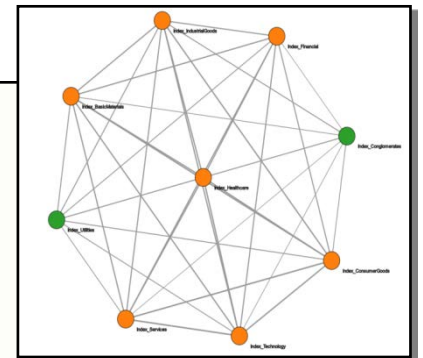
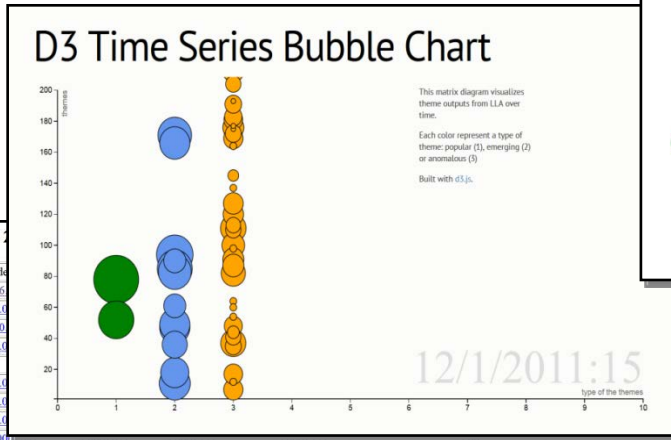


# Example LLA Reports and Visualizations

Event	Date	Theme ID	Theme Keywords	Visualization	Matched	Unique	Total	Consensus	Caps
all		85(E)	FINANCIAL CONDITION, FINANCIAL WINNERS	Visual	10	19	29	0.34	0.66
all		47(E)	GLANCE, COMPUTER, FINANCIAL, GLANCE	Visual	18	35	53	0.34	0.66
all		110(E)	ANNOUNCES SALE, ANNOUNCES MONTHLY	Visual	15	30	45	0.33	0.67
all		84(E)	SHARES DECLINE, OUTSTANDING SHARES	Visual	11	32	43	0.26	0.74
all		174(E)	WEEKLY RECAP, BUSINESS RECAP, BUSINESS WEEKLY, BUSINESS RECAP	Visual	35	96	131	0.40	0.60
all		83(E)	WORTHMLN, PAYMEN	Visual	15	34	49	0.32	0.78
all		166(E)	ENERGY CONFERENCE, HOLD CONFERENCE	Visual	24	49	73	0.33	0.67
all		61(E)	BUT BUYS, BUYS CHILDRENS	Visual	13	35	48	0.19	0.81
all		244(E)	HEALTH SERVICES, HEALTH INFORMATION, INFORMATION SERVICES	Visual	7	20	27	0.26	0.74
all		86(E)	CANADA, GANGLAS, PRIERS	Visual	8	32	40	0.14	0.86
all		89(E)	SENSOR, ANALYST, ANALYST MEETING	Visual	13	33	46	0.25	0.75
all		96(E)	AIRLINE INDUSTRY, AIRLINE COZ	Visual	10	31	41	0.24	0.76
all		256(E)	LEAD PLAINTIFF, PERCUTANEOUS LEAD	Visual	8	32	40	0.20	0.80
all		18(E)	BANKSHARES CORP, FILES, SEC FORM, RESOURCES CORP, FILES, SEC FORM	Visual	1	37	38	0.03	0.97
all		125(A)	RANKING, HOT, HOT TRENDS	Visual	9	12	21	0.25	0.75
all		267(A)	NEUSTAR, ANNOUNCED, PACIFIC, ANNOUNCED	Visual	6	29	35	0.22	0.78

Match Matrix From Lexical Link Analysis: Updated on Mon Apr 27 11:20:51

	Match Score	Index_Services	Index_Financial	Index_BasicMaterials	Index_ConsumerGoods	Index_Healthcare	Index_IndustrialGoods	Index_Utilities
1 Index_Services	252.00	-	121.00(0.26)	115.00(0.28)	123.00(0.36)	146.00(0.36)	89.00(0.25)	17.00(0.15)
2 Index_Financial	185.00	121.00(0.26)	-	97.00(0.25)	69.00(0.22)	146.00(0.36)	70.00(0.25)	20.00(0.12)
3 Index_BasicMaterials	176.00	115.00(0.28)	97.00(0.25)	-	84.00(0.29)	146.00(0.36)	70.00(0.25)	20.00(0.12)
4 Index_ConsumerGoods	154.00	123.00(0.36)	69.00(0.22)	84.00(0.29)	-	146.00(0.36)	70.00(0.25)	20.00(0.12)
5 Index_Technology	146.00	146.00(0.36)	93.00(0.25)	100.00(0.29)	87.00(0.31)	-	89.00(0.28)	17.00(0.15)
6 Index_Healthcare	114.00	89.00(0.28)	67.00(0.23)	74.00(0.28)	68.00(0.31)	146.00(0.36)	-	20.00(0.12)
7 Index_IndustrialGoods	89.00	70.00(0.25)	55.00(0.21)	57.00(0.25)	59.00(0.31)	89.00(0.28)	146.00(0.36)	-
8 Index_Utilities	17.00	20.00(0.12)	22.00(0.15)	25.00(0.19)	17.00(0.15)	20.00(0.12)	22.00(0.15)	25.00(0.19)





# Big Data and Deep Learning (BDDL)

---

- **Task 2:** We are also exploring how to use LLA jointly with other business intelligence tools especially Big Data Architecture and Analytics (BDAA) tools





# Big Data and Deep Learning

- **Data Storage**
  - Safe and fault-tolerant
  - SQL vs. NoSQL
    - Graph databases
    - Unstructured data
    - Cloud infrastructures
- **Operational Systems**
  - real-time
  - no- or low-latency of response
- **Analytics**
  - Map/Reduce paradigm related to data sciences
    - Map: parallel computing
    - Reduce: data fusion
  - Apache Mahout , Spark

- **Challenge**
  - Traditional data sciences algorithms
    - Tightly coupled Map/Reduce computations
  - Need to augment traditional data sciences with the “Map/Reduce” paradigm to take advantage of massive distributed parallel computing

- Predictive analytics
  - Identify high-valued targets
- Graph analytics
  - Core driver for Big Data analytics
- Text analysis
  - Data-driven discovery of themes and concepts
- Recommendation engine
  - Recommend high-valued targets explicitly
- **Deep learning (especially useful for CID)**
  - Self-taught learning: e.g. discover/identify interesting objects (e.g. cats and dogs) in massive images and videos  
<http://www.nytimes.com/2012/06/26/technology/in-a-big-network-of-computers-evidence-of-machine-learning.html>
  - Unsupervised machine learning
    - Pattern recognition
    - Anomaly detection
  - Data fusion
    - Abundant DoD applications

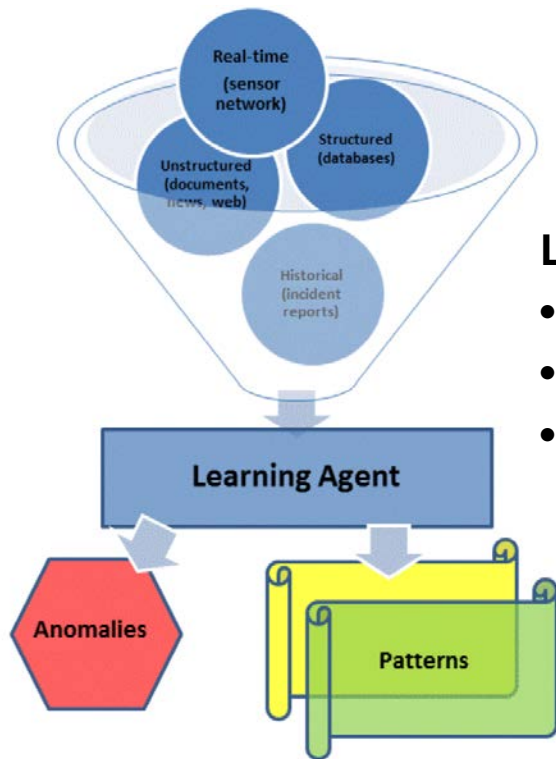
## Commercial Trends for Analytics

**Zhao, Y., MacKinnon, D.J., and Gallup, S.P. (August, 2015). Big Data and Deep Learning for Understanding DoD Data. In the *CrossTalk*, the Journal of Defense Software Engineering, Special Issue “Data Mining and Metrics.”**



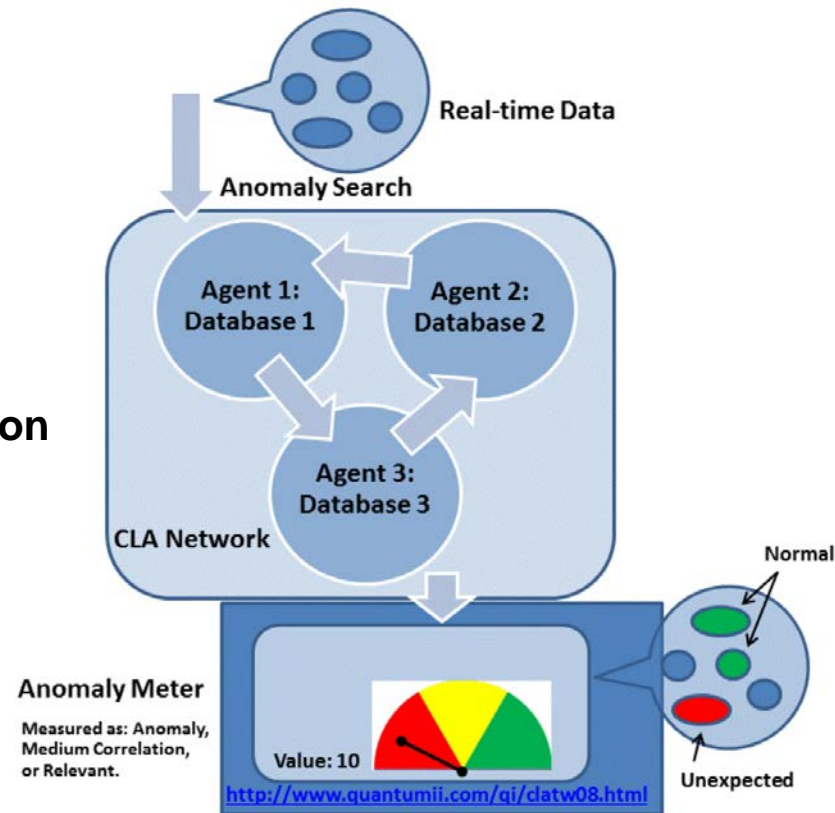


# Collaborative Learning Agent (CLA): It is Big Data and Deep Learning!



- LLA is Deep Learning**
- Pattern Recognition
  - Anomaly Detection
  - Recursive Data Fusion

A Single Collaborative Learning Agent (CLA) Patterns are graded from medium to relevant correlations.



Agent Collaboration



# How Big Data and Deep Learning can be Applied to the DoD Acquisition Process

- Examine and compare the important acquisition data sources to gain business insights:
  - Study Prime and subcontractor relationships
  - DOT&E Annual Reports for horizontal analysis
  - Compare DOT&E and DAES: What did the OT result say and did issues surface?
    - Did these issues show up somewhere in DAES earlier?
  - Time series analysis
    - Can LLA look at program ratings over time?
  - Compare budget data and contracts data



# How Big Data and Deep Learning can be Applied to the DoD Acquisition Process



(cont'd)

- In the current acquisition process, a small delay or anomaly in a contract negotiation process can have a huge impact in its performance, therefore cost the government a lot of money downstream.
- It will be very useful to apply BDAA such as LLA for pattern recognition and anomaly detection for these kind of problems and make early warnings and predictions to prevent the downstream risks.
- The Big Acquisition Data might include programs' cost, SAR, DAMIR, tech data, even outside economic environment data if the access is possible.
- The causes of the deviations from the normal behaviors for the programs/contracts might be modeled using physics (e.g., fluid dynamics theories).



# How Big Data and Deep Learning can be Applied to the DoD Acquisition Process

(cont'd)



- LLA's network perspectives, social plays among the nodes and the System Self-Awareness (SSA) theory may be used to lay out the academic vigor for the business processes, for example, answer the questions
  - Are some nodes drawn towards some other nodes because the other nodes are more powerful?
  - Is the preferential attachment growth pattern or expertise growth pattern can be used here?
  - How are the forces of the nodes modeled and mapped into the social network settings and actual business processes?





# Acknowledgements

---

- We thank Mr. Robert Flowe from OUSD(AT&L)/ARA who provided sponsorship to access the Defense Acquisition Visibility Environment, and relevant questions along with insightful guidance and discussions.

