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
Lexical Link Analysis (LLA) Application: Improving Web Service to Defense Acquisition Visibility Environment (DAVE)

Naval Postgraduate School, Monterey, CA, 93943

Approved for public release; distribution unlimited

Presented at the 12th Annual Acquisition Research Symposium held May 13-14, 2015 in Monterey, CA.

Same as Report (SAR)

16
Background: Critical Needs of Automation, Validation, and Discovery

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

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

DoD Budget Processes
How much and when to buy: control, oversight, balance

Warfighter’s Requirements
Why and what to buy: mission success at lowest cost, service advocacy

Acquisition Final Products for Procurement
How to buy: balancing cost, schedule, performance

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

1e.g. Universal Joint Tasks Lists, Urgent Needs Statements, 2e.g. Program Elements, 3e.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)
  - 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
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.
New Tools

- 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

Commercial Trends for Analytics

- 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
  - Unsupervised machine learning
    - Pattern recognition
    - Anomaly detection
  - Data fusion
    - Abundant DoD applications

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

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

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

System and Method for Knowledge Pattern Search from Networked Agents (US patent 8,903,756)
Zhao et al., Quantum Intelligence, Inc. (www.quantumii.com)
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.