



TECHNICAL ACTIVITY PROPOSAL

Panel/Group	SAS	Activity Title	
Reference Number	SAS-IST-179	Semantic Representation to Enhance Exploitation of Military Lessons Learned	
Activity Type	Task Group		
Panel Approval Date 19 Oct 2022	Board Approval Date 09 Dec 2022	Activity Start Date 31 Jan 2023	Activity End Date 31 Jan 2026
Related Activity	IST Panel, IST-ET-118, IST-ET-119, IST-ET-111, IST-192, SAS-IST-171		
Projected Meeting Location(s)	Jan 2023 – CSO Paris FRA		
Lead Nation(s)	JALLC - Joint Analysis and Lessons Learned Centre , United States		
Team Leader(s)	Mrs Jacqueline EATON (JALLC) Dr Douglas LANGE (USA)		
Panel/Group Mentor			
NATO Nations/Orgs Invited to Participate			
Non-NATO Nations/Orgs Invited to Participate	STOEOP		
Nations/Orgs Committed to Participate	DEU USA TUR ACT JALLC NCIA GBR		
Security Classification Level or Marking of the Activity	NATO Unclassified (NU)		
Keywords	AI, Deep Learning, Knowledge Management, Lessons Learned, NLP, Semantic Representation, SocialLinguistics		
NATO and National Resources Required	Facilities for meetings. Access to relevant/appropriate NATO and national exercises, experiments and hackathons. Access to NATO cloud-based software development environment.		

TECHNICAL ACTIVITY PROPOSAL

BACKGROUND

In 2018, the NATO Joint Analysis and Lessons Learned Centre (JALLC) conducted two exploratory research studies looking into the potential of modern data science techniques to support Lessons Learned (LL) analysis tasks. LL are typically documented in unstructured or semi-structured text and the hypothesis was that Natural Language Processing (NLP) techniques may help analysts to detect trends, improve the quality and utility of information held in LL repositories, and identify undocumented lessons in formal and informal communications. A key finding of these studies was that commercial-off-the-shelf NLP performed very badly when applied to NATO texts because the systems lacked the necessary semantic representation to understand the LL in a NATO context. It was recommended that a NATO LL ontology should be developed in order to enable the LL community to benefit from modern data science techniques. As JALLC has continued to explore ways to innovate in LL, under the remit of the NATO LL Capability Improvement Roadmap 2020-2025, a need for a wider range of scientific perspectives and expertise to tackle this difficult problem has been identified. Specifically, more understanding about how semantic representation can be used to enhance LL exploitation is needed in order to prepare future LL systems for the demands of AI-driven modern warfare.

MILITARY RELEVANCE

The future of warfare is increasingly automated and AI-driven. A particular challenge is to effectively process large amounts of information, both in terms of the volume and the velocity, to support decision making in a timely manner. There are lots of tools that can support this, but they are more effective when the information that is being processed is quantitative and structured. LL are knowledge artefacts, which are usually recorded as unstructured or semi-structured text and tend to be highly dependent on context, meaning that there is a high degree of prior knowledge required for a reader to be able to understand and reuse the LL. Existing tools deliver poor results when dealing with this type of data. NATO and many nations are currently working on modernizing and transforming their LL systems to increase the automation of LL activities. These innovations are intended to increase the quality, relevance and accessibility of LL stored within the system, which are all currently limited by the skills and capacity of the personnel to have to manually deal with a high volume and velocity of generated LL information. This problem will only get worse as more information is expected to be generated in the future. Semantic representations would help connect the knowledge stored in LL with wider military domain knowledge, and as such speed up the processes of implementing lessons into doctrine and standards, training and exercises and tactics, techniques and procedures.

SCIENTIFIC OBJECTIVE(S)

Enhance understanding of how to achieve the rapid, reliable, and robust automation of key LL activities, such as capture, dissemination, causal analysis, classification, summarization, search, clustering, and trend analysis of military LL, in an explainable way.

The focus areas will be on understanding how to:

- Enhance semantic representations using AI, and vice versa.
- Exploit semantic representations in military LL systems.
- Integrate socio-linguistic understanding of military LL language into semantic representations.

The objectives are to:

Objective 1: Develop use cases for semantic representations in military LL systems.

Objective 2: Develop position papers that explore key concepts, e.g., modeling and reasoning over LL, LL as a KM concept, Definition of LL “nuggets”, human expert labeling/input, notions of distance/surprise/novelty, integration with operational systems, utilization of a broad spectrum of data to capture knowledge.

Objective 3: Develop publicly releasable problem statement(s) and challenge dataset(s) with evaluation metrics, that can be used by NATO or nations to launch their own research or hackathon/innovation challenges.

Objective 4: Develop (partial) semantic representations and demonstrate them applied to (elements of) the use cases.

EXPECTED ACHIEVEMENTS

TECHNICAL ACTIVITY PROPOSAL

The results will provide NATO nations and NATO information to make best use of cutting-edge technologies to exploit explicit and tacit knowledge within KM systems in general and specifically in the context of Lessons Learned and contribute to the Innovation Line of Effort of the NATO LL Capability Improvement Roadmap 2020-2025.

The pre-capability development includes :

- Multiple use cases build the foundation for scoping the problem definition at the beginning of the RTG and highlighting the benefit of semantic knowledge representation for knowledge management in the context of LL at the end of the RTG.
- The major expected achievement is a methodological approach for extracting and continuously adapting generalized semantic representations of military language concepts from unstructured and semi-structured texts enabling the automated detection of knowledge artefacts with value to LL analysis.
- Applying the methodological approach will result in an example semantic representation of LL and military domain building the backbone for pre-capability demonstration.
- A key enabler is a challenge dataset for training, testing and validation of AI models, supplemented by the definition of evaluation metrics to be used for quality assessment.
- The technical results will be accompanied by the creation of a community of experts that is able to exploit the results within their nations.
- Overview of Knowledge Management activities, covering the base collection and capture of knowledge, its dissemination, and an evaluation mechanism.
- The final expected output is a demonstration and report on technical approaches.

SCIENTIFIC TOPICS TO BE COVERED

- The scientific topics are investigating possible ways to narrow the gap between human understanding and machine representation of information and knowledge.
- All topics aim at increasing the automation of tasks along the entire lifecycle of knowledge management.
- Exploration of semantic representations and embeddings, for guaranteeing a balance between scalability and expressivity.
- The recent advances in the projection of semantics on well-chosen mathematical spaces extend the expressivity for representing information and knowledge.
- Exploring possible combinations of semantic representations that are able to make use of the advantages of approaches from different domains of AI research.
- A thematical focus will be put on the identification and representation of causal relations within knowledge bases.
- Design of a dataset and evaluation metrics for training, testing and validating AI models.
- Use of Natural Language models to stimulate adaptation of semantic representations.
- The effects of socio-linguistic contexts resulting from different military domains
- Definition/adoption of a "notion of distance" within the context of semantic representation of KM, with the purpose of identifying novel knowledge to be integrated in LL.

SYNERGIES AND COMPLEMENTARIES

The proposed RTG work on semantic representations of a military body of knowledge has synergies with ongoing ACT capability development through the NATO Warfighting Capstone Concept. Specifically, the RTG's work is related directly to NATO's Warfighting Development Agenda (WDA) which identifies Data Exploitation as a key enabler to Cognitive Superiority (ref: NWCC & WDA) and continues to allow ACT to explore the utility of emerging and disruptive technologies. The interoperability and functionality of systems is likely to be enhanced by a common semantic framework that is informed by insights reached through this RTG study.

This RTG will contribute to the Innovation Line of Effort of the NATO LL Capability Improvement Roadmap 2020-2025.

The RTG activity also has synergies with the NATO Defence Innovation Accelerator (DIANA). DIANA can be leveraged to

TECHNICAL ACTIVITY PROPOSAL

facilitate engagement with academia and industry to further explore, develop, and implement the use-cases developed by the RTG. As such, the RTG will develop challenge problems to DIANA hubs, potentially providing a practical challenge case for DIANA.

There is a strong synergy between this activity and IST Panel activities including [Combination of Symbolic and sub-symbolic (AI IST-ET-118); Causal reasoning (IST-ET-119); Knowledge representation and reasoning (IST-ET-111); Human-AI Interaction - ANTICIPE (IST-192); and C2 Services in MDO for Federated Mission Networking (SAS-IST-171)]. Outputs from those groups, as well as direct engagement, will help to inform the RTG with respect to recent efforts in these areas within NATO and Nationally. A cross-panel Research Symposium with IST-ET-119 will be explored and proposed as a separate activity.

GBR, USA and NATO are currently involved in capability improvement/development of their LL capabilities and systems and linking the LL ontology development to those efforts represents an opportunity to enhance interoperability and functionality of the systems and contribute to LL standardization efforts.

EXPLOITATION AND IMPACT

One game changing effect would be to ensure decision makers have access to relevant lessons at the right time to avoid repeating mistakes and promote sharing of best practices across the full range of military activities. Another game changing effect would be to enable a single unified search experience that connects all national and NATO LL systems providing more comprehensive and more relevant results faster.

The RTG will develop a collaborative community that brings together wide ranging multidisciplinary technical expertise on linguistics & ontologies; data science, machine learning, natural language processing; operations research analysis; and military users (including Lessons Learned practitioners and other Knowledge Managers). In addition to participating in the RTG, the group will facilitate exploitation of findings in NATO Bodies, and National agencies, including at the NATO JALLC, and national Lessons Learned organizations, and other knowledge management organizations (e.g., NATO standardization office)

As a contribution to the Innovation Line of Effort of the NATO LL Capability Improvement Roadmap 2020-2025, and similar national LL capability development initiatives, the outputs of this RTG will inform future scoping and requirements for future capability development of next-generation/intelligent/ enhanced Lessons Learned (IKM) systems.

The RTG's exploration of state-of-art and rapidly evolving tools and techniques, through a set of use-cases will inform procurement or development of future systems, for example related to improved semantic search, knowledge integration & dissemination, trend & causal analysis, classification & clustering, and summarization.

In addition to Lessons Learned applications, the outputs of the practical utility of semantic knowledge representations explored by this RTG will be generalizable to other related knowledge management fields which can benefit from similar rapid ingestion of structured and unstructured data, its searchability and analysis, and its dissemination. This can be, for example, related to NATO and National standards and doctrine at NATO Standardization Organization, or improved semantic search on NATO/National CIS systems. This work will deliver a fundamental contribution to implementation of the NATO Data Exploitation Framework Policy.

The set of use-cases and curated challenge dataset, developed by the RTG, will provide an avenue for innovation and exploitation through further experimentation and capability development in NATO and Nations, and external partners. Engagement with external academic, industry, and national government research organizations (e.g., via DIANA). In such a rapidly evolving area, the participation of academia and industry is crucial and will be facilitated by speaking engagements, hackathons, etc.. These will provide a common platform and baseline for further development and practical implementation.

Exploitation and further development can also occur through interoperability exercises such as NATO CWIX exercises and National equivalents.