LINCOLN LABORATORY
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

1.0 TOPIC TITLE:

Text Analytics Pipeline and User Interface

2.0 SUMMARY:

The effort is to create a system for ingesting documents and making them discoverable through a web-based user interface. The system will consist of backend components including ingest pipeline, storage, and REST services implemented on top of an open-source stack, as well as a user interface implemented with a modern JavaScript library. This challenge will require familiarity with text analytics, information retrieval, and user interface design.

3.0 BACKGROUND:

MIT Lincoln Laboratory's Intelligence and Decision Technologies Group (Group 104) develops prototype systems for data exploitation and knowledge management that improve intelligence and military decision-making. Systems are infused with computer vision, natural language processing, and other machine learning algorithms to help users explore and make sense of large amounts of data.

Structured Knowledge Space (SKS)[1] is an end-to-end software system developed by MIT LL to solve a problem that has frustrated national security decision makers: "How do we take advantage of the enormous amounts of information communicated daily through a wide variety of reporting venues?" SKS builds searchable archives of text-based intelligence reports, extracts information such as named entities, times, and geographic references from free-form ("unstructured") documents, and makes the information discoverable through a keyword and faceted-search interface.

Many components of SKS initially required custom-built solutions. However, due to the rapid pace of open-source software development in the text analytics and big data domains, it is now possible to implement an SKS-like system largely from existing open-source components.

Additional concept ideas that support this challenge are also welcome.

---

[1] https://www.ll.mit.edu/publications/technotes/TechNote_SKS.pdf