

Semantic Graph Concept Detection

Challenge

The ability to prevent, mitigate, or control a biological threat depends on how quickly the threat is identified and characterized. Ensuring the timely delivery of data and analytics is an essential aspect of providing adequate situational awareness in the face of a disease outbreak.

Approach

Arcturus is a hybrid architecture, developed as part of the biosurveillance pipeline. It implements three concurrent stages during a work flow for streaming data annotation:

- » Integrated a collection of new data required to support facility utilization and functionality analysis with existing laboratory facility space data
- » Published basic facility utilization and functionality reports to display the potential of new analysis capabilities
- » Implemented processes and procedures to improve and maintain data quality
- » Satisfied new DOE Infrastructure Assessment data requirements

Arcturus is a hybrid framework capable of mining domain knowledge bases, enabling high-speed concept classification of a data stream for near-real-time event detection.



Detecting Emerging Global Epidemics

The framework is designed to enable expansion of NER algorithms and include new ontologies. The workflow results are stored in a RDF event graph with an internal Web Ontology Language (OWL) Event Reporting Ontology to enable future extensibility for detecting emerging events within a data stream.

Methodology

Arcturus enables a specified domain ontology to become a concept mine within the analytic pipeline, providing speed and accuracy in detecting terms within a data set. This functionality is possible through two possible configurations: 1) query the ontology directly with SPARQL matching, and 2) enable NER detection with a trained set of tokens and classes mined from an ontology. The second configuration uses the One Health Ontology (OHO) to create a set of training tokens for a NER library. Through a developed set of RDF rules and queries, the OHO is mined for desired terms and synonyms necessary for training the NER library. This library is then used to scan and extract classes of tokens from an English language news feed. Currently, the Arcturus biosurveillance instance assembles selected NER libraries by mining the OHO for core terms, classes, and relationships.

Due to the hybrid architecture, the number of implementations of the different technologies could create a very large number of competing applications. Furthermore, it is not necessary to use a graph database, as small updates to the storage model will create the same semantic graph structure in both relational and key-value data stores without the built in reasoning capabilities.

Impact

A version of Arcturus is currently deployed to support biosurveillance initiatives, and is trained to detect concepts and ontology relations within the OHO. The event graph created by the biosurveillance instance tracks emerging diseases from a data stream of current news feeds. Future work proposes a mine of this event graph for retraining machine learning algorithms in support of emerging pandemic detection.

CONTACT

William Smith
Semantic Data Scientist
(206) 218-7327
william.smith@pnnl.gov

Mark Greaves
Initiative Lead
(206) 528-3300
mark.greaves@pnnl.gov



Proudly Operated by **Battelle** Since 1965

