This meeting is part of the project MOOD. This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No 874850

MOOD hybrid workshop

IDENTIFY SIGNS AND DRIVERS OF ZOONOTIC DISEASES EMERGENCE AND DIGITAL DATA RESOURCES FOR EPIDEMIC INTELLIGENCE

28th-29th September 2022, Science Museum MUSE -TRENTO (Italy)

Title: How text-mining could improve epidemiological surveillance?

Author: Mathieu Roche

Author' affiliation: CIRAD, UMR TETIS, Montpellier, France

Abstract text:

The ability to identify emerging and re-emerging diseases is challenging for the health domain. In this context, event-based surveillance (EBS) gathers information from heterogenous data sources, including online news articles. EBS systems integrate text-mining methods to deal with huge amounts of textual data. Text mining aims at discovering new information from textual datasets (i.e. corpus).

This talk focuses on the use text-mining and multidisciplinary approaches in order to mine news data dealing with the health domain. These data science approaches are integrated in an EBS system called PADI-Web (Platform for Automated extraction of Disease Information from the web). PADI-Web dedicated to animal health surveillance tackles disease-based and symptom-based surveillance. To address these issues different text-mining methods associated with labeled textual datasets are integrated in the main steps of EBS systems: data acquisition, information retrieval (i.e. identification of relevant texts), epidemiological information extraction, information to communicate to end-users.

In our work, specific methods are proposed to extract epidemiological events in multilingual news. First, a fine-grained classification enables to highlight texts dealing with "Outbreak declaration" (i.e. sentence classification). Second, a group of epidemiological information (i.e. locations, hosts, symptoms, etc.) is automatically extracted in relevant documents. Based on these approaches, different strategies and associated algorithms are proposed and discussed for improving EBS systems.

Keywords: Event-based surveillance, Text mining, Data science