

MOOD Science webinar

Wednesday, 2021, April 28 · 3:30 – 5:00 pm

Text mining on COVID19 datasets - Terminology extraction

Mathieu Roche will present his latest analysis on how to use terminology and textmining for event-based surveillance systems (i.e. disease-based and symptom-based surveillance). He will discuss the use of different datasets related to COVID-19, e.g. scientific publications, news data (PADI-web, MedISys), social media data (Twitter). The extracted terminology is used (i) for surveillance systems (i.e. web crawling and information extraction tasks) and (ii) for spatio-temporal analysis of tweets dealing with COVID-19.

Biography:

Mathieu Roche is a Senior Research Scientist (PhD, HDR) at CIRAD - TETIS research unit (France). Currently he is co-leader of the MISCA group (i.e. Spatial Information, Modelling, Data Mining, and Knowledge Extraction) at TETIS. Between 2005 and 2013, he was an Associate Professor (Maître de Conférences) at the University Montpellier 2, France. Mathieu Roche obtained a PhD in Computer Science from University Paris 11 (Orsay) in 2004 and then he led several academic and industrial projects in text-mining. Currently he is involved in two H2020 projects, among which LEAP4FNSSA and MOOD. Within MOOD, Mathieu is part of the executive board and collaborates with WP2 and WP3.

Publications:

Valentin S, Mercier A, Roche M, Lancelot R, Arsevska E. Monitoring online media reports for early detection of unknown diseases: insight from a retrospective study of COVID-19 emergence. Transboundary and Emerging Diseases, 2020 <u>https://onlinelibrary.wiley.com/doi/10.1111/tbed.13738</u>

Roche M. COVID-19 and Media Datasets: Period- and location-specific textual data mining. Data in Brief. Elsevier. Volume 33, 2020 https://doi.org/10.1016/j.dib.2020.106356

Decoupes R., Kafando R., Roche M, Teisseire M. H-TFIDF: What makes areas specific over time in the massive flow of tweets related to the covid pandemic? In Proc. of AGILE conference (Association of Geographic Information Laboratories in Europe), *to appear* 2021