A dark green vertical bar on the left side of the page, with a light green arrow pointing to the right, overlapping the title text.

1st International Workshop on Horizon scanning for Plant health

Participants booklet

9th - 10th of April 2024, Maisons-Alfort, France



Mathieu Roche, Research Scientist - Text Mining, CIRAD, TETIS

Epidemic intelligence integrates two components: indicator-based (e.g. official disease reports) and event-based surveillance (EBS) based on unofficial data. EBS tools may monitor digital media and other unofficial outlets. The Platform for Automated extraction of Disease Information from the web (PADI-web - <https://padi-web.cirad.fr>) has been developed initially for animal health surveillance, and recently for plant disease surveillance.



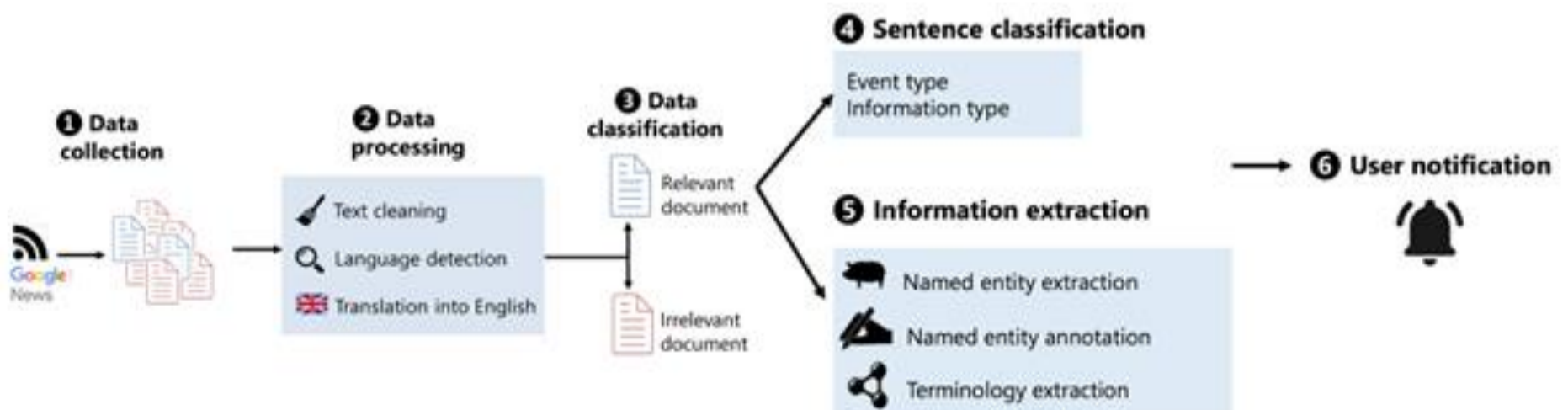
PADI-web automatically collects news via customized multilingual queries, classifies them and extracts epidemiological information. PADI-web retrieves articles daily from the news aggregator Google News through customized RSS feeds. An RSS feed is a combination of terms (disease names, symptoms or hosts). These terms have been identified by an approach combining text mining and domain experts. The RSS feeds are of two types:

- Disease-based surveillance consists of disease names and target seven animal diseases.
- Symptom-based surveillance includes clinical signs and hosts without any disease names.

RSS feeds are implemented in 28 languages (e.g. English, French, Chinese, Arabic, Italian, Russian, Turkish, etc.).

In order to identify relevant news and epidemiological information, machine learning approaches and language models (RoBERTa) have been integrated for monitoring animal and plant diseases. In order to implement these machine learning approaches (for learning classification models), labeled data (i.e. corpora) are used and/or created [1].

PADI-web pipeline, detailed in [2], involves six steps ranging from online news collection to the extraction of epidemiological features: (1) data collection, (2) data processing, (3) data classification, (4) sentence classification, (5) information extraction and (6) user notification.



The PADI-web pipeline collected more than 500,000 news dealing with animal health (Avian Influenza, African swine fever, West Nile Virus, TBE, etc.) to extract epidemiological events. The new instance of PADI-web dedicated to plant health (<https://plant.padi-web.cirad.fr>) collected 8,900 news items (monitored diseases: *Xylella fastidiosa* and *Fusarium Oxysporum Tropical*).

PADI-web is integrated into 2 international projects:

- H2020 MOOD (MONitoring Outbreaks for Disease surveillance in a data science context): <https://mood-h2020.eu/about-mood/>

- CARIBGreen - AFD (Plateforme collaborative caribéenne de santé globale en agroenvironnement pour le développement de la Caraïbe): <https://www.afd.fr/fr/carte-des-projets/caribgreen-plateforme-collaborative-caribeenne-sante-globale-agroenvironnement-developpement-de-la-caraibe>

Web page: <https://www.padi-web-one-health.org>

PADI-web team: <https://www.padi-web-one-health.org/presentation/team2/multidisciplinary-team>

References:

[1] Valentin, S., Arsevska, E., Vilain, A., De Waele, V., Lancelot R., Roche M. (2022). Elaboration of a new framework for fine-grained epidemiological annotation. *Scientific Data*, 9:655, 12 p. <https://doi.org/10.1038/s41597-022-01743-2>

[2] Valentin, S., Arsevska, E., Rabatel, J., Falala, S., Mercier, A., Lancelot, R., and Roche, M. (2021). PADI-web 3.0: A new framework for extracting and disseminating fine-grained information from the news for animal disease surveillance. *One Health*, 13:100357. <https://doi.org/10.1016/j.onehlt.2021.100357>

Roche, M., Rabatel, J., Trevennec, C., Pieretti, I. (2024). PADI-web for plant health surveillance. In Proceedings of CAISE'24 Forum, Demonstration Session, to appear