

Système de veille sanitaire pour analyser l'émergence et la propagation de maladies animales

Sylvain Falala¹, Jocelyn De Goër², Elena Arsevska¹, Mathieu Roche^{3,4}, Julien Rabatel⁴, David Chavernac¹, Pascal Hendrikx⁵, Barbara Dufour⁶, Renaud Lancelot¹, Thierry Lefrancois¹

¹ CIRAD & INRA, UMR CMAEE, Montpellier

² INRA, UR EPIA, CLERMONT-FERRAND

³ CIRAD, UMR TETIS, Montpellier

⁴ LABEX NUMEV, LIRMM, Montpellier

⁵ ANSES, UCAS, Maisons-Alfort

⁶ ENVA, EpiMAI, Maisons-Alfort

Résumé : La veille en santé animale, et notamment la détection précoce d'émergences au niveau mondial d'agents pathogènes, est l'un des moyens permettant de prévenir l'introduction en France de dangers sanitaires (Paquet *et al.*, 2006). Dans ce contexte, cet article présente une plateforme dédiée à la veille automatique allant du recueil des données textuelles (dépêches) jusqu'à la restitution synthétique des informations extraites.

Mots-clés : Veille sanitaire, Recherche d'Information, Extraction d'informations

1 Contexte

Dans le cadre de la thématique "Veille sanitaire internationale" de la Plateforme nationale d'épidémiologie en santé animale (Plateforme ESA), le Cirad, l'ANSES et la Direction générale de l'alimentation (DGAI) développent depuis 2013 un système de veille automatique du Web qui effectue : (1) le recueil quotidien de dépêches épidémiologiques provenant de sources non officielles, incluant les médias électroniques, (2) l'extraction automatique d'informations (nom de maladie ou symptômes, lieu, date et espèce touchée) issues de ces dépêches et (3) une restitution synthétique et agrégée de l'information : cartes, séries spatiotemporelles.

Les maladies actuellement surveillées sont la peste porcine africaine, l'Influenza aviaire, la fièvre catarrhale ovine, la fièvre aphteuse et la maladie de Schmallenberg. L'outil est développé de façon générique et permet la surveillance d'autres maladies. Ce système sera utilisé par la Plateforme ESA pour la France et par le réseau de vétérinaires CaribVet situé dans les Caraïbes.

2 Approche mise en œuvre

Le but de notre système de veille est de disposer d'un outil très réactif qui se veut complémentaire aux sources officielles comme l'Organisation mondiale de la santé animale (OIE) ou l'Organisation des Nations unies pour l'alimentation et l'agriculture (FAO).

Le recueil des dépêches s'appuie sur des requêtes constituées de mots-clés de maladies, d'hôtes et de symptômes pour collecter des articles issus de Google News. Ces mots-clés ont été définis par des experts et/ou par des méthodes de fouille de textes (Arsevska *et al.*, 2016).

Chaque article est prétraité et normalisé (suppression de balises HTML, reconnaissance de la langue, etc.) avant d’être stocké dans une base de données MySQL. Une interface Web permet de paramétrer le processus de recueil et de consulter les articles collectés (cf. Figure 1).

L’extraction d’information dans les dépêches collectées identifie les éléments clés (noms de maladies, lieux, dates, nombres et espèces d’animaux touchés). Elle repose sur des dictionnaires dédiés et des règles préalablement construites par un processus de fouille de données. Les premiers résultats sur un corpus de 357 dépêches montrent des scores d’exactitude d’environ 70% pour les informations spatiales et d’au moins 80% pour les autres types d’informations.

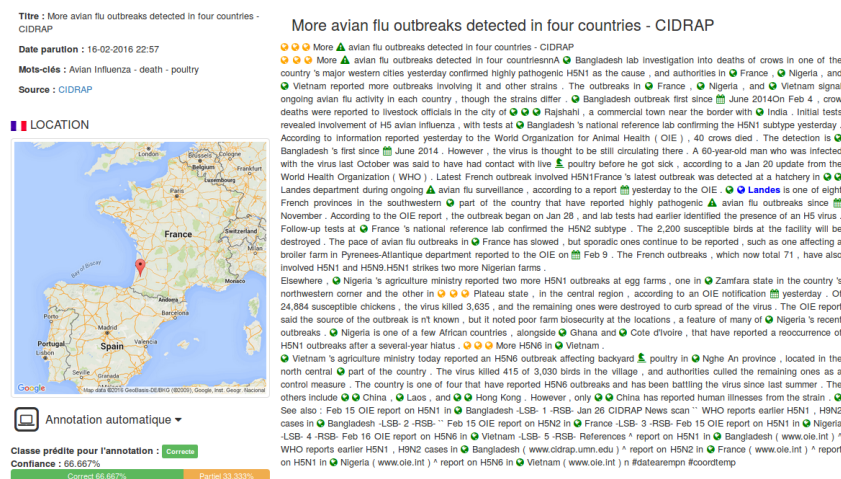


FIGURE 1 – Interface de consultation des dépêches.

3 Conclusion et perspectives

Cet article résume les travaux liés au développement d’une plateforme dédiée à la veille automatique allant du recueil des données textuelles (dépêches) jusqu’à la restitution synthétique des informations extraites dans les textes. Les informations actuellement extraites à partir des dépêches seront prochainement comparées aux informations issues des données officielles (OIE) afin de mettre en relief la découverte de l’émergence de maladies animales.

Remerciements : Les auteurs remercient les étudiants ayant participé au développement de l’outil : Max Devaud, Thomas Filiol, Baptiste Belot et Clément Hemeury. Ce travail est en partie financé par la DGAI et le Labex Numev (convention ANR-10-LABX-20).

Références

ARSEVSKA E., ROCHE M., HENDRIKX P., CHAVERNAC D., FALALA S., LANCELOT R. & DUFOUR B. (2016). Identification of terms for detecting early signals of emerging infectious disease outbreaks on the web. *Computers and Electronics in Agriculture*, **123**, 104 – 115.

PAQUET C., COULOMBIER D., KAISER R. & CIOTTI M. (2006). Epidemic intelligence : a new framework for strengthening disease surveillance in europe. *Euro surveillance*, **11**(12), 212–214.