

Système de collecte de données Web 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, Thierry Lefrancois¹, Barbara Dufour⁶, Renaud Lancelot¹

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). Cet article présente une plateforme dédiée à la collecte de données (dépêches) utiles pour la veille automatique. Le recueil des dépêches s'appuie sur des requêtes constituées de mots-clés de maladies, d'hôtes et de symptômes appliquées à Google News. Une interface Web a été développée pour consulter les articles collectés et paramétrer le processus de recueil en définissant de nouvelles combinaisons de mots-clés.

Mots-clés : Veille sanitaire, Collecte de données Web, Recherche d'information

1 Introduction

Dans le cadre de la thématique "Veille sanitaire internationale" de la Plateforme nationale d'épidémiosurveillance 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 :

- le recueil quotidien de dépêches épidémiologiques provenant de sources non officielles, incluant les médias électroniques.
- l'extraction automatique d'informations issues de ces dépêches.
- 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 Méthodes et outils d'acquisition de données Web

Le but de notre système de veille est de disposer d'un outil très réactif qui se veut complémentaire aux sources officielles telles que 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 appliquées à Google News (par exemple, la requête en anglais "*high fever AND mortality AND pigs*") qui combine deux symptômes et un hôte).

Ces mots-clés ont été préalablement définis par des experts et/ou par des méthodes de fouille de textes sur la base du logiciel BioTex (Lossio-Ventura *et al.*, 2016) développé dans le cadre du projet ANR SIFR¹. BioTex prend en compte deux facteurs pour extraire, de manière automatique, la terminologie dans des corpus textuels. Dans un premier temps, l'approche extrait des termes selon des patrons syntaxiques définis (nom-adjectif, adjectif-nom, nom-préposition-nom, etc.). Après un tel filtrage linguistique, un autre filtrage statistique est appliqué. Celui-ci mesure l'association entre les mots composant un terme. Enfin, des pondérations selon les sources de données sont proposées (Arsevska *et al.*, 2016).

Chaque article est prétraité et normalisé (suppression de balises HTML et Javascript, reconnaissance de la langue, etc.) avant d'être stocké dans une base de données MySQL.

Une interface Web (cf. Figure 1) a été développée en HTML, CSS, PHP et JavaScript. Elle permet de :

- *consulter les articles collectés*. Une section de recherche avancée permet de sélectionner les dépêches en combinant des critères comme les noms de maladie, les hôtes, les symptômes, les noms de source/du média et les dates de publication. Une section "Statistiques" permet de connaître le nombre d'articles recueillis pour une maladie donnée sur une période donnée et d'en observer la distribution dans le temps.
- *paramétrer le processus de recueil* en définissant de nouvelles combinaisons de mots-clés pour Google News.

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.

L'extraction d'informations sera prochainement intégrée au système. L'extraction 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 informations extraites à partir des dépêches seront comparées aux informations issues des données officielles (OIE) afin de mettre en relief la découverte de l'émergence de maladies animales.

1. Semantic Indexing of French Biomedical Data Resources (SIFR) project : <http://www.lirmm.fr/sifr>

Système de collecte de données Web pour la veille sanitaire des maladies animales émergentes

Recherche par titre :

Recherche contenu :

Résultat Recherche

The Horror of African Swine Fever in Eastern Europe NewsThe Horror of African Swine Fever in Eastern Europe09 September 2015 EUROPE - Over this su...

panama holds swine fever simulation - globalmeatnews.com
Panama holds swine fever simulation Panama is holding a simulation exercise to prepare its response in the event of an outbreak of class...

danish crown slaughterhouse closed on suspicion of swine fever - the copenhagen post - danish news in english
UPDATE: Danish Crown slaughterhouse closed on suspicion of swine fever UPDATE: Danish Crown's slaughterhouse in Herning was authorised to resume p...

pig farmers seek kcca help over swine fever outbreak - uganda radio network
Pig Farmers seek KCCA help over Swine Fever outbreak :: Uganda Radio Network In shortPig farmers in Lunguja, Busega and Nateete parishes in Rubaga...

2 swine flu cases among 3 fever deaths in tiruchy - the new indian express
2 Swine Flu Cases Among 3 Fever Deaths In Tiruchy - The New Indian Express TIRUCHY: In what has become a continuing saga, three fever-related d...

Titre : Pig Farmers seek KCCA help over Swine Fever outbreak - Uganda Radio Network

Date parution : 16-01-2016 08:26

Source : Uganda Radio Network

Pig Farmers seek KCCA help over Swine Fever outbreak :: Uganda Radio Network

In shortPig farmers in Lunguja, Busega and Nateete parishes in Rubaga division have appealed to Kampala Capital City Authority KCCA help them fight the African Swine Fever ASF that broke out four days ago.Pig farmers in Lunguja, Busega and Nateete parishes in Rubaga division have appealed to Kampala Capital City Authority (KCCA) help them fight the African Swine Fever (ASF) that broke out four days ago. Speaking to URN, Mrs Kyamufumba Regina, a farmer in Wakaliga, says that she has lost over 25 pregnant pigs in three days due to swine fever. She is unsure how to handle the situation. //Cue in: They haven't...// Cue out...hurting// She says that when she saw her first pig die, she was puzzled because her pigs had never suffered from swine fever before. //Cue in: I had...// Cue out...confirmed// Kyalagonza Amon, another farmer, says that he has lost over 10 pigs. He adds that he has now sold off the remaining pigs despite a quarantine that has been enforced by at Kampala Capital Authority (KCCA). According to Dr Emilian Ahimbislowe, who is in charge of veterinary services at KCCA, the disease is transmitted[1] by pigs eating infected pork or pork products; contact with infected pigs or their faeces. He says that quarantine for pigs and their products has been enforced in the affected area and advises farmers to take precaution in order to prevent the disease. //Cue in: When the...// Cue out...jk. // He adds that the authority cannot do any more than enforcing quarantine because there is no vaccine or treatment for swine fever. The disease is characterized by high fever, loss of appetite, bleeding under the skin and internal organs, and death occurs within 2-10 days. KCCA says that over 20 farmers have reported cases of sickly pigs and estimates that over 50 pigs to have died since the outbreak. References^ the disease is transmitted (www.merckvetmanual.com)

#datearemp
#cooortemp

FIGURE 1 – Interface de consultation des dépêches.

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 Numev² (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.
- LOSSIO-VENTURA J. A., JONQUET C., ROCHE M. & TEISSEIRE M. (2016). Biomedical term extraction : overview and a new methodology. *Inf. Retr. Journal*, **19**(1-2), 59–99.
- 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.

2. <http://www.lirmm.fr/numev>