

## GROUPE DE TRAVAIL « VEILLE SANITAIRE INTERNATIONALE » - VSI

La "Synthèse" résume l'état actuel des connaissances sanitaires et scientifiques des végétaux au travers de l'actualité en Europe et à l'International.

### OUTILS POUR LA VEILLE SANITAIRE

La veille scientifique et sanitaire de la Plateforme ESV vise à acquérir et transmettre des informations sur des thèmes spécifiques liés à la santé végétale et aux effets non intentionnels des pratiques agricoles sur l'environnement. Elle constitue une des **missions transversales** de la plateforme : *assurer une veille internationale sur les dangers sanitaires susceptibles d'avoir un impact sanitaire et/ou économique*. Pour obtenir une veille de manière relativement automatique, certains sites spécialisés (**EFSA**, **AlimAgri**, **Chambre d'Agriculture**...) dans les domaines d'intérêt de la plateforme proposent des abonnements ou des flux RSS permettant de recevoir régulièrement leurs lettres d'information ou des alertes dans une boîte de messagerie électronique. Les principales limites de cette approche sont la saturation rapide de la boîte de messagerie lorsque le nombre d'abonnements est important et les difficultés de tri des informations qui en résultent. Afin de gagner en efficacité, la veille sanitaire de la plateforme est en partie fondée sur des démarches informatiques telles que le web scraping et la fouille de texte permettant de collecter et pré-trier les informations jugées pertinentes au sein d'un corpus de sources défini.

### WEB SCRAPING

#### ► Description

Le *web scraping* est l'action de récupérer des informations sur le web de manière automatique via un script ou un programme, en se basant sur le code source des pages internet. Les packages *rvest* ou *xml2* du logiciel R permet d'implémenter le *web scraping*.

Grâce à des requêtes sur des mots clés référencés par intérêt, de multiples informations émanant de sites web (source, journal, date de parution, titre etc.) peuvent être récupérées. Elles sont ensuite stockées dans une base de données pour être traitées dans le cadre d'analyses ou de synthèses.

#### ► Exemple d'application

Dans le cadre de la veille sanitaire sur l'organisme nuisible *Xylella fastidiosa*, la Plateforme ESV récupère les informations parues sur le web. Par exemple, le 14 septembre 2019, *Nice-Matin* a diffusé un article sur le désaccord des oléiculteurs envers les mesures de lutte mises en place contre *Xylella fastidiosa*. Au moyen d'un script R de web scraping sur le moteur de recherche Google et en utilisant des mots clés définis au préalable, l'article de *Nice-Matin* a été récupéré, les informations intéressantes (le titre de l'article, la date de diffusion, la source, l'auteur, etc.) ont été sélectionnées et stockées dans une base de données.

#### ► Notes et références

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Wickham, H. (2019). *rvest*: Easily Harvest (Scrape). R package.

Wickham, H. (2019). xml2: Parse XML. R package.

## FOUILLE DE TEXTE

### ► Description

La fouille de texte est une technique qui permet d'extraire de manière automatique des informations à partir d'un corpus de textes. Cette méthode se base sur des algorithmes d'apprentissage automatisés, des arbres de décisions, des référentiels ou encore des corpus annotés.

Dans le but de développer et mettre en place un outil appliqué à la santé végétale, l'équipe opérationnelle de la Plateforme ESV collabore avec l'unité INRAE MaIAGE, l'unité CIRAD / INRAE ASTRE, et l'unité CIRAD / INRAE TETIS. Cet outil a pour objectif de récupérer des informations de types dates, localisations, détections issues de différents corpus de textes afin d'alimenter de manière automatique une base de données. Cette base de données, complémentaires des données de surveillance officielle, servira à enrichir la veille sanitaire et les analyses de la Plateforme ESV.

La construction de l'outil appliqué à la santé végétale prend notamment ses fondations sur PADI-web (Platform for Automated extraction of Disease Information from the web), qui a été développé dans le cadre de la veille sanitaire en santé animale, et sur Florilege.

### ► Exemple d'application

La Plateforme ESV prévoit d'utiliser cet outil de *text-mining* dans le cadre de la VSI (Veille Sanitaire Internationale) pour recueillir et extraire quotidiennement des informations issues d'articles médiatiques parus sur le web. Lorsqu'un article paraît par exemple dans un journal local pour informer de la nouvelle détection d'un type de maladie sur une espèce végétale, l'outil récupérera les informations clés (la source (lien url), la date de publication de l'article, la date de détection, le lieu, l'espèce hôte de la maladie ou de l'organisme nuisible, etc.) si elles sont mentionnées dans le texte. Les informations seront stockées dans la base de données de la Plateforme ESV et utilisées pour le bulletin de VSI ou pour une analyse de risque sur cette maladie ou cet organisme nuisible.

### ► Notes et références

Valentin, S. et al. (2020). PADI-web: A multilingual event-based surveillance system for monitoring animal infectious diseases. *Computers and Electronics in Agriculture*, 169(8) 105163. [PADI-web page](#).

Falentin, H. et al. (2017). Florilege: a database gathering microbial phenotypes of food interest. In 4th International Conference on Microbial Diversity 2017 , Bari, ITA (2017-10-24 - 2017-10-26).

Chaix, E. et al. (2018). Text-mining tools for extracting information about microbial biodiversity in food. *Food Microbiology*, pp. 1-13.

Conformément aux productions réalisées par la Plateforme d'Épidémiologie en Santé Végétale (ESV), celle-ci donne son droit d'accès à une utilisation partielle ou entière par les médias, à condition de ne pas apporter de modification, de respecter un cadre d'usage bienveillant et de mentionner la source © <https://www.plateforme-esv.fr/>

Nédellec, C. et al. (2018). L'ontologie OntoBiotope pour l'étude de la biodiversité microbienne. Presented at EGC'2018: 18èmes Conférence Internationale sur l'Extraction et la Gestion des Connaissances, Paris, FRA (2018-01-22 - 2018-01-26).

Conformément aux productions réalisées par la Plateforme d'Épidémiosurveillance en Santé Végétale (ESV), celle-ci donne son droit d'accès à une utilisation partielle ou entière par les médias, à condition de ne pas apporter de modification, de respecter un cadre d'usage bienveillant et de mentionner la source © <https://www.plateforme-esv.fr/>

Attribution - Pas d'Utilisation Commerciale - Pas de modification  
CC BY-NC-ND

