

L'IA pour le classement automatique des articles collectés via web scraping – cas de la VSI de la Plateforme ESV



« La Plateforme est en charge d'assurer une veille internationale sur les dangers sanitaires susceptibles d'avoir un impact sanitaire et/ou économique »
(extrait de la convention cadre).



<https://plateforme-esv.fr/>

[linkedin](#) 



Isabelle Pieretti (CIRAD) et Antoine Marullaz (INRAE)

Journée Vectopole Sud, Montpellier, 18-09-2025

Objectifs

Anticiper et mettre à jour des connaissances pour le gestionnaire du risque et les acteurs de la surveillance sur le territoire métropolitain et dans les DROMs.

- 1) Augmenter le nombre d'Organismes Nuisibles ciblés (couverture optimale de la VSI)
- 2) Produire une veille adaptée et la valoriser

Comment ?

Utiliser et développer des outils basés sur l'IA et la fouille de texte pour améliorer en continu le pipeline de la VSI (collaborations VSI avec SI, autres informaticiens, projets de recherche connexes).

Pour :

1. Réduire le nombre d'articles « non pertinents » (hors scope VSI)
2. Automatiser les tâches qui peuvent l'être
3. Optimiser les requêtes et les sources

Productions (exemple)

S'abonner gratuitement ici —>



BHV-SV 2025/17
Mois d'Avril 2025
semaine 17
du 21 au 27 avril 2025

Bulletin Hebdomadaire de Veille Sanitaire Internationale en Santé Végétale

Le Bulletin Hebdomadaire de Veille sanitaire internationale en Santé Végétale (BHV-SV) s'inscrit dans l'activité de veille sanitaire internationale menée dans le cadre de la Plateforme ESV (Plateforme d'Épidémiologie en Santé Végétale - <https://www.plateforme-esv.fr/>). Le BHV-SV sélectionne et résume des actualités sanitaires et scientifiques en santé végétale qui sont parues dans la semaine.

ATTENTION : Le contenu du document n'engage pas les membres de la Plateforme ESV.



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

Code juridique

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://plateforme-esv.fr/>

Confiance + est un indicateur sur la crédibilité des sites de diffusion (+ : peu fiable à +++ : très fiable, source officielle majoritairement).

Sommaire

Veille non ciblée	2
<i>Agrilus planipennis</i>	3
<i>Agrilus anxius</i>	3
<i>Xylella fastidiosa</i>	4

Popillia japonica

Veille sanitaire

En Suisse, dans le canton de Schwytz, différentes mesures dans la zone d'infestation et dans la zone tampon ont été prises via un arrêté général depuis la découverte de *Popillia japonica* au cours de l'été 2024.

Titre	Categorie	PaysSujet	Fiabilite	Lien
Mesures contre les scarabées japonais dans la région de Sägel	Réglementation, Evolution de l'état sanitaire	Suisse	Officielle	lien

résumé de l'article

autres informations

Veille scientifique

Voici un rapport de synthèse sur la biologie, la propagation, le potentiel de nuisance, les mesures de surveillance et de lutte concernant *Popillia japonica* (langues EN, FR, IT).

Titre	Categorie	Lien
Japanese beetle: current information on its biology, legal bases and control measures	Synthèse et sensibilisation	lien

Ceratocystis platani

Veille sanitaire

Plante & Cité a publié, le 18 mars 2025, une version enrichie du guide destiné à aider les gestionnaires et propriétaires de platanes à prévenir et limiter la dissémination la maladie due au champignon *Ceratocystis platani*.

Titre	Categorie	PaysSujet	Fiabilite	Lien
Un kit pour aider à lutter contre le chancre coloré du platane	Méthode et mesure de lutte, Méthode, outil et mesure de surveillance	France	Agronomique	lien



Productions (exemple)

<i>Spodoptera frugiperda</i>
<i>Popillia japonica</i>
<i>Candidatus Liberibacter</i> et ses vecteurs
<i>Toumeyella parvicornis</i>
<i>Xylella fastidiosa</i>

BULLETIN MENSUEL N°68

Plateforme ESV

Mars 2025



Le bulletin d'Épidémiosurveillance en Santé Végétale est une revue des actualités concernant la santé du végétal en Europe et à l'international. Il contribue à faciliter l'accès aux informations concernant la santé des végétaux et leur diffusion. Le bulletin est validé au préalable par une cellule éditoriale composée d'experts scientifiques et de collaborateurs partenaires ayant un rôle de conseillers.



Attribution - Pas d'Utilisation Commerciale - Pas de modification

CC BY-NC-ND

Code juridique

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/>

Le bulletin est enregistrable dans son format html sur votre ordinateur. Toutes les cartes du bulletins sont interactives (zoom/dézoom, informations cliquables directement sur la carte).

Sommaire

Sujet phytosanitaire	Zone géographique	Cultures	Nature de l'information
<i>Xylella fastidiosa</i>	Espagne	Multi-espèces	Évolution de l'état sanitaire et réglementaire
<i>Popillia japonica</i>	Italie	Multi-espèces	Évolution de l'état sanitaire et réglementaire
<i>Spodoptera frugiperda</i>	Grèce	Multi-espèces	Évolution de l'état sanitaire
<i>Toumeyella parvicornis</i>	Italie	Pins	Évolution de l'état sanitaire et réglementaire
<i>Toumeyella parvicornis</i>	Italie	Pins	Outil de détection des symptômes de dépérissement forestier causés par la cochenille tortue du pin
<i>Candidatus Liberibacter</i> spp. et ses vecteurs	Monde	Agrumes	Gestion collective d'une maladie végétale incurable : cas du Huanglongbing
<i>Candidatus Liberibacter</i> spp. et ses vecteurs	Portugal	Agrumes	Estimation du risque épidémiologique

Spodoptera frugiperda

Grèce / Multi-espèces / Évolution de l'état sanitaire

La présence de *Spodoptera frugiperda* en Europe a été établie pour la première fois en 2021 dans toutes les îles espagnoles de l'archipel des Canaries (Lanzarote, Fuerteventura, Gran Canaria, Tenerife, La Palma, El Hierro et La Gomera, source OEPP), puis en 2023 en Roumanie (comté de Calarasi, source OEPP), à Chypre (districts de Limassol et Larnaca, source OEPP), sur l'île portugaise de Madère (OEPP) et en Grèce. En Grèce, l'organisme de quarantaine prioritaire pour l'UE a été découvert dans le cadre de la surveillance officielle via un dispositif de piégeage à phéromones. Le ravageur a ainsi été capturé dans diverses préfectures du pays. En Laconie et Attique orientale pour la zone continentale, et dans les îles suivantes: Crète (Héraklion, Chania ou La Canée, Lassithi), Eubée, Lesbos, Salamine, Kos, Chios, Samos, Naxos et Syros. Dans la préfecture de Xanthi (nord de la Grèce), le réseau de pièges a été établi dans des cultures de maïs, de riz et d'autres céréales (blé, avoine, orge) ainsi que des cultures de luzerne. Le ravageur n'a pas été découvert à ce jour à Xanthi.

Sources : pamth.gov.gr, gd.eppo.int.

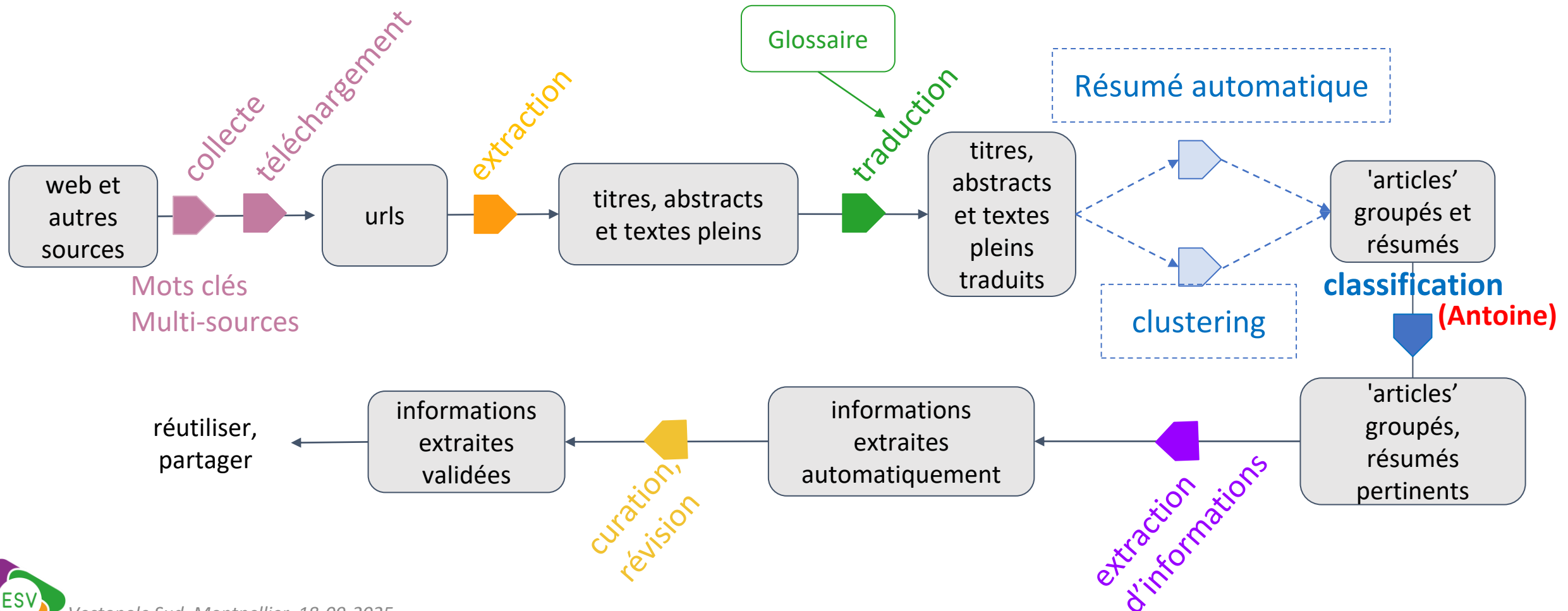


Figure 1 : Carte des détections de *Spodoptera frugiperda* dans l'union européenne, à noter que les zones notées en "Détecté" pour 2025 sont issues d'un document de 2025 dans lequel les dates de détection ne sont pas précisées. Sources pour les zones non relayées dans les BM précédents : pamth.gov.gr, gd.eppo.int.

texte

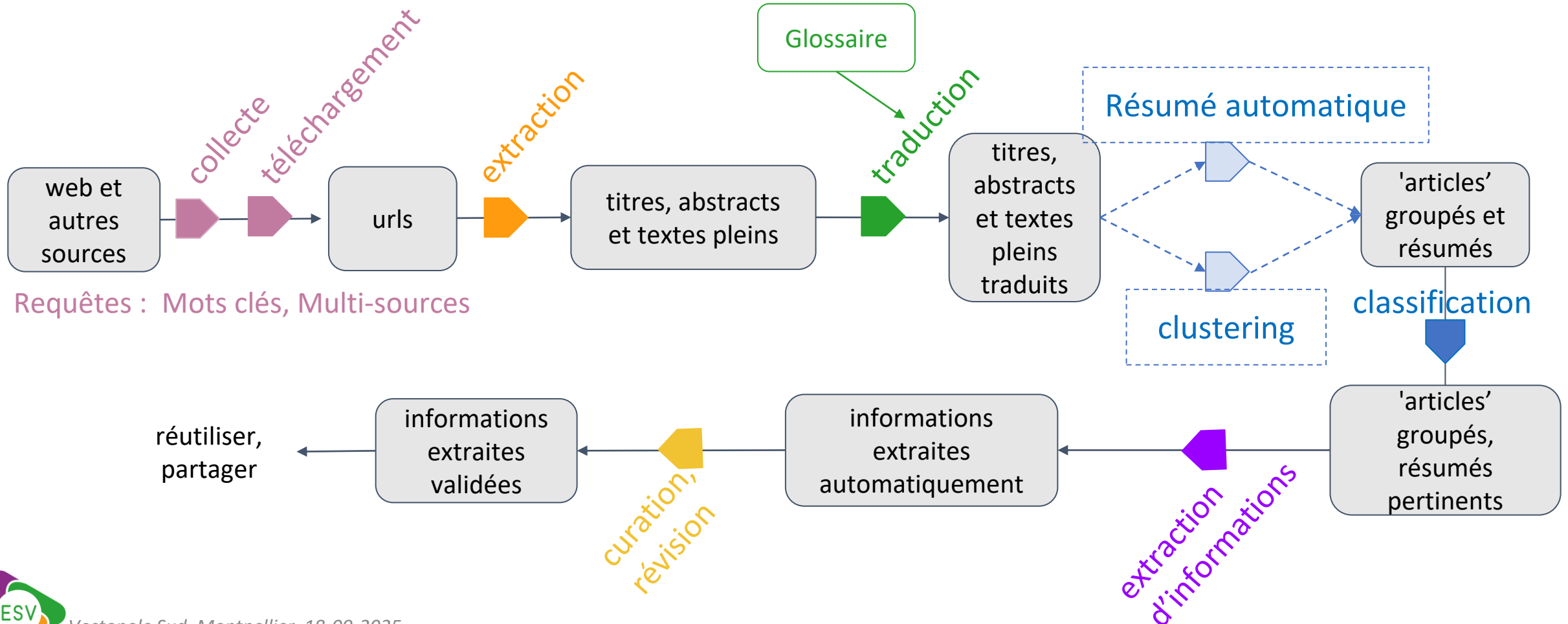
carte
interactive

PIPELINE VSI → Lancé tous les lundis matins



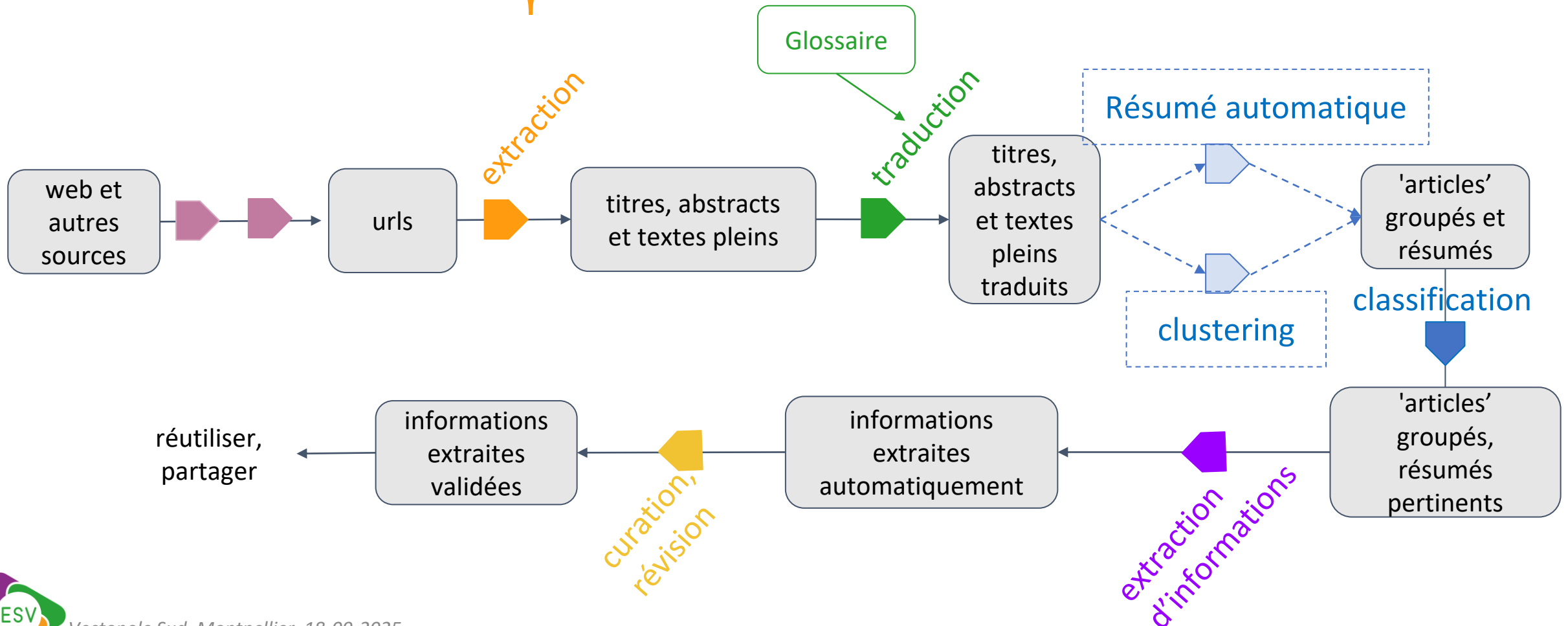
PIPELINE VSI

Collecte, téléchargement automatique des URLs avec scripts python
[package Scrapy pour le webscrapping et API scaleSERP pour les sources Google]



PIPELINE VSI

Extraction document via script python utilisant librairie Trafilatura + extraction contenu page web au format XML_TEI



Étape 1 : OUVERTURE d'UNE URL (exemple)



<https://www.vivicastellanagrotte.it/index.php/notizietop/16838-monitoraggio-xylella-tre-piante-positive-a-castellana-grotte>

HOME

WEBTV

DAL PALAZZO

CULTURA ▾

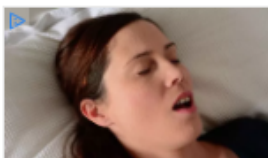
SPORT

GROTTE DI CASTELLANA ▾

LA CITTADINA ▾

NEWSLETTER

CONTATTI



Apnée du sommeil : un oreiller ingénieux fait fureur au France

NaturallyYou

OUVRIR



COMUNICATO STAMPA NOTIZIE 09 SETTEMBRE 2022

Monitoraggio xylella - Tre piante positive a Castellana-Grotte

Riceviamo comunicazione del rinvenimento di tre piante di ulivo positive nella periferia nord-est di Castellana-Grotte. Dopo Locorotondo, Monopoli, Polignano a mare e Alberobello - dunque - la Città

delle Grotte è il quinto Comune della Città Metropolitana di Bari nel quale sia stata rilevata la presenza del batterio xylella fastidiosa. Le tre piante, tra loro adiacenti, sono al confine tra la "Zona cuscinetto Polignano a mare" e la zona indenne. Il primo aggiornamento dopo la pausa estiva comprende anche altre 13 piante infette, due delle quali nella (ormai ex) zona indenne di Polignano a mare, altre 6 in zona contenimento (5 a Fasano e 1 a Martina Franca); le restanti cinque sono in zona cuscinetto a Polignano a mare (3), a Monopoli (1) e una ad Alberobello. In quest'ultimo caso si tratta del primo ritrovamento a ovest del centro abitato della capitale dei trulli, a circa 1 Km in direzione Noci. Le analisi, effettuate dal Centro di ricerca sperimentazione e formazione in agricoltura "Basile Caramia" di Locorotondo su campioni rilevati il 23 agosto 2022, sono datate al 29 agosto u.s.



◀ Indietro

Avanti ▶

Cerca

Q cosa cerchi?

SERVIZI ALLA
PERSONA

EVENTI
Musica • Teatro • Cinema

ARTIGIANI &
ARTIGIANATO

SALUTE E
BELLEZZA

PRODOTTI

Étape 2 : OUVERTURE
code source en HTML de la
page web



HOME WEBTV DAL PALAZZO CULTURA ▾ SPORT GROTTE DI CASTELLANA ▾ LA CITTADINA ▾ NEWSLETTER CONTATTI

```
129
130     <h2 itemprop="name">
131         Monitoraggio xylella - Tre piante positive a Castellana-Grotte
132     </h2>
133
134
135
136
137
138     <div itemprop="articleBody">
139         <div class="l7ghb35v kjdcldyq kmwttqpk gh25dzvf jikcssrz n3t5jt4f">Riceviamo comunicazione del rinvenimento di tre piante di ulivo positive nella periferia nord-est di Castellana
140     <a class="a2a_button_facebook"></a>
141     <a class="a2a_button_twitter"></a>
142     <a class="a2a_button_telegram"></a>
143     <a class="a2a_button_whatsapp"></a>
144     <a class="a2a_button_threads"></a>
145     <a class="a2a_button_email"></a>
146     <a class="a2a_button_print"></a>
147 </span>
148 </div> </div>
149
150
151
152 <ul class="pager pagonav">
153 <li class="previous">
154     <a class="hasTooltip" title="I.I.S.S. ''dell'Erba'' - A.S. 2022-2023: si comincia" aria-label="Articolo precedente: I.I.S.S. ''dell'Erba'' - A.S. 2022-2023: si comincia" href="#">
155         <span class="icon-chevron-left" aria-hidden="true"></span> <span aria-hidden="true">Indietro</span> </a>
156 </li>
```

◀ Indietro

Avanti ▶

SALUTE E
BELLEZZA

PRODOTTI
A KM 0

Étape 3 : EXTRACTION CONTENU TEXTUEL au format XML_TEI via Trafilatura



COMUNICATO STAMPA NOTIZIE
09 SETTEMBRE 2022

Monitoraggio xylella - Tre piante positive a Castellana-Grotte

Grotte

Riceviamo comunicazione del rinvenimento di tre piante di ulivo positive nella periferia nord-est di Castellana-Grotte. Dopo Locorotondo, Monopoli, Polignano a mare e Alberobello - dunque - la Città delle Grotte è il quinto Comune della Città Metropolitana di Bari nel quale sia stata rilevata la presenza del batterio xylella fastidiosa. Le tre piante, tra loro adiacenti, sono al confine tra la "Zona cuscinetto Polignano a mare" e la zona indenne. Il primo aggiornamento dopo la pausa estiva comprende anche altre 13 piante infette, due delle quali nella (ormai ex) zona indenne di Polignano a mare, altre 6 in zona contenimento (5 a Fasano e 1 a Martina Franca); le restanti cinque son in zona cuscinetto a Polignano a mare (3), a Monopoli (1) e una ad Alberobello. In quest'ultimo caso si tratta del primo ritrovamento a ovest del centro abitato della capitale dei trulli, a circa 1 Km in direzione Noci. Le analisi, effettuate dal Centro di ricerca sperimentazione e formazione in agricoltura "Basile Caramia" di Locorotondo su campioni rilevati il 23 agosto 2022, sono datate al 29 agosto u.s.



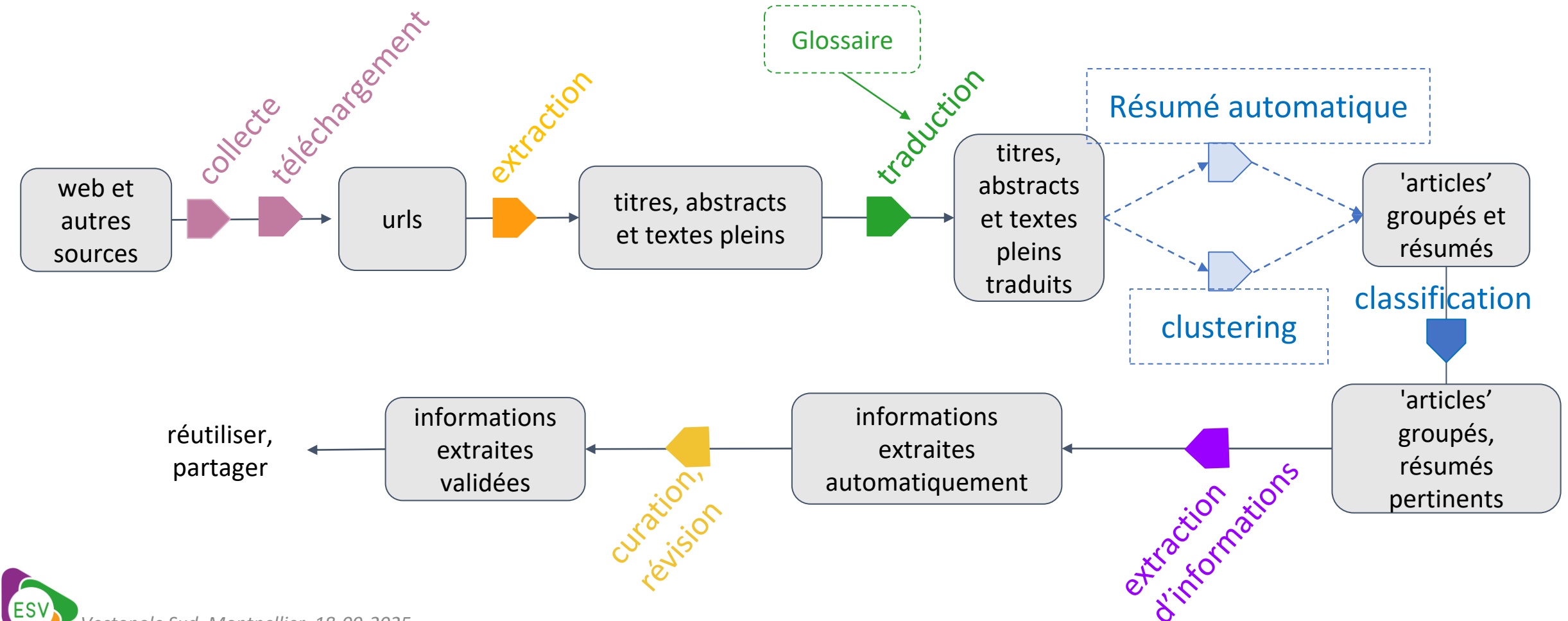
◀ Indietro

Avanti ▶

```
"<TEI xmlns="http://www.tei-c.org/ns/1.0">\n <teiHeader>\n <fileDesc>\n <titleStmt>\n <title type="main">Monitoraggio xylella - Tre piante positive a Castellana-Grotte</title>\n <author>Comunicato Stampa</author>\n </titleStmt>\n <publicationStmt>\n <p/>\n </publicationStmt>\n <notesStmt>\n <note type="fingerprint">ASwdniSgGiFYGI51Y9b3Kxau9AM=</note>\n </notesStmt>\n <sourceDesc>\n <bibl>Monitoraggio xylella - Tre piante positive a Castellana-Grotte. vivicastellanagrotte.it, 2022-09-09</bibl>\n <bibl type="sigle">vivicastellanagrotte.it, 2022-09-09</bibl>\n <biblFull>\n <titleStmt>\n <title type="main">Monitoraggio xylella - Tre piante positive a Castellana-Grotte</title>\n <author>Comunicato Stampa</author>\n </titleStmt>\n <publicationStmt>\n <publisher>vivicastellanagrotte.it (vivicastellanagrotte.it)</publisher>\n <ptr type="URL" target="http://www.vivicastellanagrotte.it/index.php/notizietop/16838-monitoraggio-xylella-tre-piante-positive-a-castellana-grotte"/>\n <date>2022-09-09</date>\n </publicationStmt>\n </biblFull>\n </sourceDesc>\n </fileDesc>\n <profileDesc>\n <abstract>\n <p>Riceviamo comunicazione del rinvenimento di tre piante di ulivo positive nella periferia nord-est di Castellana-Grotte. Dopo Locorotondo, Monopoli, Poligna...</p>\n </abstract>\n <textClass>\n <keywords>\n <term type="tags">webtv, web, tv, castellana, grotte, fanove, madonna della vetrana, puglia, bari</term>\n </keywords>\n </textClass>\n </profileDesc>\n <encodingDesc>\n <applInfo>\n <application version="1.3.0" ident="Trafilatura">\n <label>Trafilatura</label>\n <ptr target="https://github.com/adbar/trafilatura"/>\n </application>\n </applInfo>\n </encodingDesc>\n </teiHeader>\n <text>\n <body>\n <div type="entry">\n <p>Riceviamo comunicazione del rinvenimento di tre piante di ulivo positive nella periferia nord-est di Castellana-Grotte. Dopo Locorotondo, Monopoli, Polignano a mare e Alberobello - dunque - la Città delle Grotte è il quinto Comune della Città Metropolitana di Bari nel quale sia stata rilevata la presenza del batterio xylella fastidiosa.</p>\n <hi rend="H1">Le tre piante, tra loro adiacenti, sono al confine tra la "Zona cuscinetto Polignano a mare" e la zona indenne. Il primo aggiornamento dopo la pausa estiva comprende anche altre 13 piante infette, due delle quali nella (ormai ex) zona indenne di Polignano a mare, altre 6 in zona contenimento (5 a Fasano e 1 a Martina Franca); le restanti cinque son in zona cuscinetto a Polignano a mare (3), a Monopoli (1) e una ad Alberobello. In quest'ultimo caso si tratta del primo ritrovamento a ovest del centro abitato della capitale dei trulli, a circa 1 Km in direzione Noci.</hi>\n <p>Le analisi, effettuate dal Centro di ricerca sperimentazione e formazione in agricoltura "Basile Caramia" di Locorotondo su campioni rilevati il 23 agosto 2022, sono datate al 29 agosto u.s.</p>\n </div>\n <div type="comments">\n </div>\n </body>\n </text>\n</TEI>
```

PIPELINE VSI

RENDRE L'INFORMATION COMPREHENSIBLE : traduction automatique du titre et du texte plein (EN) via script python qui appelle API Google Translate

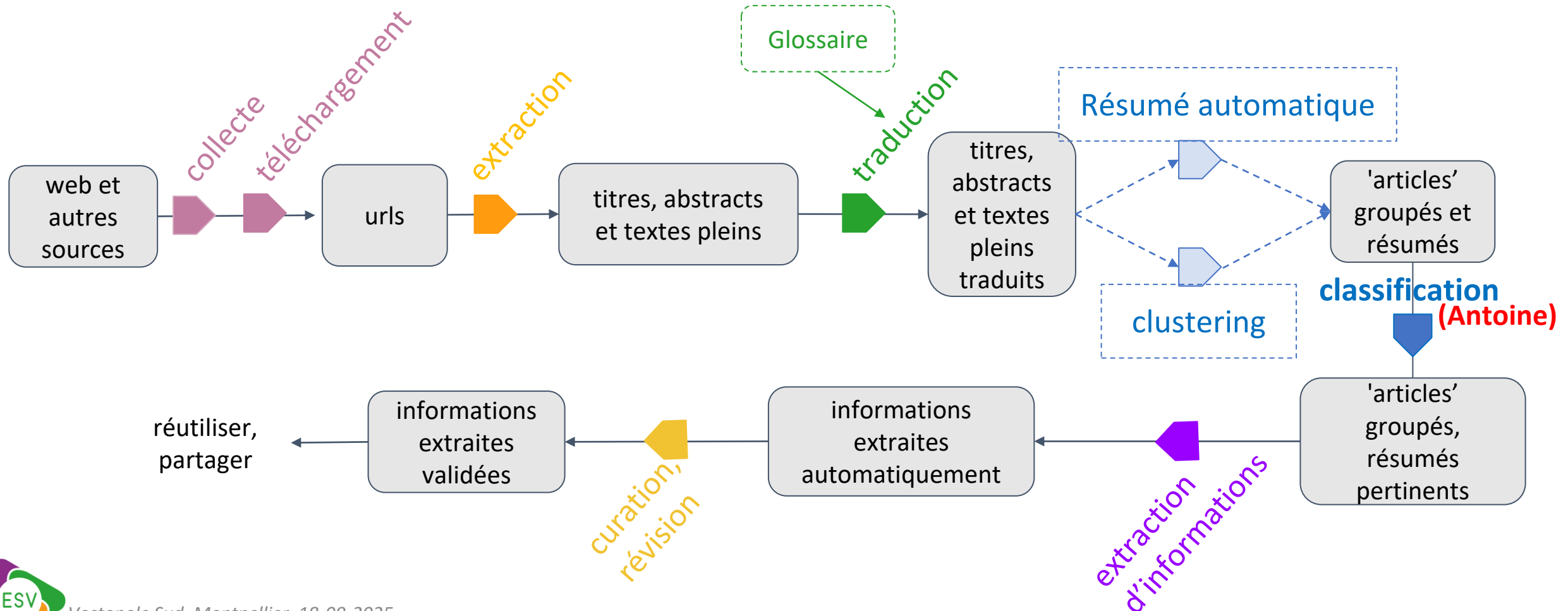


Titre publié (anglais)			Titre collecté (multi-langues)	
https://www.eleftheria.gr/peth...	true	Les platanes meurent dans la vallée de Tempi	Πεθαίνουν τα πλατάνια στην Κοιλάδα των Τεμπών - Ελευθερία	grec
https://33.fsvps.gov.ru/news/...	true	Under the supervision of the Office of Rosselkhoznadzor in Vladimir, measures have been taken to localize and eli...	Под контролем Управления Россельхознадзора во ...	russe
https://rezeknesnovads.lv/osu...	true	Surveillance de l'agrire du frêne dans la région de Rēzekne	Ošu smaragdzaļās krāšņvaboles monitorings ...	letton
https://rosleshoz.gov.ru/news/...	true	Emerald ash borer found in Krasnoarmeyskoye forestry in Saratov region	В Красноармейском лесничестве Саратовской области ...	
https://www.gov.pl/web/wior...	true	Control activities within the EABRACE project - monitoring of the emerald ash borer in the Podkarpackie Voivodesh...	Działania kontrolne w ramach projektu EABRACE ...	polonais
https://www.gov.me/clanak/la...	true	Analysis of grapevine leaf samples did not reveal the presence of the phytoplasma Flavescence dorée	Analizom uzoraka listova vinove loze nije utvrđeno ...	serbe
https://dentrolanotiziabreak.it/...	true	Agriculture alert: small, but devastating, beetle invasion in Piedmont - Inside the news	è piccolo, ma devasta il Piemonte, l'invasione del coleottero	italien
https://www.juntadeandalucia...	true	First detections of Popillia japonica in Galicia and France - RAIF	Primeras detecciones de Popillia japonica en Galicia y ...	espagnol
https://www.yvorne.ch/	true	Bienvenue sur le site Internet officiel	Commune d'Yvorne: Bienvenue sur le site Internet officiel	français
Ex. semaine 33 (2025) : 963 articles capture image OVIDE (IHM)/comité éditorial			Citrus Forum on pests and technology	
			es for the control of HLB in citrus plants was established.	
https://www.lagaceta.com.ar/...	true	HLB: Pilcomayo regained its vector-free status	HLB: Pilcomayo recuperó el status de área libre del vector	
https://www.fecier.org.ar/notic...	true	The citrus leafhopper, carrier of the dreaded HLB bacteria, is already roaming the orange groves of San Pedro and...	Federacion del Citrus de Entre Rios	
https://papers.ssrn.com/sol3/...	true	Identification of Soil Microbial Communities in Huanglongbing Infected Mandarin Orange Orchards Through Metag...	Identification of Soil Microbial Communities in ...	anglais
https://www.frontiersin.org/jo...	true	Frontiers The Application of Bacillus amyloliquefaciens and Arbuscular Mycorrhizal Fungi Displays Curative Effect...	The Application of Bacillus amyloliquefaciens and ...	
https://gdqinshuoyuan.com/3...	true	General Administration of Customs Announcement No. 165 of 2025 (Announcement on Plant Quarantine Requirem...	海关总署公告2025年第165号 (关于中国柚子出口新西兰植物 ...	chinois
https://www.sciencedirect.co...	true	Explainable ensemble learning for predicting pine wilt disease spread	Explainable ensemble learning for predicting pine wilt ...	
https://www.haber236.com/ya...	true	Study on leaf blight disease	Yaprak yanıklığı hastalığına yönelik çalışma	turc

PIPELINE VSI

TRIER-MODIFIER-CLASSER LES INFOS PERTINENTES :

- **Classification automatique (présentation Antoine)**
- Ecriture d'un résumé automatique du texte plein (à tester en routine)
- Clustering automatique des doublons (à tester en routine).
- Modification via interface OVIDE (IHM)



PIPELINE VSI

- **résumé automatique** via DistilBART (model encodeur-décodeur) *(à tester en routine)*

- **clustering automatique** des doublons via algo DBSCAN *(à tester en routine)*

Type de sujet
Scientifique

Paragraphe sur le sujet

Au Portugal, la force du secteur oléicole repose sur le riche patrimoine génétique des cultivars traditionnels. Cette étude visait à évaluer la réponse olfactive de P. spumarius à cinq cultivars d'oliviers traditionnels portugais : « Cobrançosa », « Negrinha de Freixo », « Santulhana », « Madural » et « Verdeal Transmontana » au cours de deux saisons distinctes (printemps et automne). Ces résultats ont montré que P. spumarius présentait des réponses olfactives significativement différentes envers les différents cultivars étudiés.

Date d'évènement
Nécessite d'être de type évènement

Mettre à jour Nouveau sujet

Filtrer les sujets par Paragra... Type --- Types de publication Éléments observés Période Date de ... 21/08/2025 Date de fin Export Réinitialiser

PARAGRAPHE	DATE DE CRÉATION	TYPE DE SUJET	DATE D'ÉVÈNEMENT
Vidéo d'un symposium sur la situation de X. fastidiosa dans les îles Baléares (date?)	25/08/2025	Actualité	
Un nouveau cas de la sous-espèce Xylella pauca a été identifié à Giovinnazzo, dans la zone de « Pietre rosse ...	25/08/2025	Sanitaire	
Plusieurs nouveaux foyers de Xylella fastidiosa subsp. pauca ST53 ont été découverts dans les campagnes ...	25/08/2025	Sanitaire	
Au Portugal, la force du secteur oléicole repose sur le riche patrimoine génétique des cultivars traditionnels. ...	25/08/2025	Scientifique	

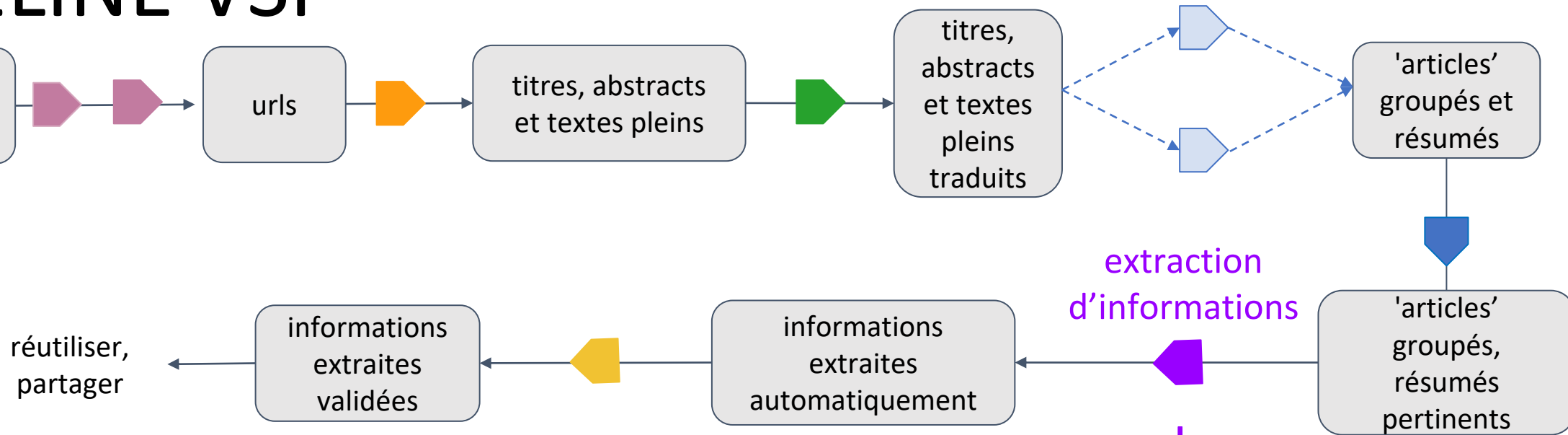
ID	URL	ENVOYÉ AU CE	TITRE PUBLIÉ
254509	https://www.letemps.ch/scienc...	true	Terreur des vignobles, le scarabée japonais débarque en Suisse romande
254507	https://www.rts.ch/info/regions/...	true	Un premier foyer de scarabée japonais identifié à Genève
253966	https://www.watson.ch/fr/suiss...	false	Un foyer de cet insecte nuisible identifié à Genève
253964	https://www.tdg.ch/geneve-un-...	false	Un premier foyer de cet insecte invasif identifié dans le canton de Genève
253960	https://www.letemps.ch/scienc...	false	Terreur des vignobles, le scarabée japonais débarque en Suisse romande - Le Temps

PIPELINE VSI

[https://doi.org/10.57745/ZDNOGF
et https://doi.org/10.57745/YKSEPY]



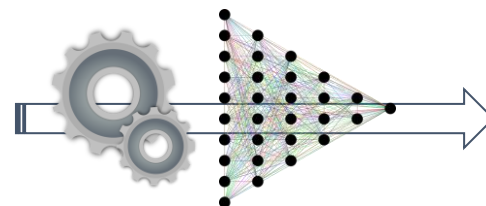
Vectopole Sud, Montpellier, 18-09-2025



Campagne d'Annotation par des experts* pour assister la VSI par des méthodes d'Intelligence Artificielle, notamment via des outils issus de la recherche en traitement automatique de la langue (TAL) ? **Corpus Epidemiomonitoring Of Plant (EPOP)** [Dataverse Plateforme ESV] ? **Challenge CLEF** (Conference and Labs of the Evaluation Forum) 2025 ? **Modèle prédictif obtenu par apprentissage automatique** (à implémenter)

Abstract The cyanobacterium *Planktothrix rubescens* Anagnostidis & Komarek (previously *Oscillatoria rubescens* DC ex Gomont) is present in several Italian lakes and it is known to produce cyanotoxins. The dynamics and toxin production of *P. rubescens* population in Lake Albano, a volcanic crater lake in Central Italy, has been studied for 5 years (January 2001-April 2005). Winter-

Article non-annoté



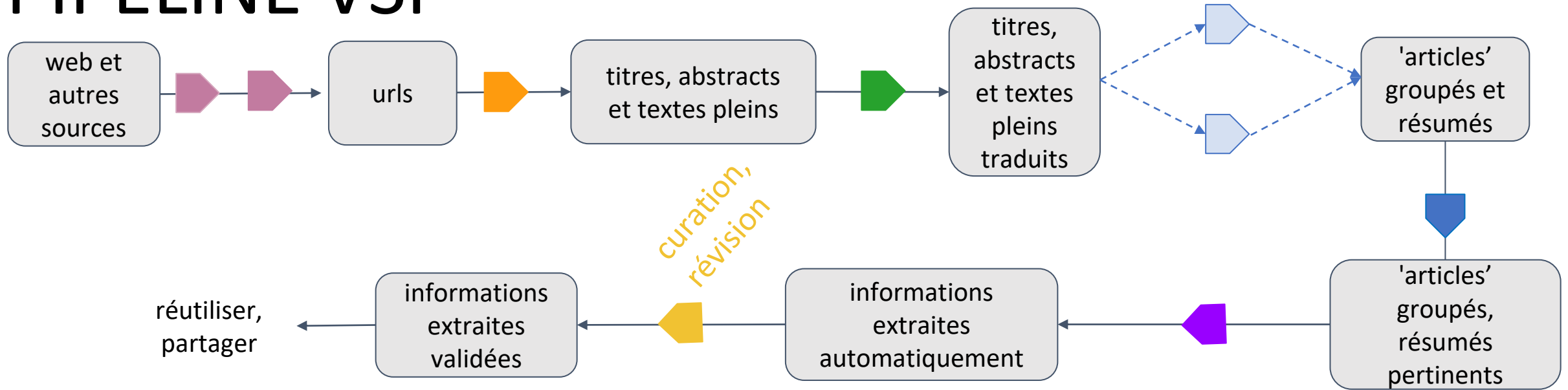
Modèle prédictif

Abstract The cyanobacterium *Planktothrix rubescens* Anagnostidis & Komarek (previously *Oscillatoria rubescens* DC ex Gomont) is present in several Italian lakes and it is known to produce cyanotoxins. The dynamics and toxin production of *P. rubescens* population in Lake Albano a volcanic crater lake in Central Italy, has been studied for 5 years (January 2001-April 2005). Winter-

Article annoté automatiquement

(Execution modèle et enregistrement données dans BDD via API)

PIPELINE VSI



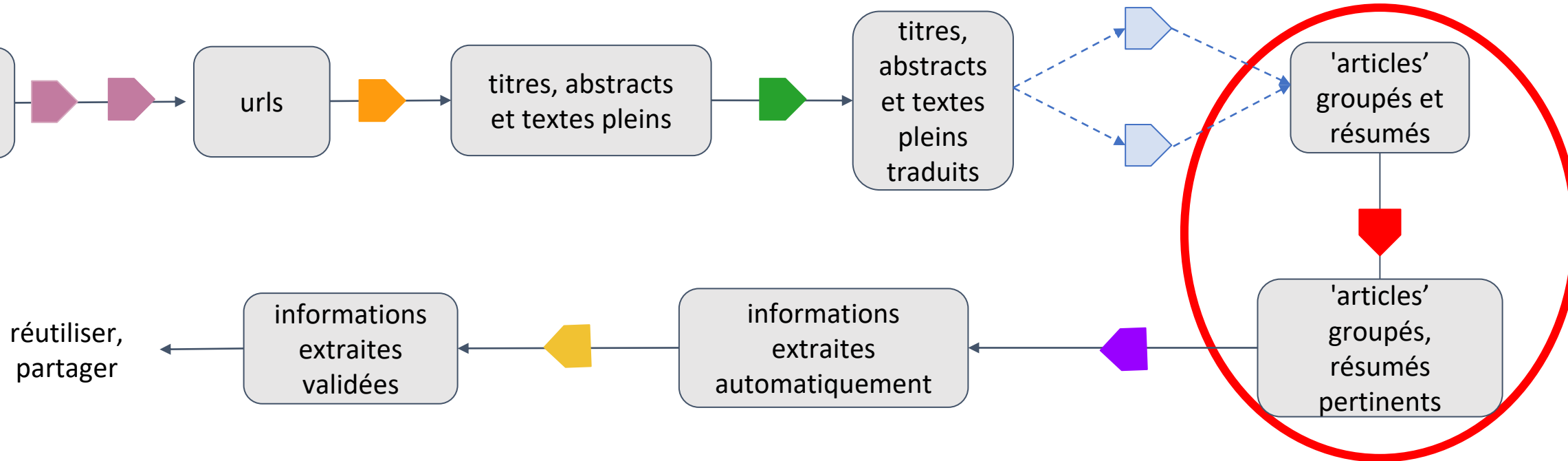
- Comité éditorial via interface OVIDE – production manuelle des bulletins (scriptR)

PERSPECTIVES

- Production automatique des bulletins.
- Mise en place d'un moteur de recherche VSI directement associé à OVIDE - production de bulletins à la carte
- Synthèse spatio-temporelle automatique (cartographie) automatique
-

CLASSIFIEUR

Présentation



Classification : sélection des articles pertinents

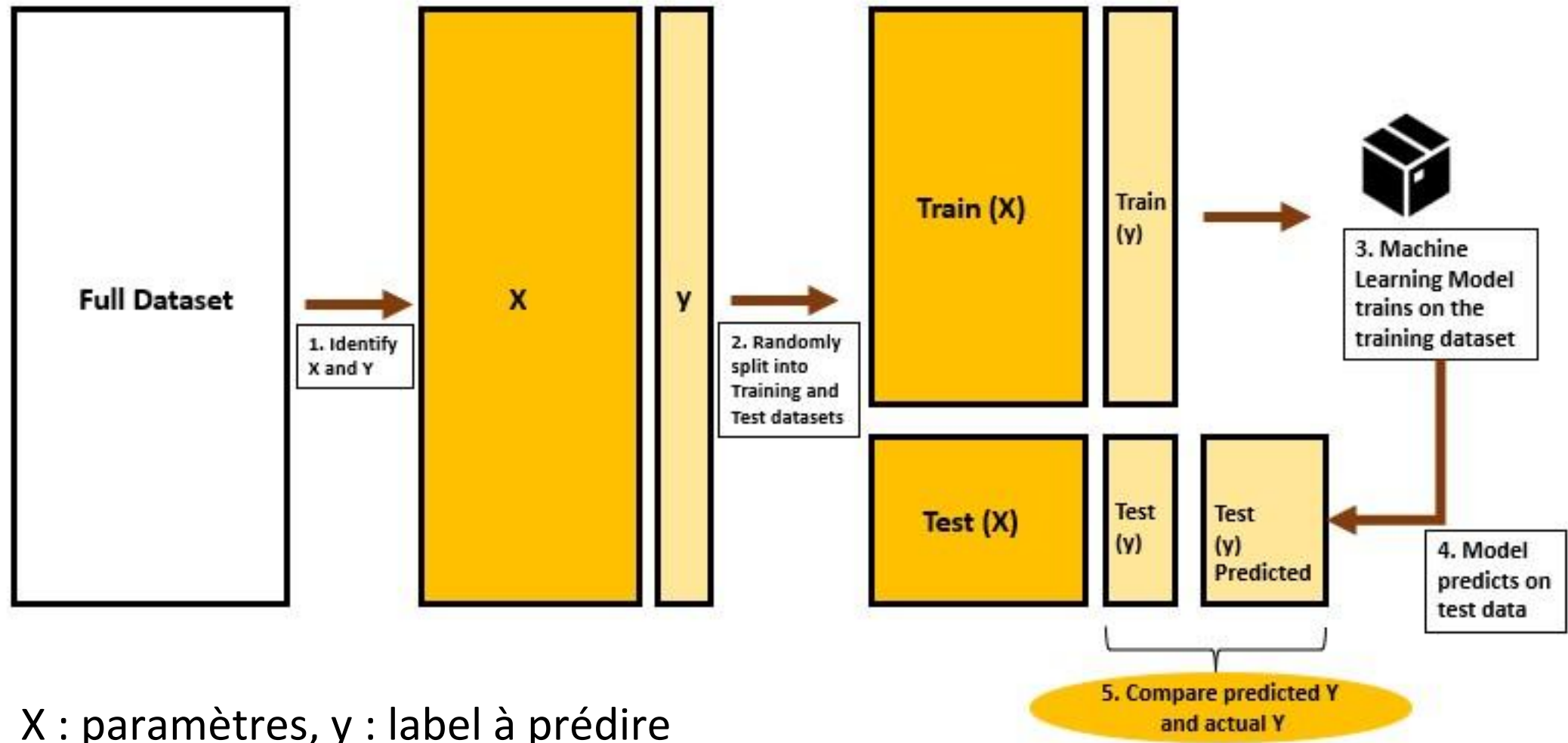
Environ 1000 articles collectés par semaine, dont ~15% d'articles pertinents.

~1 jour de travail par semaine par veilleuse —> objectif automatisation partielle

Grandes quantités de données labellisées par les veilleuses

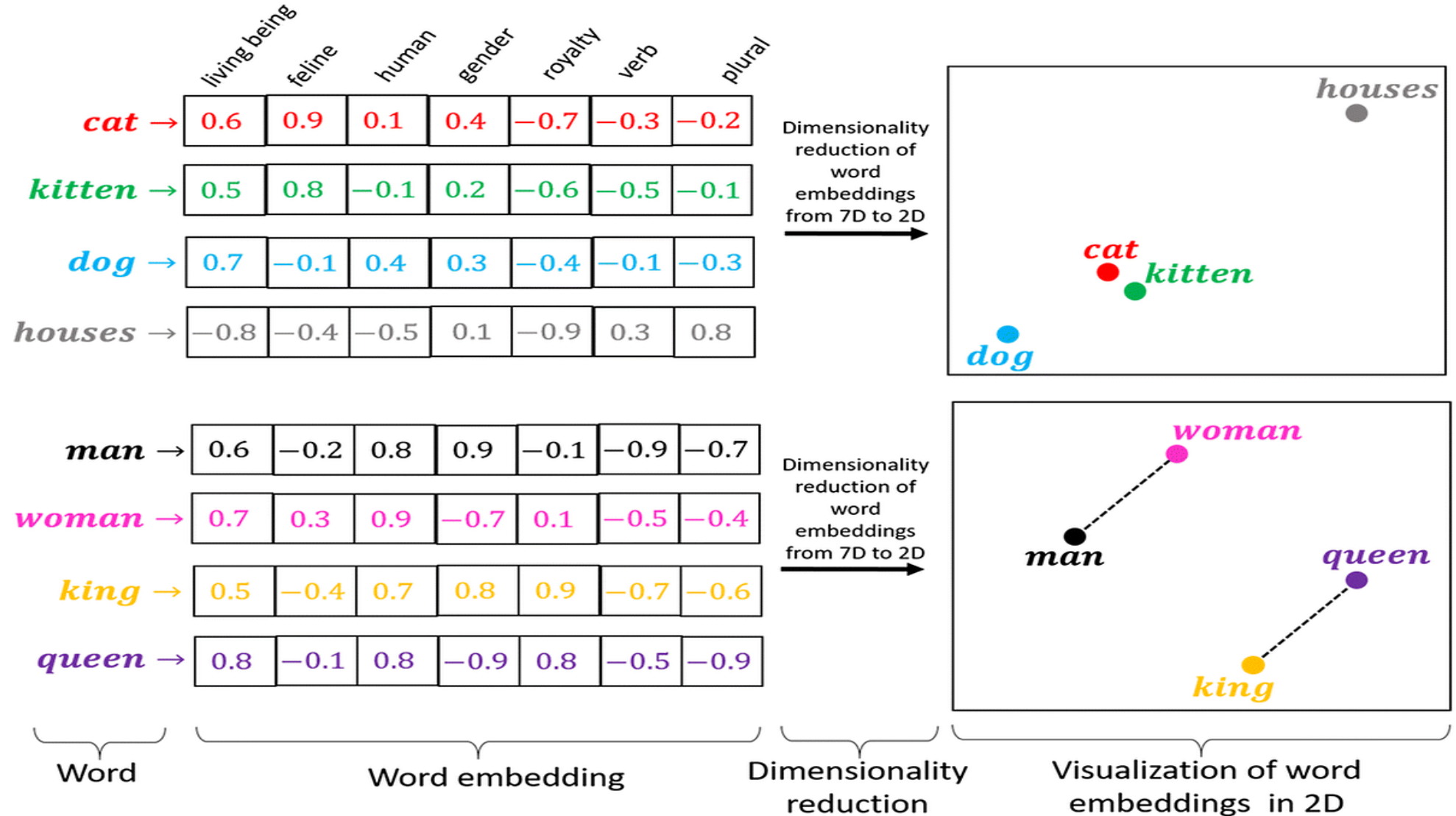
CLASSIFIEUR

Présentation : approche machine learning



CLASSIFIEUR

Entraînement : word embeddings



CLASSIFIEUR

Présentation : exemples

HLB: Se erradicaron 19 plantas positivas en dos localidades de la provincia de Corrientes

Lo hizo el Senasa en el marco de su Programa para la prevención de la enfermedad más destructiva de los cítricos.

08 de octubre de 2024



ID 217095 : **Pertinent**



Submit to this Journal

Review for this Journal

Propose a Special Issue

Article Menu

Academic Editor



Hui Sun

Subscribe SciFeed

Related Info Link

More by Author Links

Article Views

668

IK

Order Article Reprints

Open Access

Opinion

Two Infectious Diseases: "COVID-19" and "Pine Wilt Disease"

by Kazuyoshi Futai

Kyoto University, Kyoto 606-8501, Japan

Forests 2024, 15(10), 1724; <https://doi.org/10.3390/f15101724>

Submission received: 21 August 2024 / Revised: 27 September 2024 / Accepted: 27 September 2024 /

Published: 29 September 2024

(This article belongs to the Section Forest Health)

Download

Browse Figures

Versions Notes

Abstract

Two epidemics, a new coronavirus disease (hereafter COVID-19) and pine wilt disease (hereafter PWD) whose causal agent is the pine wood nematode *Bursaphelenchus xylophilus*, are similar in terms of disease spread, particularly in an important role of the latent carrier in the spread of the disease. In Japan, damage caused by PWD has gradually decreased over the past 40 years, not because of successful control of PWD, but because pine forests susceptible to pine wilt have been drastically reduced by the disease. Meanwhile, COVID-19, which was first identified in China in 2019, became a pandemic in 2020 and caused long-term disruptions to people's lives worldwide. Even after the declaration of the end of the emergency by the WHO in May 2023, the coronavirus continues to produce new variants and maintains the potential for new waves of infection. In this paper, I would like to discuss the issues of control measures for the forest disease PWD, especially for the measures against asymptomatic carrier trees, with reference to the efforts implemented for COVID-19, which is more closely related to human society. This is because an asymptomatic carrier has been a blind spot in conventional PWD control

ID 214962 : **Non pertinent**

CLASSIFIEUR

Entraînement : Pré-traitement des données

- Label 1 si pertinent, 0 sinon
- Concaténation du title + abstract + text (pas de lien / nom du site web)
- Tri des lignes vides —> environ 16% des articles, pertinents par défaut
- « Accept cookies » ou « javascript cant download » gardés
- Si deux articles identiques/similaires ont des labels différents, ils sont relabellisés 1

CLASSIFIEUR

Entraînement

1	id	date	lang	url	input	labels
2	213331	2024-09-16T05	es	https://www.elespanol.com/alica	Farmers in Alicante are furious over the government's "permissiveness" in the "avalanche of pests" from South Africa [SEP] The employers' association Asaja criticises Madrid's "passivity" in the face of phytosanitary problems affecting citrus fruits and calls for the Mercosur agreement to be reconsidered. Farmers in Alicante are outraged by the government's "permissiveness" in the face of the	0
3	213814	2024-09-16T05	es	https://www.todoalicante.es/ecor	Farmers demand cold treatment for citrus fruits arriving from Africa to stop the pests that ravage Alicante crops TodoAlicante [SEP] La Unió estimates that the last six diseases reported in the last 15 years have led to a 40% increase in production costs. Sections Services We highlight You need to be registered to access this feature. Sharing options The Banski spider or the South African thrips are just two of the latest pests that are ravaging	1

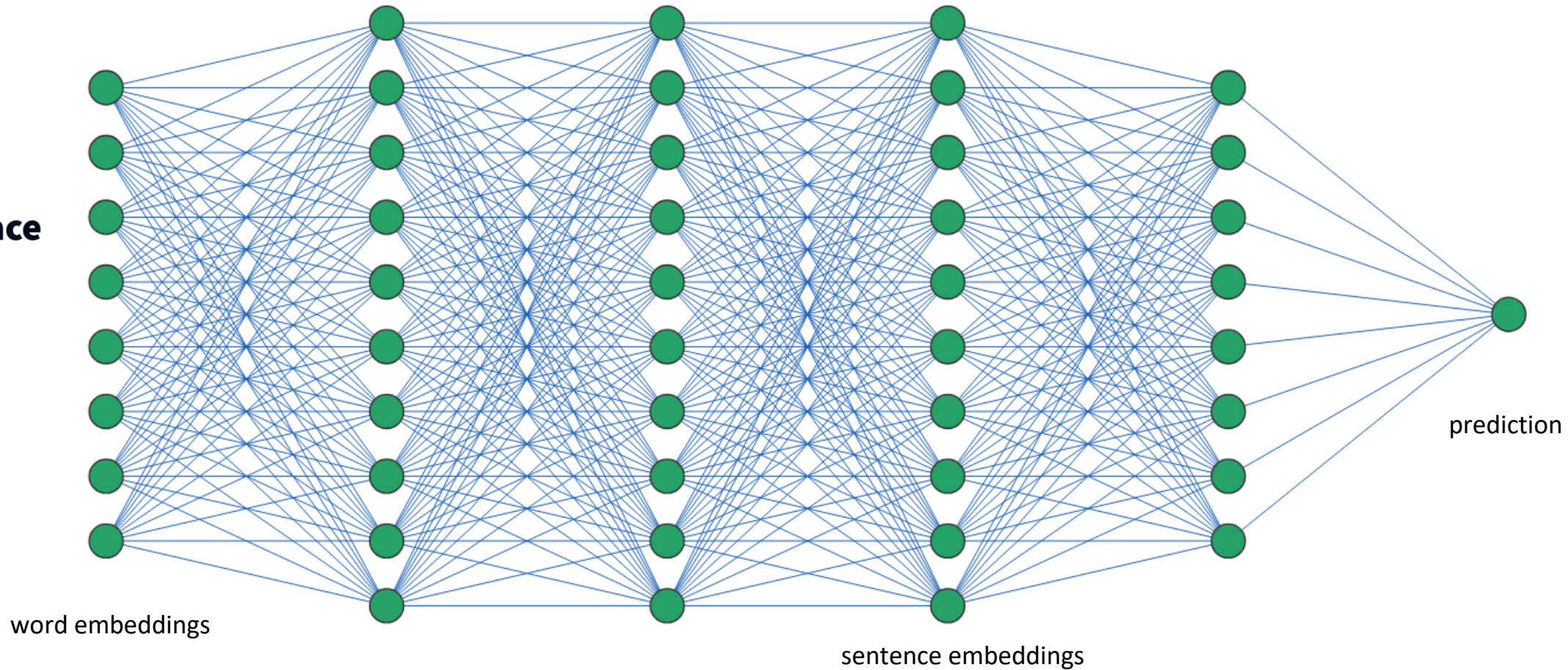
CLASSIFIEUR

Entraînement : architecture testées

- fine-tuning complet (100M paramètres)



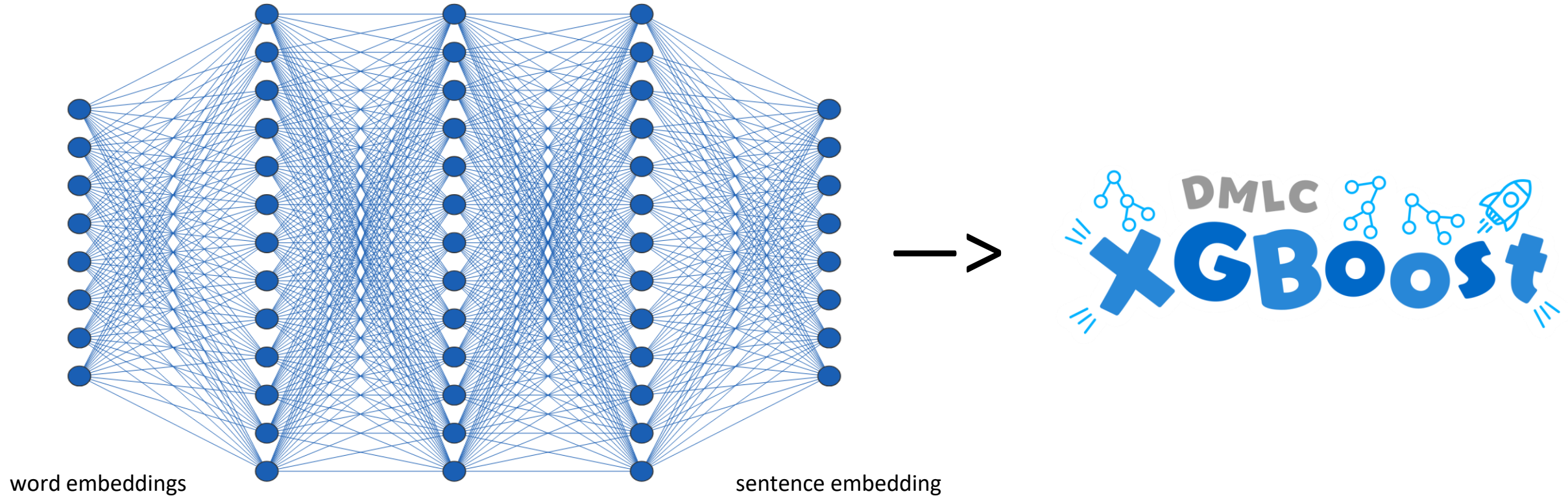
Hugging Face



CLASSIFIEUR

Entraînement : architecture testées

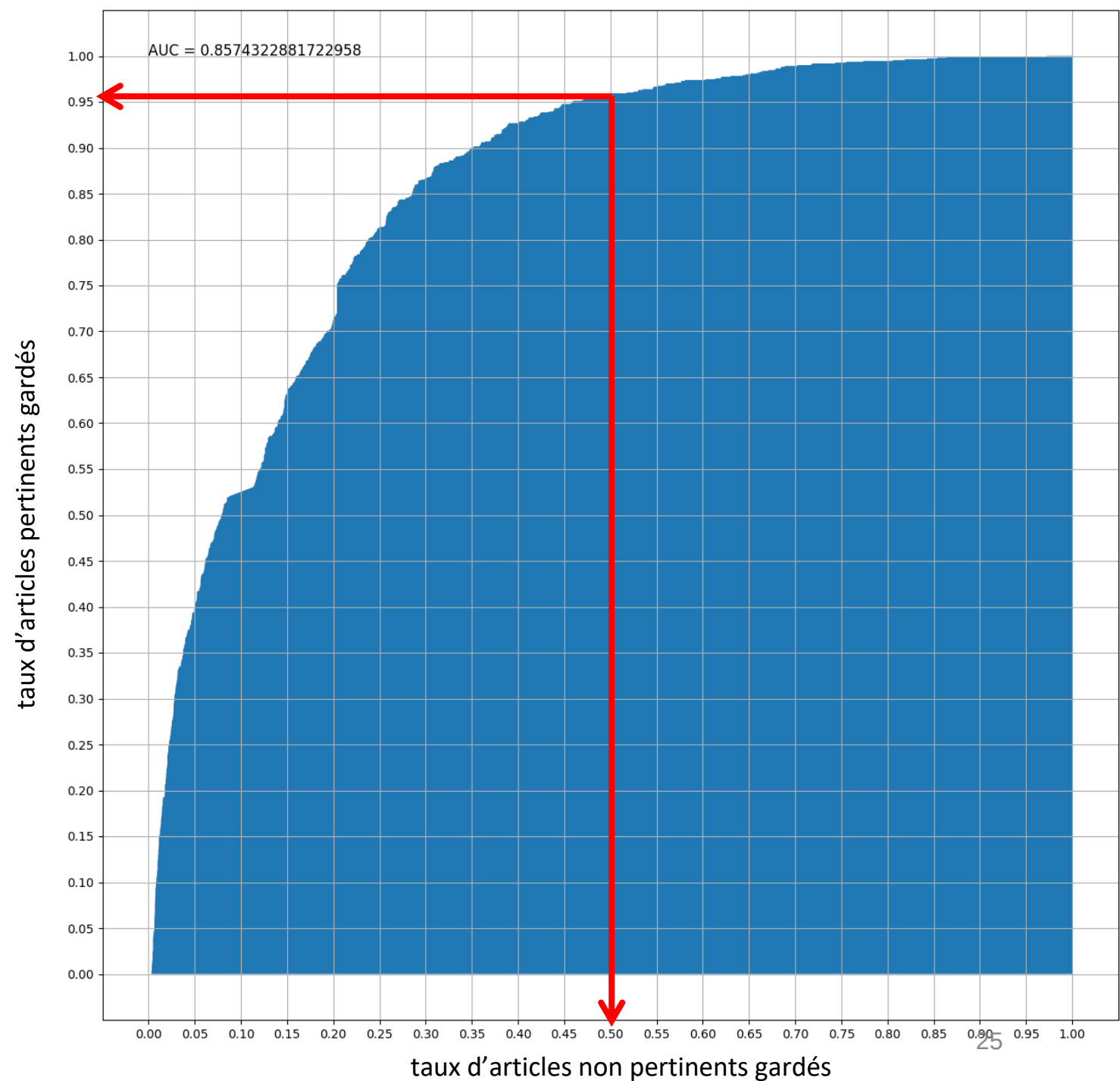
- modèle de langue pré-entraîné (1B paramètres + entraînement local d'un modèle XGBoost)



CLASSIFIEUR

Test : Résultats

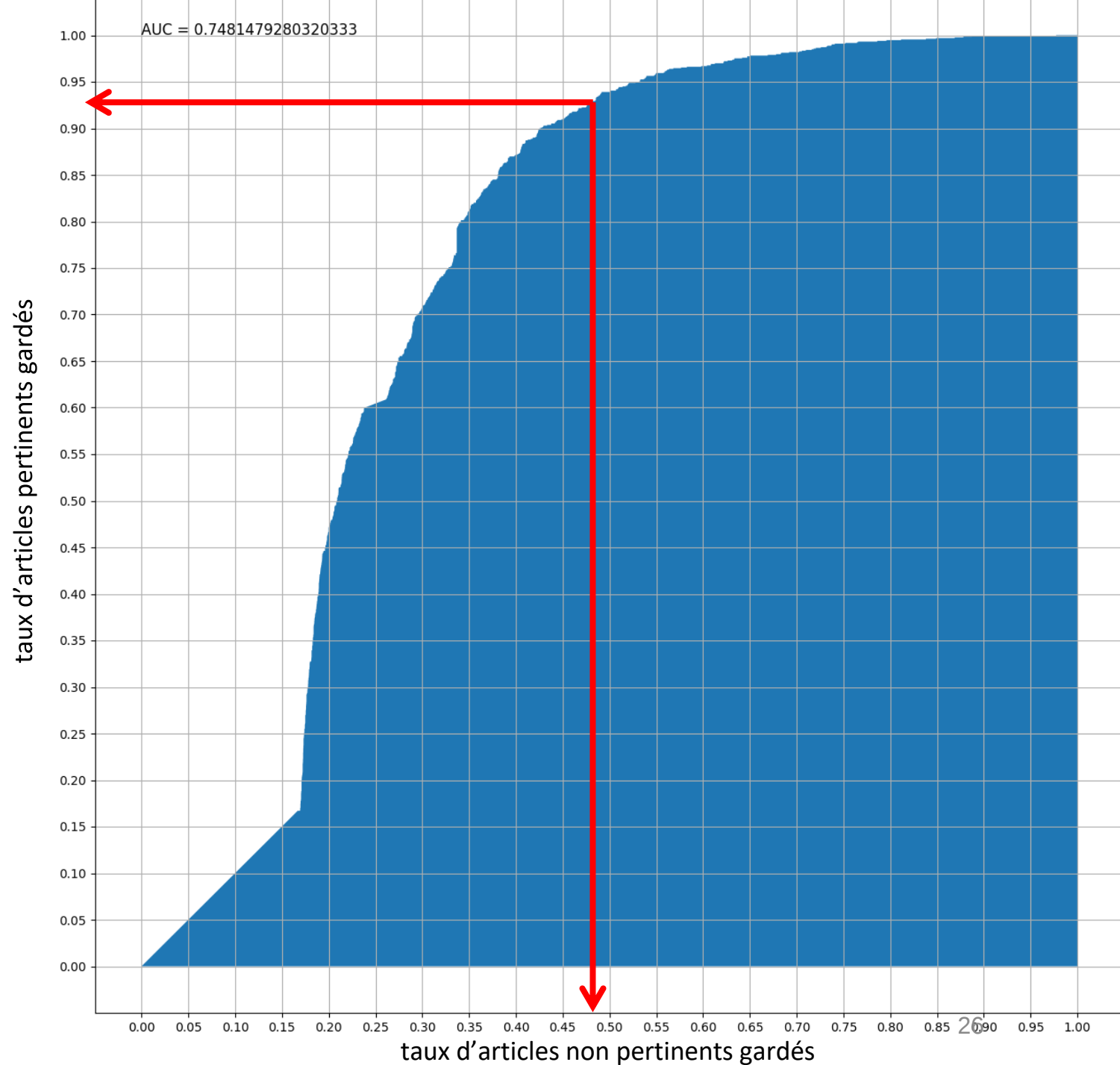
4% perte d'articles pertinents
50% articles non pertinents éliminés



CLASSIFIEUR

Test : Résultats (lignes vides incluses)

6% perte d'articles pertinents —> 47%
articles non pertinents éliminés



CLASSIFIEUR

Validation : Sélection du seuil discriminant

input	label	prediction	%_articles_pertinents_perdus	%_articles_non_pertinents_éliminés
NT+ Local Authorities & Construction [SI	1	0,0057	0,20%	0,50%
NT+ Local Authorities & Construction [SI	1	0,0057	0,41%	0,57%
Cantons - imTicker staht's [SEP] Top nev	1	0,0069	0,61%	14,88%
Greenery Scanner, the electronic eye to	1	0,0076	0,82%	21,95%
Destination Canal du Midi Facebook [S	1	0,0076	1,02%	22,17%
Salmonella in imported cucumbers. 18 p	1	0,0081	1,22%	23,49%
Multiscale Modelling of European Beech	1	0,0083	1,43%	24,18%
Transcriptomic response of citrus psyllid	1	0,0091	1,63%	28,29%
Webinar – Let's Talk About Planted Fore	1	0,0097	1,84%	29,29%
Emergency permit granted in the fight ag	1	0,0111	2,04%	30,39%
The State supports the agricultural trans	1	0,0126	2,24%	31,08%
Biological control of citrus pests: A syste	1	0,0149	2,45%	31,30%
Longhorn beetles from postharvest wood	1	0,0162	2,65%	33,42%
Influence of Different Mango Phenologic	1	0,0169	2,86%	36,86%
Flavescence dorée, regional tender work	1	0,0187	3,06%	37,27%
Moderate Phosphorus Addition to Field-c	1	0,0195	3,27%	37,58%
Update of the delimited area for the pr				
Information Session: What is Xylella fa				
ToBREV virus in Poland - assessment				

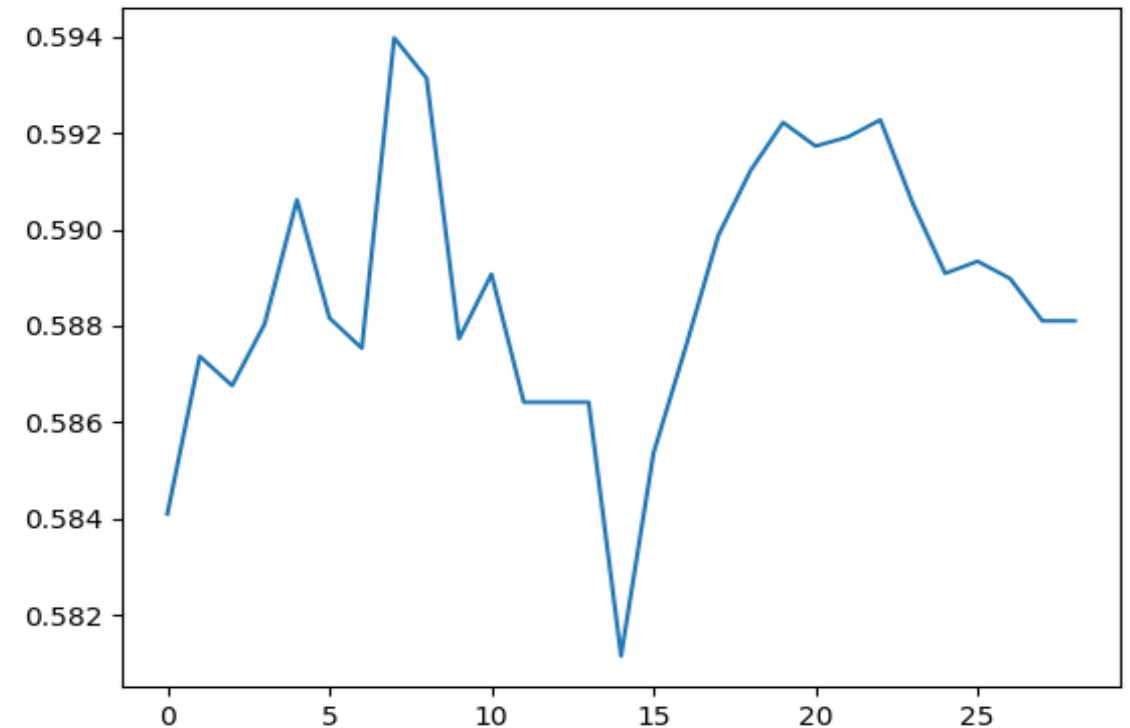
Acceptable de ne pas voir ces 13 articles pour réduire le volume à traiter d'un tiers ?

CLASSIFIEUR

Mise en production : amélioration continue

- Réorganisation de la BDD VSI, intégration à la pipeline VSI et à OVIDE
- Prévention d'une dérive du modèle (10% d'articles négatifs ré-échantillonnés)
- suivi des performances
- Ré-entraînement régulier

Évolution de l'AUC en fonction du nombre de semaines



CLASSIFIEUR

Mise en production : tracking des expériences

- Framework dédié au machine learning pour suivre les expériences (optimisation des hyper-paramètres, historique)



				Metrics	Parameters			
☐	Run Name	Created ⌵	Duration	AUC	embedding_model	learning_rate	n_estimators	run_date
☐	shivering-asp-446	✓ 1 month ago	26.3s	0.83726205...	NovaSearch/jasper_en_vision_language_v1	0.1	900	2025-07-11 17:51:38
☐	charming-bass-54	✓ 1 month ago	29.7s	0.85459902...	NovaSearch/jasper_en_vision_language_v1	0.1	900	2025-07-11 17:51:12
☐	righteous-duck-449	✓ 1 month ago	35.3s	0.83560318...	NovaSearch/jasper_en_vision_language_v1	0.05	900	2025-07-11 17:50:42
☐	monumental-elk-366	✓ 1 month ago	38.0s	0.85618435...	NovaSearch/jasper_en_vision_language_v1	0.05	900	2025-07-11 17:50:07
☐	redolent-colt-661	✓ 1 month ago	22.3s	0.83645208...	NovaSearch/jasper_en_vision_language_v1	0.1	600	2025-07-11 17:49:29
☐	valuable-gnat-801	✓ 1 month ago	24.3s	0.85375877...	NovaSearch/jasper_en_vision_language_v1	0.1	600	2025-07-11 17:49:06
☐	sincere-geese-721	✓ 1 month ago	30.4s	0.83460830...	NovaSearch/jasper_en_vision_language_v1	0.05	600	2025-07-11 17:48:42
☐	legendary-snipe-945	✓ 1 month ago	31.8s	0.85457198...	NovaSearch/jasper_en_vision_language_v1	0.05	600	2025-07-11 17:48:12

CLASSIFIEUR

Mise en production : test grandeur nature juillet-août 2025

Classifieur comparaison veilleuses	Nombre d'articles
Vrai Positif	147
Vrai Négatif	369
Faux Positifs	424
Faux Négatifs	8

Semaines 28 à 35 (2025)
~ **999 articles** /semaine

CLASSIFIEUR

Conclusion

- Données d'entraînement de qualité en quantité —-> bons résultats
- Meilleurs résultats avec des modèles pré-entraînés
- Cependant, modèles open source de taille raisonnable suffisants
- Résultats satisfaisants, passage en production définitif semaine prochaine ?

CLASSIFIEUR

Perspectives

- Ajout d'ONs pour voir le changement envisageable
- Périodicité de ré-entraînement
- Curation des données d'entraînement
- Amélioration du modèle open source
- Modèle plus gros ? Nouvelle architecture (RAG) ?
- Amélioration de la pipeline en amont



Isabelle Pieretti (animatrice du GT-VSI et veilleuse)



Marie Grosdidier (animatrice du GT-VSI et veilleuse)



Sandy Dupérier (veilleuse)



Jean-Baptiste Louvet (informaticien)



Simon Nicoux (informaticien)



Antoine Marullaz (Ingénieur en traitement automatique de données de télédétection et textuelles)

MERCI



isabelle.pieretti@cirad.fr
antoine.marullaz@inrae.fr