



Méthode de construction de bases de connaissances à partir d'alignements entre ontologies de référence : Annotation des Bulletins de Santé du Végétal

**Tayeb Ghorfi^{1,2}, Catherine Roussey¹, Haïfa Zargayouna²,
Adeline Nazarenko²**

¹ UR TSCF, Irstea, 9 av. Blaise Pascal CS 20085, 63172 Aubière, France
prénom.nom@irstea.fr

² LIPN, UMR CNRS 7030 Institut Galilée - Université Paris 13 99, avenue
Jean-Baptiste Clément 93430 Villetaneuse *prénom.nom@univ-paris13.fr*





Sommaire

- 1 **Sujet de thèse**
 - Cadre
 - Objectif
- 2 **Processus d'annotations des Bulletins de Santé du Végétal**
 - Bulletins de Santé du Végétal
 - Annotations Thématico Spatio-Temporelles
 - Annotation du contenu des BSV
- 3 **Conclusion**



I. Sujet de thèse

1. Cadre

- Diminuer l'usage des produits chimiques
- Passer à une agriculture durable
- Disponibilité des données

Enjeux

- Santé publique
- Ecologie
- Economie et production

Solution

Détection d'alertes agricole



I. Sujet de thèse

2. Objectif

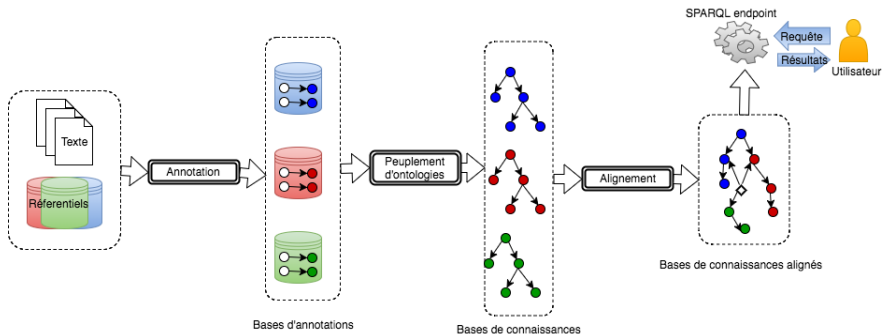
Comment détecter une alerte agricole ?

- Construire et rassembler des référentiels (Thésaurus, Ontologies, volubulaires ...) du domaine de l'agriculture.
- Collecte d'un corpus de bulletins d'information agricole : **Bulletins de Santé du Végétal**
- Annotation des Bulletins de Santé du Végétal
- Construction de différentes bases de connaissances :
 - Les principaux agresseurs des cultures cultivées en France
 - Les observations des stades de développements des cultures
 - Les niveaux de risque des attaques d'une culture.
- Alignement des bases de connaissances
- Interogation de la base de connaissance

L'observation d'une attaque d'une culture par un agresseur est une relation **N-aire** liant **une culture** ayant atteint **un stade de développement** à **un agresseur** donné.

I. Sujet de thèse

2. Objectif






II. Processus d'annotations des Bulletins de Santé du Végétal

1. Bulletins de Santé du Végétal

- 1 Sujet de thèse
 - Cadre
 - Objectif
- 2 Processus d'annotations des Bulletins de Santé du Végétal
 - Bulletins de Santé du Végétal
 - Annotations Thématico Spatio-Temporelles
 - Annotation du contenu des BSV
- 3 Conclusion




II. Processus d'annotations des Bulletins de Santé du Végétal

1. Bulletins de Santé du Végétal


- Alertes sur les attaques des cultures et les stades de développement des cultures.
- Archive publiée dans le LOD : collection de bulletins de toutes les régions de France et sur plusieurs années. Cette archive a été réalisé en lien avec le projet Vespa par Irstea [Roussey et al., 2016].
- Document pdf décrivant l'état des cultures dans chaque région.

II. Processus d'annotations des Bulletins de Santé du Végétal

1. Bulletins de Santé du Végétal



BULLETIN DE SANTÉ DU VÉGÉTAL
BOURGOGNE
Grandes cultures n° 14 du 22 février 2011



Colza

A RETENIR CETTE SEMAINE

- Lente progression des stades de colza (stade C2 sur 36% des parcelles)
- Régression des captures de charançon de la tige.
- Conditions météorologiques annoncées pour la semaine a priori peu favorables aux vols des ravageurs.

Réseau 2010-2011

Sur les 74 parcelles que constituent actuellement le réseau colza 39 ont fait l'objet d'au moins une observation cette semaine.

Stade des colzas

Les stades évoluent lentement :

C1 reprise de végétation : apparition de jeunes feuilles	: 64%
C2 entre-nœuds visibles : étrangement vert clair à la base des nouveaux pétioles	: 36%


Les conditions climatiques de la semaine dernière avec des températures peu élevées n'ont pas permis de faire évoluer les stades de colza.

Il est urgent de remettre en place les cuvettes jaunes.

Charançon de la tige du colza

3 parcelles renseignées cette semaine

Des captures ont été enregistrées cette semaine dans moins d'un tiers des cuvettes du réseau (principalement dans l'Yonne et la Nièvre). En cas de capture, on dénombre 3,9 charançons en moyenne par cuvette avec un maximum de 17 à Sully-la-Tour (Nièvre). La semaine dernière les captures concernaient 1 cuvette sur deux avec un nombre moyen de charançons de la tige piégés de 6,4. Le manque d'ensoleillement de la semaine passée, la persistance des brouillards dans plusieurs secteurs et des températures peu élevées expliquent le faible niveau de piégeage de ce ravageur.



II. Processus d'annotations des Bulletins de Santé du Végétal

1. Bulletins de Santé du Végétal

Réseau 2010-2011

Sur les 74 parcelles qui constituent actuellement le réseau colza 39 ont fait l'objet d'au moins une observation cette semaine.

Stade des colzas

Les stades évoluent lentement :

C1 reprise de végétation : apparition de jeunes feuilles : 64%

C2 entre-nœuds visibles : étranglement vert clair à la base des nouveaux pétioles : 36%

Les conditions climatiques de la semaine dernière avec des températures peu élevées n'ont pas permis de faire évoluer les stades de colza.


Il est urgent de remettre en place les cuvettes jaunes.

Charançon de la tige du colza

39 parcelles renseignées cette semaine

Des captures ont été enregistrées cette semaine dans moins d'un tiers des cuvettes du réseau (principalement dans l'Yonne et la Nièvre). En cas de capture, on dénombre 3,9 charançons en moyenne par cuvette avec un maximum de 17 à Sully la Tour (Nièvre). La semaine dernière les captures concernaient 1 cuvette sur deux avec un nombre moyen de charançons de la tige piégés de 6,4. Le manque d'ensoleillement de la semaine passée, la persistance des brouillards dans plusieurs secteurs et des températures peu élevées expliquent le faible niveau de piégeage de ce ravageur.





II. Processus d'annotations des Bulletins de Santé du Végétal

2. Annotations Thématico Spatio-Temporelles

- **Annotations spatiales** : région de publication
Création d'un jeu de données des différentes régions de France
- **Annotations temporelles** : date de publication
- **Annotations thématiques** : Grande catégorie de la culture
Création du thésaurus FrenchCropUsage¹

Remarque

- Ces annotations ont été obtenues à partir des sites web des chambres d'agriculture régionales.
- Annotations automatiques et manuelles

1. <http://agroportal.lirmm.fr/ontologies/CROPUSAGE>

II. Processus d'annotations des Bulletins de Santé du Végétal

2. Annotations Thématico Spatio-Temporelles

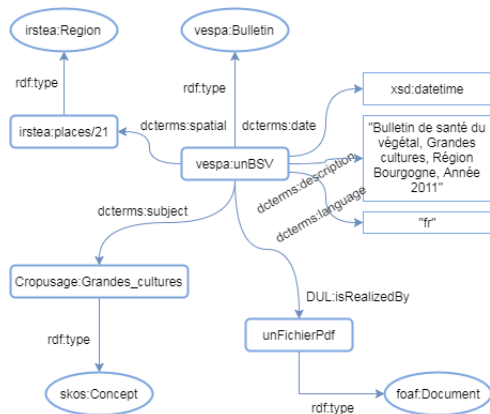




FIGURE – Instanciation du schéma d'annotation des BSV [Roussey et al., 2016]

II. Processus d'annotations des Bulletins de Santé du Végétal

3. Annotation du contenu des BSV




BULLETIN DE SANTÉ DU VÉGÉTAL BOURGOGNE
Grandes cultures n° 14 du 22 février 2011



À RETENIR CETTE SEMAINE

- Lente progression des stades de colza (stade C2 sur 36% des parcelles)
- Régression des captures du charançon de la tige.
- Conditions météorologiques annoncées pour la semaine a priori peu favorables aux vols des ravageurs.



Réseau 2010-2011

Sur les 74 parcelles que constituent actuellement le réseau colza 39 ont fait l'objet d'au moins une observation cette semaine.

Stade des colzas

Les stades évoluent lentement :

C1 reprise de végétation : apparition de jeunes feuilles	: 64%
C2 entre-nœuds visibles : étrangement vert clair à la base des nouveaux pétioles	: 36%


Les conditions climatiques de la semaine dernière avec des températures peu élevées n'ont pas permis de faire évoluer les stades de colza.

Il est urgent de remettre en place les cvettes jeunes.

Charançon de la tige du colza

3 parcelles renseignées cette semaine

Des captures ont été enregistrées cette semaine dans moins d'un tiers des cvettes du réseau (principalement dans l'Yonne et la Nièvre). En cas de capture, on dénombre 3,9 charançons en moyenne par cvette avec un maximum de 17 à Sully la Tour (Nièvre). La semaine dernière les captures concernaient 1 cvette sur deux avec un nombre moyen de charançons de la tige piégés de 6,4. Le manque d'ensoleillement de la semaine passée, la persistance des brouillards dans plusieurs secteurs et des températures peu élevées expliquent le faible niveau de piégeage de ce ravageur.



II. Processus d'annotations des Bulletins de Santé du Végétal

3. Annotation du contenu des BSV

Réseau 2010-2011

Sur les 74 parcelles que constituent actuellement le réseau colza 39 ont fait l'objet d'au moins une observation cette semaine.

Stade des colzas

Les stades évoluent lentement :

C1	reprise de végétation : apparition de jeunes feuilles	: 64%
C2	entre-nœuds visibles : étranglement vert clair à la base des nouveaux pétioles	: 36%

Les conditions climatiques de la semaine dernière avec des températures peu élevées n'ont pas permis de faire évoluer les stades de colza.

Il est urgent de remettre en place les cuvettes jaunes.

II. Processus d'annotations des Bulletins de Santé du Végétal

3. Annotation du contenu des BSV

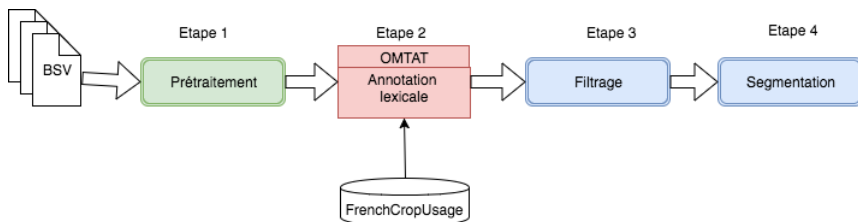



FIGURE – Processus d'annotation



II. Processus d'annotations des Bulletins de Santé du Végétal

3. Annotation du contenu des BSV

Etape 1 : Prétraitement

Résultats attendus

- Détection des paragraphes et découpage
- Elimination de :
 - Pieds de pages : notes et numéro de page
 - Hauts de pages
 - Tableaux
 - Légendes des images

Résultats évaluation :

Précision	Rappel
64.86%	89.69%

II. Processus d'annotations des Bulletins de Santé du Végétal

3. Annotation du contenu des BSV

Etape 2 : Annotation lexicale automatique

- Outil : OMTAT - *One More Text Annotation Tool*
- Thésaurus : FrenchCropUsage

Résultats évaluation :

Précision	Rappel
59.67%	93.48%

T0	Colzas 132 137	colza	
T3	Colzas 1378 1383	colza	
T13	Cultures_florales 1015 1021		fleurs
T18	Choux_potagers 4048 4052		chou
T2	Colzas 720 726	colzas	
T9	Colzas 6817 6822	colza	
T10	Colzas 6864 6869	colza	
T17	Choux 7249 7253	chou	
T6	Colzas 3334 3339	colza	
T14	Moutardes 6069 6078	moutardes	
T7	Colzas 3727 3732	colza	
T11	Colzas 7127 7132	colza	
T5	Colzas 3120 3125	colza	

II. Processus d'annotations des Bulletins de Santé du Végétal

3. Annotation du contenu des BSV

Etape 3 : Filtrage

A RETENIR CETTE SEMAINE

- Grande majorité de parcelle avec des fleurs
- Avec le développement rapide des colzas le risque méligèthes

s'estompe

- Signalement de quelques dégâts de charançons de la tige du colza
- Premiers charançons des siliques piégés mais sans incidence car stade de sensibilité des colzas pas atteint (à partir du stade G2)


T38	Cereales_paille 29 46	Céréales à paille
T6	Cereales 29 37	Céréales
T29	Cereales_paille 40 46	paille

Problématique

- Annotation de cultures générales (fleurs, orange)
- Annotations multiples (Etiquettes longues)

Solution

- Cultures Générales :
 - Filtrage basé sur la fréquence
 - Filtrage basé sur une ressource externe (Annotation thématique)
- Etiquettes longues : Label le plus long



II. Processus d'annotations des Bulletins de Santé du Végétal

3. Annotation du contenu des BSV

Etape 4 : Segmentation

Résultats attendus

Annotation de chaque paragraphe avec le concept correspondant dans le thésaurus en se basant sur la fréquence d'occurrence de la culture dans le paragraphe.



II. Processus d'annotations des Bulletins de Santé du Végétal

3. Annotation du contenu des BSV

Résultats et Observations :

- L'étape de pré-traitement produit du bruit ;
- Problèmes de filtrage de la sortie d'OMTAT.



III. Conclusion

Conclusion et Perspectives

- Annotations thématico Spatio-temporelles ;
- Annotation du contenu BSV.

Perspectives

- Proposer un nouveau schéma d'annotation ;
- Reprendre le processus de reconnaissance automatique : Travaux Synda Ouardani ;
- Construction de nouveau référentiels ;
- Construction de différentes bases de connaissances ;
- Découverte de relations complexes.



III. Conclusion

Bibliographie



Roussey, C., Bernard, S., Pinet, P., Reboud, X., and Cellier, V. (2016).
Gestion sémantique des bulletins de santé du végétal dans le projet vespa.
In *Atelier IN-OVIVE@ IC 2016*, pages 12–p.