



Enrichissement d'un module ontologique : proposition d'une méthode pour le cas de l'agriculture

**F. Amarger^{1,2}, C. Roussey², J.P. Chanet², O. Haemmerlé¹,
N. Hernandez¹**

¹ IRIT

UMR 5505, UTM, Département de Mathématiques-Informatique, 5 allées Antonio Machado, F-31058 Toulouse Cedex - prenom.nom@univ-tlse2.fr

² IRSTEA

Équipe COPAIN, 24 Av. des Landais CS 20085, 63178, Aubière, France - prenom.nom@irstea.fr

Pour mieux
affirmer
ses missions,
le Cemagref
devient Irstea



www.irstea.fr

2 juillet 2013





Plan

- I. **Motivations**
- II. **Etat de l'art**
- III. **Proposition**
 - Hypothèses générales
 - Méthode générale
 - Processus de sélection
 - Processus de filtrage
 - Processus de transformation
 - Processus de validation
- IV. **Perspectives**
- V. **Conclusion**



2 juillet 2013



I. Motivations

- Volonté d'aider les agriculteurs à diminuer l'usage des produits phytosanitaires
- Augmentation du nombre de données dans le domaine de l'agriculture
 - Bulletins de Santé du Végétal
 - thésaurus AGROVOC
 - base de données publique e-phy
 - etc.
- Volonté de contribuer au Linked Open Data (LOD)
- Interrogation de l'ontologie par requête en langage naturel (projet SWIP)



2 juillet 2013



I. Motivations

Objectifs

- Méthodologie de construction d'ontologies à partir de divers sources
- Interrogation intelligente des informations collectées



2 juillet 2013

F. Amarger, C. Roussey, J.P. Chanet, O. Haemmerlé, N. Hernandez



II. Etat de l'art

- Utilisation que d'une seule source
- Possibilité d'extraction :
 - Classes
 - Relations non nommées
 - Intuition de relation de subsomption
- Techniques de désambiguïisation
 - Contraignantes ou pas suffisantes
- Validation peu existante
 - Validation de la conceptualisation manuelle
 - Validation structurelle automatique

Proposition

Méthodologie de construction d'ontologie exploitant la multiplication des sources d'informations disponibles





III. Proposition

1. Hypothèses générales

Hypothèse 1 :

Lors de la création d'une ontologie, toutes les sources n'ont pas le même intérêt

Hypothèse 2 :

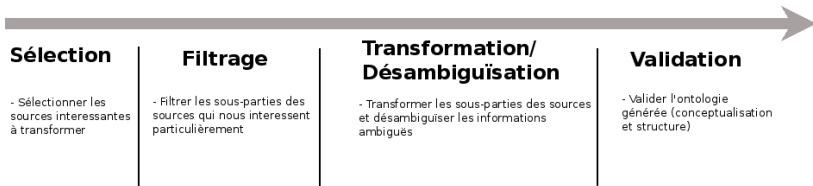
La présence de la même information dans plusieurs sources implique une augmentation de la confiance attribuée à cette information



III. Proposition

2. Méthode générale

- Ontologie modulaire :
 - Ontologie de domaine de haut niveau centrée sur une fonctionnalité
 - Premier cas d'étude module sur la taxonomie agricole
- Enrichissement de module :



2 juillet 2013



III. Proposition

3. Processus de sélection

- Définition par un expert d'un score de confiance pour chaque type de découverte par source (d'après hypothèse 1) :
 - Découverte de classes
 - Découverte de relations de subsomption (subClassOf)
 - Découverte d'autres relations du domaine (objectProperty)
 - Découverte d'attributs (dataProperty)
 - Découverte d'individus



2 juillet 2013



III. Proposition

3. Processus de sélection

Exemple

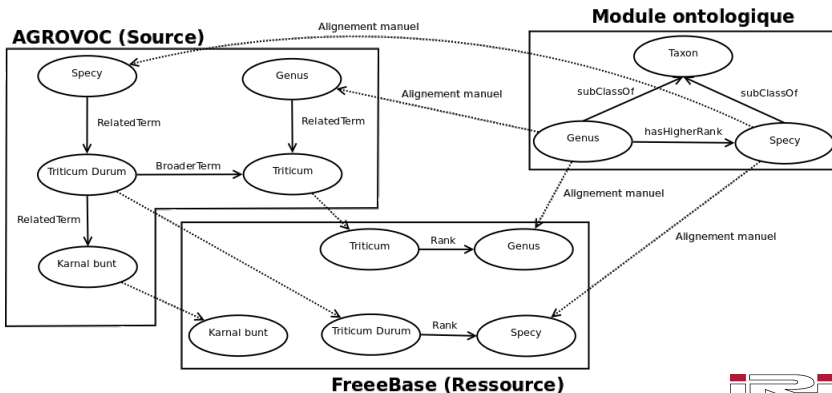
- Agrovoc (Source)
 - Confiance classe : 0.8
 - Confiance relation de subsomption (subClassOf) : 0.1
 - Confiance relation du domaine (ObjectProperty) : \emptyset
- FreeBase (Ressource)
 - Confiance classe : 0.4
 - Confiance relation de subsomption (subClassOf) : 0.5
 - Confiance relation du domaine (ObjectProperty) : 0.7
- NCBI (Ressource)
 - Confiance classe : 0.6
 - Confiance relation de subsomption (subClassOf) : 0.5
 - Confiance relation du domaine (ObjectProperty) : 0.8



III. Proposition

4. Processus de filtrage

- Limites du domaine posées par le module à enrichir
- Utilisation d'une ressource externe pour alignements
- Exemple : filtrage par FreeBase (relation hiérarchique : "Rank")





III. Proposition

5. Processus de transformation

- Transformation naïve des sources
- Désambiguïsation par multiplication des ressources externes
 - Alignement des désambiguïsation sur les relations du module
- Création d'une matrice d'hypothèses



2 juillet 2013

III. Proposition

5. Processus de transformation

Exemple

Matrice de passage :

Source/Ressource	T.D. -> T.
AGROVOC	subClassOf (0.1)
FreeBase	HigherClassification (0.7)
NCBI	subClassOf (0.5)

Matrice d'hypothèses (après alignement des relations) :

T.D. -> T.	hasHigherRank (0.7)	subClassOf (0.6)
----------------------	---------------------	------------------

Validation des classes :

Triticum Durum	1.8
Triticum	1.8

Légende

T. = Triticum
T.D. = Triticum Durum





III. Proposition

6. Processus de validation

- Validation des classes par alignement sur des ressources externes
- Validation des désambiguïisations par addition des scores de confiance
- Possibilité d'ajouter une nouvelle source sans refaire tout le traitement (réutilisation de la matrice d'hypothèses)
- Système de validation structurelle automatique :
 - Anti-Pattern
 - Moteur d'inférences



2 juillet 2013



IV. Perspectives

- Concrétisation de l'idée
 - Implémentation
 - Expérimentations
- Développement d'autres modules
- Système d'interrogation intelligente



2 juillet 2013



V. Conclusion

- Méthodologie d'enrichissement d'un module ontologique
 - Exploitation de nombreuses sources et ressources
 - Traitement simplifié de sources de grandes tailles
- Création d'une ontologie sur les observations d'attaques de bio-agresseurs



2 juillet 2013



VI. Remerciements

Merci pour votre attention.
Avez-vous des questions ?



2 juillet 2013

F. Amarger, C. Roussey, J.P. Chanet, O. Haemmerlé, N. Hernandez