



La confiance consensuelle au service de la connaissance

**Fabien Amarger^{1,2}, Jean-Pierre Chanet², Ollivier Haemmerlé¹,
Nathalie Hernandez¹, Catherine Roussey²**

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, 9 av. Blaise Pascal CS 200 85, 63178, Aubière, France - prenom.nom@irstea.fr

Pour mieux
affirmer
ses missions,
le Cemagref
devient Irstea



www.irstea.fr

13 mai 2014





Plan

I. Motivations

- Contexte
- Etat de l'art
- Problématique

II. Proposition

- Schéma général
- Méthode de transformation de la source
- Alignement et fusion des sources
- Filtrage

III. Conclusion

- Actuellement



13 mai 2014

I. Motivations

1. Contexte



13 mai 2014

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

I. Motivations

1. Contexte

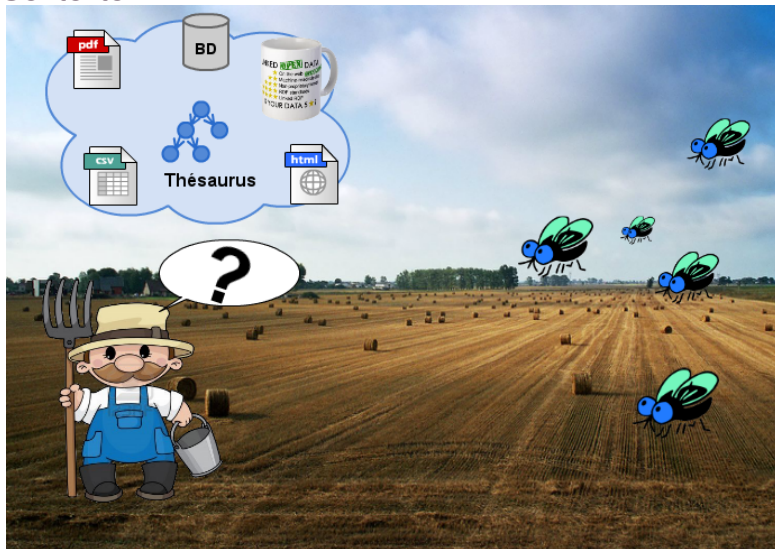


13 mai 2014

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

I. Motivations

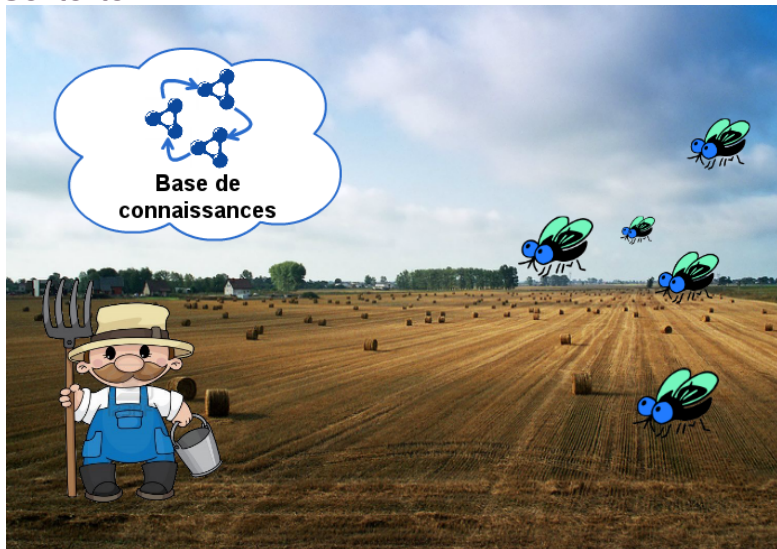
1. Contexte



13 mai 2014

I. Motivations

1. Contexte



13 mai 2014



I. Motivations

2. Etat de l'art

- Transformation de thésaurus :
 - [Amarger et al. 2013 INFORSID]
- Transformation de bases de données :
 - [Sequeda et al. 2011]

Analyse générale

- Utilisation d'une seule source
- Techniques dépendantes de la qualité de la source
- Techniques de désambiguïsation
 - Contraignantes ou insuffisantes
- Aide à la validation peu existante
 - Validation manuelle de la conceptualisation
 - Validation logique automatique





I. Motivations

3. Problématique

Problématique générale

Comment intégrer la transformation de plusieurs sources pour générer une base de connaissances sur une thématique donnée ?

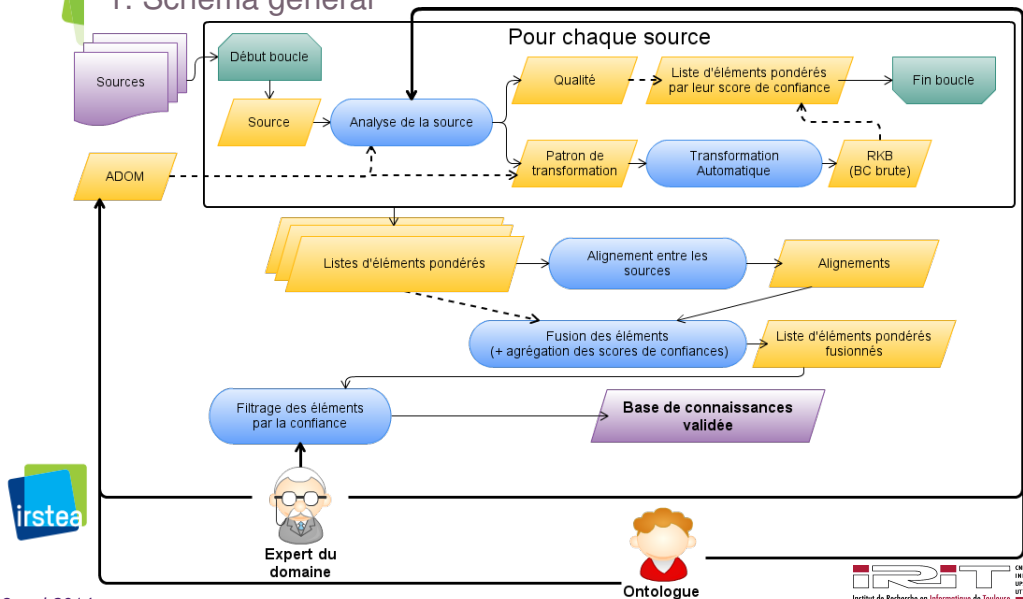
Proposition

- Évaluer la confiance à accorder à une connaissance extraite, à partir de :
 - La qualité de la source de la connaissance
 - Son degré de consensus



II. Proposition

1. Schéma général



Expert du
domaine



Ontologue



II. Proposition

2. Méthode de transformation de la source *Patron de transformation*

Raw Knowledge Base (RKB)

Une RKB est une base de connaissances obtenue par une méthode automatique

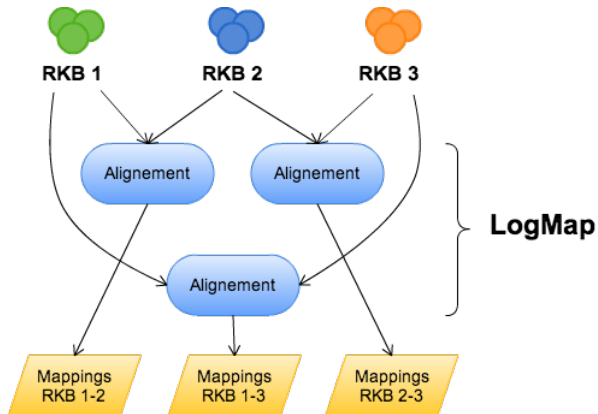
- Définition de règles permettant la transformation d'une source non ontologique en RKB spécialisant l'ADOM
 - Les règles simples sont privilégiées
- Des erreurs et des incohérences dans la RKB mais corrigées par le nombre de sources disponibles
- Par exemple : les relations hiérarchiques d'un thésaurus en subClassOf



13 mai 2014

II. Proposition

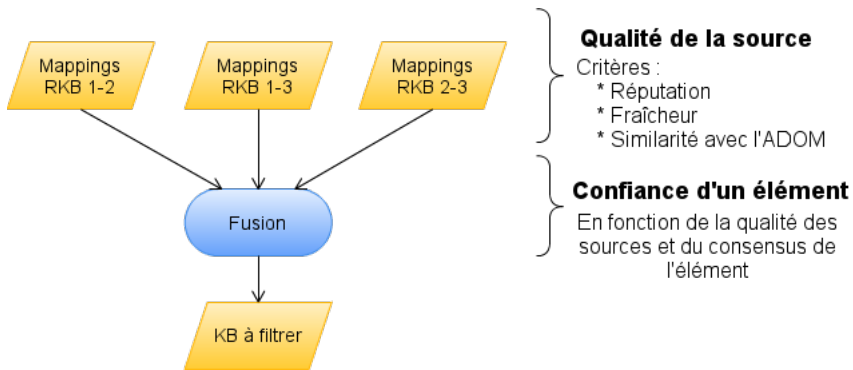
3. Alignement et fusion des sources



13 mai 2014

II. Proposition

3. Alignement et fusion des sources



13 mai 2014

II. Proposition

3. Alignement et fusion des sources

Proposition de fusion

Trust for Merge - New Project

Project name :

ADOM :

Criteria weight

Freshness : ▲▼

Reputation : ▲▼

ADOM similarity : ▲▼



13 mai 2014

II. Proposition

3. Alignement et fusion des sources

Proposition de fusion

Trust for Merge - Project 1

Sources :

	Freshness	Reputation	ADOM similarity	Source quality score
Agrovoc_OWL	<input type="button" value="Change source"/> <input type="button" value="Pretty good"/>	<input type="button" value="Bad - 0"/> <input type="button" value="Pretty bad - 0.25"/> <input type="button" value="Medium - 0.5"/> <input type="button" value="Pretty good - 0.75"/> <input type="button" value="Good - 1"/>	Medium	0.75



13 mai 2014

II. Proposition

3. Alignement et fusion des sources

Proposition de fusion - Calcul du score ADOM similarity

ADOM sim

$$ADOMSim(ADOM, S) = \frac{nbElemAlign(ADOM, S)}{nbElem(ADOM)}$$

$nbElemAlign(X, Y)$: nombre de correspondances trouvées entre X et Y

$nbElem(X)$: nombre d'éléments présents dans X

Variables linguistiques :

0]0, 0.25[[0.25, 0.5[[0.5, 0.75[[0.75, 1[1
Error	Bad	Pretty bad	Medium	Pretty good	Good



II. Proposition

3. Alignement et fusion des sources

Proposition de fusion - Calcul du score de qualité

Score de qualité d'une source (moyenne pondérée) :

$$\text{SourceQual}(S) = \frac{\text{imp}(cf) * \text{val}(cf, S) + \text{imp}(cr) * \text{val}(cr, S) + \text{imp}(csa) * \text{val}(csa, S)}{\text{imp}(cf) + \text{imp}(cr) + \text{imp}(csa)}$$

$\text{imp}(x)$: Importance du critère x pour le projet

$\text{val}(x, y)$: Valeur du critère x pour la source y



13 mai 2014

II. Proposition

3. Alignement et fusion des sources

Proposition de fusion - Calcul de la confiance d'un élément

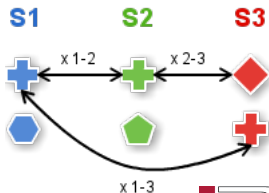
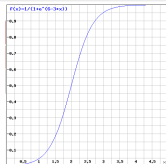
Confiance d'un élément

$$\text{trust}(e) = \sigma\left(\sum_{i=1}^n (\text{SourceQual}(S_i) + \text{MAX}(S_i, \text{salign}(e)))\right)$$

$$\sigma(x) = \frac{1}{1 + \exp^{6-3x}}$$

n = nombre de sources dans lequel l'élément e apparaît

$\text{MAX}(S_i, \text{salign}(x))$ = Valeur max des scores d'alignements dans lequel l'élément x est repéré dans la source S_i





II. Proposition

4. Filtrage

- Utilisation de seuils pour filtrer les connaissances :
 - $trust(e) \geq 0.90 \rightarrow$ Garder la connaissance
 - $trust(e) \leq 0.10 \rightarrow$ Ne pas garder la connaissance
 - $0.10 < trust(e) < 0.90 \rightarrow$ Validation manuelle
- Permet la validation d'un nombre restreint de connaissances



13 mai 2014



III. Conclusion

1. Actuellement

- Actuellement :
 - Prototype fonctionnel
 - Expérimentation encourageante sur un jeu de données simple (Agrovoc, TaxRef, NCBI)
- Travaux futurs :
 - Théorie de l'argumentation pour gérer les éléments contradictoires
 - Ajouter une propagation de confiance lors d'une validation manuelle (phase de filtrage)



13 mai 2014

IV. Expérimentations

1. Présentation des sources

- AGROVOC :
 - Thésaurus multilingue de 32 000 concepts créé par la FAO¹
- TaxRef :
 - Référentiel taxonomique contenant 80 000 taxons créé par le Muséum national d'histoire naturelle²
- NCBI taxonomie :
 - Taxonomie créée par National Center for Biotechnology Information (NCBI) des états unis, contenant 1 000 000 de taxons³

MiniRKB - Triticum

- AGROVOC : 10 instances
- TaxRef : 11 instances
- NCBI : 80 instances

1. <http://aims.fao.org/standards/agrovoc/about>
2. <http://inpn.mnhn.fr/programme/referentiel-taxonomique-taxref>
3. <http://www.ncbi.nlm.nih.gov/taxonomy>

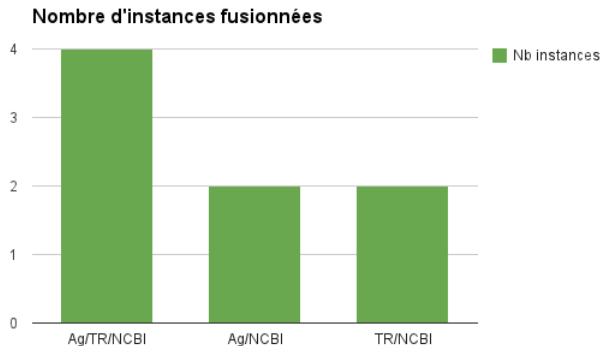


13 mai 2014

IV. Expérimentations

1. Présentation des sources

Résultats



13 mai 2014