



**HAL**  
open science

## Apports des bases de connaissances RDF à la détection de vérité et vice versa

Valentina Beretta, Sébastien Harispe, Sylvie Ranwez, Isabelle Mougenot

### ► To cite this version:

Valentina Beretta, Sébastien Harispe, Sylvie Ranwez, Isabelle Mougenot. Apports des bases de connaissances RDF à la détection de vérité et vice versa. 30es Journées Francophones d'Ingénierie des Connaissances, IC 2019, AFIA, Jul 2019, Toulouse, France. pp.97-98. hal-02329518

**HAL Id: hal-02329518**

**<https://hal.science/hal-02329518>**

Submitted on 23 Oct 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Apports des bases de connaissances RDF à la détection de vérité et vice versa

Valentina Beretta<sup>1</sup>, Sébastien Harispe<sup>2</sup>, Sylvie Ranwez<sup>2</sup>, and  
Isabelle Mougenot<sup>3</sup>

<sup>1</sup> MIDN, IRD, Marseille, France

<sup>2</sup> LGI2P, IMT Mines Ales, Univ Montpellier, Ales, France

<sup>3</sup> UMR 228 Espace Dev UM, Maison de la Télé-détection, Montpellier, France

**Résumé** : La détection de vérité sur le Web (ou Truth Discovery), est un domaine de recherche qui a émergé ces dernières années pour contrer les dangers de la désinformation. Elle vise à identifier les faits (assertions vraies) lorsque des assertions contradictoires sont émises par différentes sources. Partant de l'hypothèse que les assertions vraies sont fournies par des sources fiables, et que les sources fiables fournissent des assertions vraies ; les modèles de recherche de vérité de la littérature calculent itérativement la confiance dans une assertion et la fiabilité des sources. Il est ainsi possible d'identifier les assertions qui se révèlent vraies.

Dans l'étude présentée dans Beretta *et al.* (2018), nous avons exploité les connaissances du domaine exprimées sous la forme de bases de connaissances RDF, pour améliorer les performances des méthodes actuelles de recherche de vérité. Notre postulat est que la connaissance *a priori* d'un domaine peut conforter certaines assertions et donc influencer le processus de détection de vérité. En étudiant les cooccurrences entre les faits présents dans une base de connaissances RDF, il est possible d'identifier des motifs sous forme de règles qui renforcent la confiance accordée à certaines affirmations exprimées sous la forme de triplets <subjects, prédicat, valeur>. Par exemple, une analyse de la base de connaissances peut permettre de déduire que la majorité des personnes qui parlent espagnol, sont nées en Espagne. Ce constat peut être pris en compte dans un processus de recherche de vérité concernant le lieu de naissance du peintre Pablo Picasso. Si on observe que Pablo Picasso parle couramment espagnol, la confiance attribuée à l'assertion <Pablo Picasso, est né en, Espagne> doit être renforcée. Il en va de même pour les assertions qui contiennent des valeurs plus générales, à l'exemple de <Pablo Picasso, est né en, Europe>.

Dans notre approche, l'identification des règles est réalisée à l'aide d'AMIE+ (Galárraga *et al.*, 2015). A partir de ces règles, nous avons calculé un coefficient propulseur (ou booster) qui traduit les cooccurrences récurrentes entre différents faits sous la forme de mesures de qualité, afin de renforcer la confiance dans certaines valeurs (et donc dans certaines assertions) lors du processus de détection de vérité. Ce coefficient propulseur représente le degré de soutien (ou caution) conféré à une assertion en exploitant le contenu de la base de connaissances. Ainsi le postulat de base du processus de recherche de vérité présenté en introduction en sera modifié et nous considérerons désormais que les faits (vérités) sont des assertions proposées par des sources fiables et/ou qui sont renforcées par un coefficient booster élevé, en considération de règles d'association extraites de la base de connaissances. Il est à noter que les motifs récurrents n'ont pas tous le même degré d'expressivité et ne doivent donc pas avoir le même impact sur le processus de détection de vérité. L'influence du coefficient booster sur le calcul de confiance dans une assertion sera donc paramétrable afin d'accorder plus d'importance soit à la fiabilité des sources, soit à l'information contenue dans la base de connaissances en fonction du contexte et/ou de la qualité de la base.

Une évaluation de l'approche a été réalisée à l'aide de jeux de données réels. L'objectif principal était d'évaluer l'impact de la prise en compte des connaissances *a priori* lors de l'utilisation des modèles de recherche de vérité dans un scénario réel, doté de spécificités propres. Nous avons adapté le modèle de découverte de vérité Sums, défini dans (Pasternack & Roth, 2010). La méthode Sums adopte une procédure itérative dans laquelle le calcul de la fiabilité associée à une source et le calcul de la confiance associée à une assertion sont alternés jusqu'à atteindre une convergence. Nous l'avons adaptée en modifiant le calcul de la confiance d'une assertion pour qu'il prenne en compte le coefficient propulseur. Nous montrons notamment que les modifications apportées au modèle Sums permettent d'obtenir des gains de performance de l'ordre de 18% par rapport à l'approche Sums seule. Par comparaison avec l'existant, ce gain de performances permet de placer notre modèle parmi les modèles les plus performants du domaine de la recherche de vérité.

## Références

BERETTA V., HARISPE S., RANWEZ S. & MOUGENOT I. (2018). Combining truth discovery and RDF knowledge bases to their mutual advantage. In *The Semantic Web - ISWC 2018 - 17th International Semantic Web Conference, Monterey, CA, USA, October 8-12, 2018, Proceedings, Part I*, p. 652–668.

*IC 2018*

- GALÁRRAGA L., TEFLIOUDI C., HOSE K. & SUCHANEK F. M. (2015). Fast rule mining in ontological knowledge bases with amie+. *The VLDB Journal—The International Journal on Very Large Data Bases*, **24**(6), 707–730.
- PASTERNAK J. & ROTH D. (2010). Knowing what to believe (when you already know something). In *Proceedings of the 23rd International Conference on Computational Linguistics*, p. 877–885.