

Pharmacovigilance du Web Social par une approche fondée sur les bases de connaissances du Web Sémantique

Damien Leprovost, Marie-Christine Jaulent

Laboratoire Limics, INSERM U-1142, Univ. UPMC Paris 6, Sorbonne Universités
Université Paris 13, Sorbonne Paris Cité, 75006 Paris, France – {prenom.nom}@inserm.fr

1 Introduction

La pharmacovigilance s'intéresse à la surveillance des effets indésirables des médicaments autorisés sur le marché. On estime que moins de 5% de ces événements sont effectivement déclarés, ce qui entraîne un coût économique et humain très important (Hazell et Shakir, 2006).

Nous présentons ici nos premiers travaux visant à améliorer tant la détection de ces effets que celle des signaux. Nous utilisons la structure des données du Web Sémantique pour réaliser une pré-évaluation automatique des données du Web Social, afin de fournir un ensemble limité de signaux potentiels. La figure 1 présente l'architecture du framework LINPHAS développé.

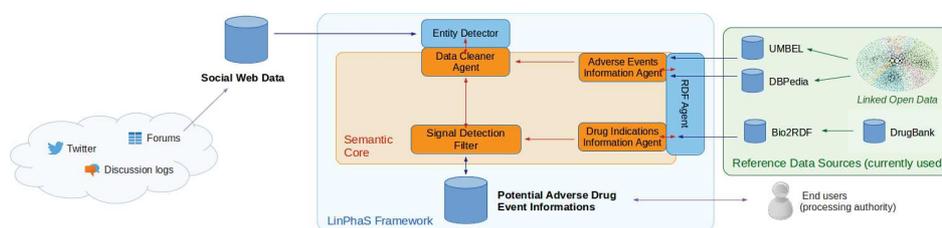


FIG. 1 – Architecture fonctionnelle du framework LINPHAS

2 Données utilisées et processus

Les données sociales peuvent être n'importe quel contenu Web généré par des utilisateurs. Dans notre étude, 82 924 tweets sont extraits entre septembre 2014 et mars 2015 à partir de 34 médicaments ciblés¹. Les données ouvertes et liées utilisées proviennent de trois jeux de données : *UMBEL*, pour extraire et naviguer entre les concepts ; *DBpedia*, pour déterminer la terminologie et *DurgBank via Bio2RDF* pour accéder aux informations des médicaments.

Après nettoyage et normalisation, les messages sont mis en correspondance avec les données extraites du Web Sémantique (états de santé et médicaments). Ne sont alors conservées

1. Le détail des termes retenus est disponible à <https://goo.gl/pKXUkX>.