

# INTRODUCTION

La pharmacovigilance est l'étude des effets indésirables liés aux médicaments. La mise en évidence d'un signal (une relation inattendue entre un médicament et un effet indésirable) s'appuie sur des méthodes d'apprentissage statistique. En France, plus de 250 000 cas sont actuellement enregistrés dans la base nationale et il en existe environ 3,5 millions dans la base de l'Organisation mondiale de la Santé.

En l'absence d'une représentation formelle des termes qui codent l'information, le processus d'extraction des connaissances à partir des données n'est pas efficace. Nous posons l'hypothèse qu'un pré-traitement des données au moyen d'un raisonnement terminologique permet d'améliorer les performances des algorithmes de détection du signal.

Notre objectif est de comparer trois modèles de raisonnement terminologique : distance sémantique, subsomption au moyen de logiques de description, analyse formelle des concepts. Dans chaque cas, nous réalisons un pré-traitement des données au moyen d'une de ces méthodes et nous comparons les résultats obtenus par apprentissage statistique avant et après le pré-traitement terminologique.

Ce travail devrait mettre en évidence l'intérêt d'une représentation formelle des éléments terminologiques et des modèles de raisonnement terminologique associés avant toute démarche d'apprentissage statistique sur des données de pharmacovigilance.

## RÉFÉRENCES

BATE A, LINDQUIST M, EDWARDS IR, ORRE R (2002). A data mining approach for signal detection and analysis. *Drug Safety*;25(6):393-7.

BOUSQUET C, JAULENT M, CHATELLIER G, DEGOULET P. (2000) Using semantic distance for the efficient coding of medical concepts. *Proc AMIA Symp.*;20(Suppl):96-100.

BOUSQUET C, LAGIER G, LILLO-LE LOUËT A., LE BELLER C., VENOT A., JAULENT MC. Appraisal of MedDRA conceptual structure for describing and grouping adverse drug reactions. *Drug Safety* 2005 ;28(1) :19-34.

BOUSQUET C., HENEGAR C., LILLO-LE LOUËT A., DEGOULET P., JAULENT MC. Implementation of automated signal generation in pharmacovigilance using a knowledge based approach. *Int J Med Inform* 2005;74(7-8):563-71.

BROWN EG (2002). Effects of coding dictionary on signal generation: a consideration of use of MedDRA compared with WHO-ART. *Drug Safety* ;25(6):445-52.

HENEGAR C., BOUSQUET C., LILLO-LE LOUËT A., DEGOULET P., JAULENT MC. Building an ontology of adverse drug reactions for automated signal generation in pharmacovigilance. *Comput Biol Med*. 2005

EGBERTS ACG, MEYBOOM RHB, VAN PUJENBROEK EP (2002). Use of measures of disproportionality in pharmacovigilance: Three dutch examples. *Drug Safety*;25(6):453-8.

EVANS SJ, WALLER PC, DAVIS S (2001). Use of proportional reporting ratios (PPRs) for signal generation from spontaneous adverse drug reaction reports. *Pharmacoepidemiol Drug Saf*;10(6):483-6.

MOORE N, KREFT-JAIS C, HARAMBURU F, NOBLET C, ANDREJAK M, OLLAGNIER M, BÉGAUD B (1997). Reports of hypoglycemia associated with the use of ACE inhibitors and other drugs: a case/non case study in the French pharmacovigilance system database. *Br J Clin Pharmacol*;44(5):513-8.

SZARFMAN A, MACHADO SG, O'NEILL RT (2002). Use of screening algorithms and computer systems to efficiently signal higher-than-expected combinations of drugs and events in the US FDA's spontaneous reports Database. *Drug Safety*;25(6):381-92.

# EI-Xplora : évaluation de trois méthodes de raisonnement terminologique pour la détection du signal en pharmacovigilance

Cédric BOUSQUET <sup>1</sup>, Iulian ALECU <sup>1</sup>, Anita BURGUN <sup>2</sup>, Catherine DUCLOS <sup>3</sup>, Nicolas GARCELON <sup>2</sup>, Jimison IAVINDRASANA <sup>1</sup>, Agnès LILLO-LE LOUËT <sup>4</sup>, Fleur MOUGIN <sup>2</sup>, Alain VENOT <sup>3</sup>, Yannick TOUSSAINT <sup>5</sup> et Marie-Christine JAULENT <sup>1</sup>

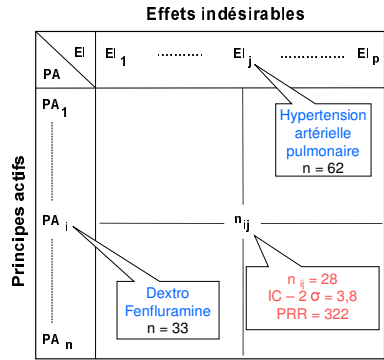
<sup>1</sup> INSERM U729, Paris [cedric.bousquet,marie-christine.jaulent@spim.jussieu.fr](mailto:cedric.bousquet,marie-christine.jaulent@spim.jussieu.fr)

<sup>2</sup> EA2888, Département d'information médicale, Rennes

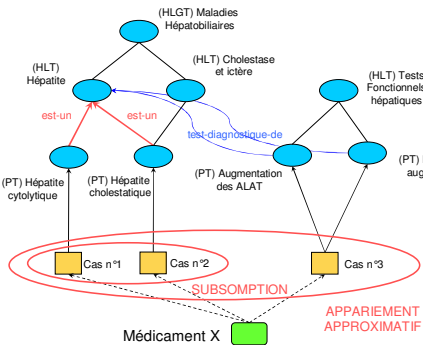
<sup>3</sup> LIM&BIO, Laboratoire d'informatique médicale et bioinformatique, Bobigny

<sup>4</sup> Centre régional de pharmacovigilance, Hôpital Européen Georges Pompidou, Paris

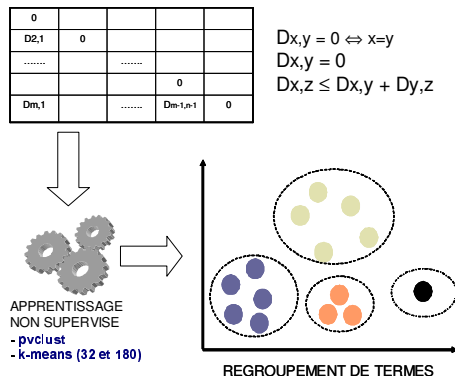
<sup>5</sup> LORIA, Laboratoire lorrain en informatique et ses applications, Nancy



**Figure 1** De façon rétrospective l'association entre la dextrofenfluramine et l'hypertension artérielle pulmonaire est statistiquement significative (données du centre de pharmacovigilance, Hôpital Européen Georges Pompidou). La borne inférieure de l'indice « information component (IC) » proposé par l'OMS est nettement supérieure à 0.

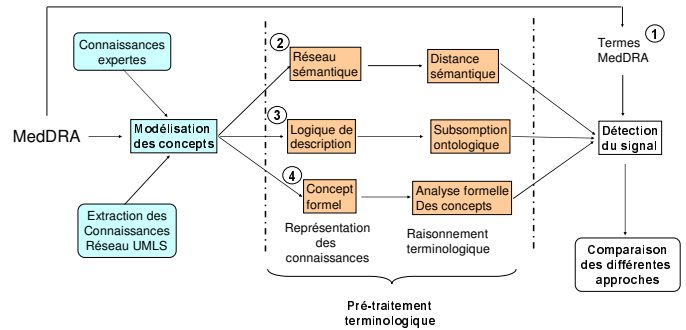


**Figure 3** Trois modes de raisonnement terminologique : 1) regroupement des cas n°1 et n°2 au niveau du HLT Maladies hépatobiliaires au moyen des relations taxinomiques MedDRA. 2) Regroupement des cas n°1 et n°2 au niveau du HLT Hépatite grâce aux relations de subsomption dans l'ontologie formelle. 3) Regroupement du cas n°3 avec les deux autres par appariement approximatif.

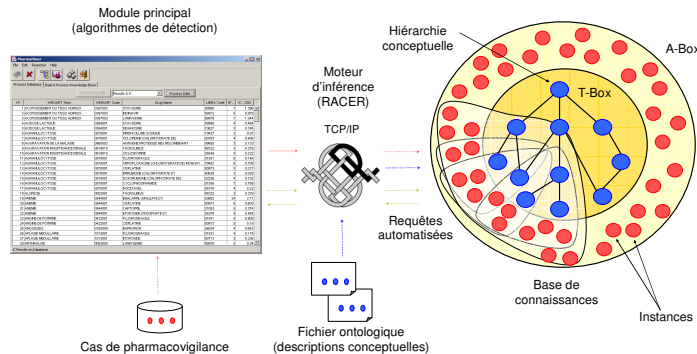


**Figure 5** Les regroupements sont obtenus au moyen de deux techniques d'apprentissage non supervisé.

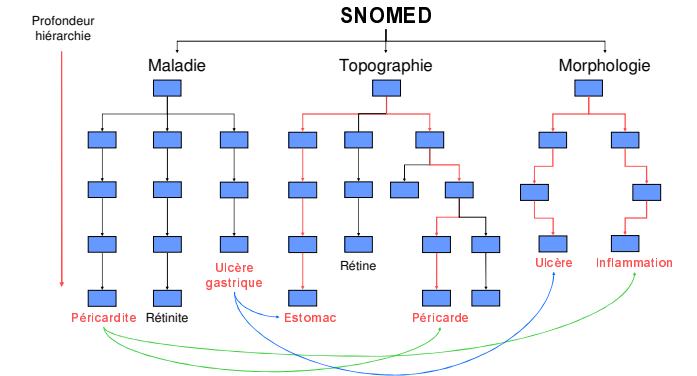
- 1) La classification hiérarchique avec l'algorithme pvcust en faisant varier la valeur du p-value de 0.97 à 0.99.
- 2) La technique de réallocation itérative pour partitionner les termes en 180 classes, puis en 32 classes au moyen de la méthode kmeans.



**Figure 2** Notre objectif est de comparer les avantages et inconvénients des formalismes mis en œuvre. Les résultats sont comparés selon les modalités suivantes : 1) Signaux détectés par une méthode quantitative directement sur les termes MedDRA 2) Signaux détectés par une méthode quantitative après formalisation par un réseau sémantique 3) Signaux détectés par une méthode quantitative après formalisation par une ontologie 4) Signaux détectés par une méthode quantitative après formalisation par un treillis de concept.



**Figure 4** Le module principal de l'outil PharmaMiner intègre l'interface utilisateur, les méthodes quantitatives de détection du signal et la communication avec les autres modules. Le moteur d'inférence communique avec le module principal au moyen d'une connexion TCP/IP. Ce moteur construit la base de connaissances au niveau conceptuel (T-Box) et factuel (A-Box). Le fichier ontologique contient la description des termes MedDRA au format DAML+OIL et permet de constituer la T-Box. Les objets contenus dans la A-Box sont les descriptions simplifiées des cas de pharmacovigilance tels qu'ils sont présentés dans les données d'entrée.



**Figure 6** Les effets indésirables sont modélisés au moyen de projections multiples sur la nomenclature SNOMED. Par exemple l'effet indésirable « Ulcère gastrique » se projette sur « Ulcère » dans l'axe morphologique et « Estomac » dans l'axe topographique. La distance entre « Péricardite » et « Ulcère gastrique » est une combinaison des distances : - entre « Estomac » et « Péricarde » sur l'axe topographique et - entre « Ulcère » et « Inflammation » sur l'axe morphologique.