



HAL
open science

VigiTermes : une plateforme de recherche et d'analyse des publications scientifiques au service de la pharmacovigilance

Mathilde Thebault, Florence Amardeilh, Denis Delamarre, Sylvie
Guillemin-Lanne, Anne Jamet, Agnès Lillo-Le Louet

► To cite this version:

Mathilde Thebault, Florence Amardeilh, Denis Delamarre, Sylvie Guillemin-Lanne, Anne Jamet, et al.. VigiTermes : une plateforme de recherche et d'analyse des publications scientifiques au service de la pharmacovigilance. 21es Journées Francophones d'Ingénierie des Connaissances, Jun 2010, Nîmes, France. pp.195 - 206. hal-00488145

HAL Id: hal-00488145

<https://hal.science/hal-00488145v1>

Submitted on 1 Jun 2010

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.

VigiTermes : une plateforme de recherche et d'analyse des publications scientifiques au service de la pharmacovigilance

Mathilde Thébault¹, Florence Amardeilh², Denis Delamarre³, Sylvie Guillemin-Lanne¹, Anne Jamet⁴, Agnès Lillo-Le Louët⁴

¹TEMIS, 164 rue de Rivoli, 75001 Paris,
(mathilde.thebault, sylvie.guillemin-lanne)@temis.com

²Mondeca, 3 cité Nollez, 75018 Paris, France
florence.amardeilh@mondeca.com

³INSERM U936, Université Rennes 1, IFR 140, Rennes, France,
denis.delamarre@chu-rennes.fr

⁴Centre Régional de Pharmacovigilance, Hôpital Européen Georges Pompidou, Assistance
Publique-Hôpitaux de Paris, Paris, France
Agnès.Lillo-LeLouet@hop.egp.ap-hop-paris.fr

Résumé : La réglementation impose une identification systématique des effets indésirables des médicaments. La plupart des pays se sont dotés de systèmes d'information d'aide à la documentation de ces effets. La tâche est rendue difficile par la multiplicité des sources et le peu d'outils fédérateurs existant pour accéder, rechercher et analyser l'information autour des médicaments. L'objectif du projet VigiTermes est de développer une plateforme pour améliorer la documentation des rapports de cas de pharmacovigilance et de proposer des outils d'accès et d'analyse pour les experts pharmacovigilants, dont l'objectif est d'enquêter sur la détection de nouveaux cas (signaux). Dans ce cadre, nous avons développé un prototype qui reproduit et standardise des stratégies de recherche documentaire formulées par les pharmacovigilants, récupère les résumés PubMed pertinents et en extrait de l'information autour des médicaments et de leurs effets secondaires potentiels.

Mots-clés : annotation sémantique, text mining, modélisations des connaissances, bases de connaissance, pharmacovigilance, recherche PubMed.

1 Les besoins en pharmacovigilance

La pharmacovigilance a pour objet la surveillance du risque d'effet indésirable. Elle comprend en particulier l'identification, l'évaluation et la prévention du risque de l'utilisation des produits à finalité sanitaire à usage humain. Cette activité de veille sanitaire s'exerce notamment sur tous les médicaments bénéficiant d'une autorisation de mise sur le marché, qu'elle soit française ou européenne. La pharmacovigilance a désormais une dimension européenne afin de prendre en compte l'évaluation et la gestion d'un problème survenant sur le territoire communautaire [BPPV, 2005]. En France, le réseau national de pharmacovigilance est organisé en 31 centres régionaux

de pharmacovigilance (CRPV) coordonnés par l'unité de pharmacovigilance de l'AFSSAPS (Agence Française de Sécurité Sanitaire des Produits de Santé). La mission des CRPV est de surveiller, évaluer, prévenir les risques médicamenteux potentiels ou avérés, et de promouvoir le bon usage du médicament.

La principale source d'information en pharmacovigilance est fondée sur la notification spontanée des EI par les professionnels de santé vers un système national de pharmacovigilance, et/ou vers les industries pharmaceutiques. D'autres sources d'information peuvent se révéler intéressantes pour la documentation et la caractérisation de nouveaux effets secondaires ou pour compléter les informations existantes. La littérature biomédicale est, à ce titre, intéressante car elle enregistre de manière pérenne des informations vérifiées scientifiquement dans un contexte qui peut être autre que la pharmacovigilance. Elle peut ainsi offrir de nouvelles pistes pour peu que son information soit correctement exploitée. Avec l'avènement d'Internet, d'autres sources de données, plus subjectives, ont émergé. La pharmacovigilance peut également s'appuyer sur ces nouvelles approches (blogs, revues non médicales, forums de discussion voire sites de consommateurs ou de patients) pour enrichir sa connaissance des médicaments. La pharmacovigilance requiert donc des stratégies spécifiques pour intégrer ces sources et fédérer l'ensemble de l'information. Ces outils doivent être suffisamment flexibles et puissants pour permettre d'explorer le vaste champ des médicaments et leurs effets secondaires tout en répondant à la fois à des questions très générales (*les effets secondaires attendus avec un traitement antidépresseur*) ou très précises (*le risque de suicide chez les adolescents prenant un inhibiteur de la recapture de la sérotonine (SSRI)*).

Le projet VigiTermes (ANR-07-TECSAN-026-03) est dédié à la mise en œuvre d'une plateforme capable d'assister le pharmacovigilant dans sa recherche d'informations scientifiques autour des effets secondaires des médicaments. Le CRPV de l'Hôpital Européen Georges Pompidou (HEGP) expérimente en réel la plateforme VigiTermes, destinée à offrir un portail sémantique à la validation fonctionnelle des cas relevant de la pharmacovigilance. La solution permet de rechercher de façon automatique des publications d'intérêt et de les annoter ; elle donne au pharmacovigilant les moyens d'analyser au mieux ces données.

2 La plateforme VigiTermes

La plateforme VigiTermes se présente sous la forme d'un portail sémantique permettant l'accès à l'annotation automatique de résumés d'articles issus de PubMed traitant potentiellement de cas de pharmacovigilance similaires et l'accès à la documentation des cas rapportés grâce aux instances de connaissance créées à partir des annotations, en fonction d'une ontologie de la pharmacovigilance et de terminologies médicales standard. Ce portail sémantique fournit donc une interface utilisateur, à partir de laquelle l'expert pharmacovigilant soumet ses requêtes. Celles-ci comportent au moins le couple (effet indésirable, principe actif), notifié dans le cas qu'il souhaite documenter. Le flux d'information du portail VigiTermes est représenté ci-dessous.

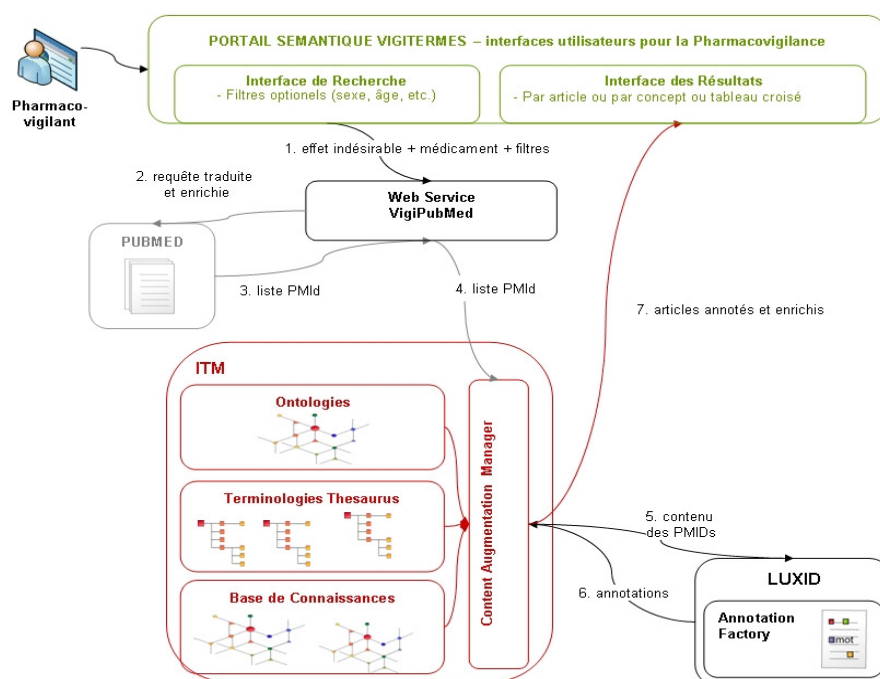


Fig. 1 – VigiTermes: schéma global du flux de l'information

2.1 Web service de récupération des publications PubMed

La première étape de la chaîne consiste en l'interrogation de la base bibliographique PubMed et la récupération automatique des publications correspondant à un nouveau cas de pharmacovigilance. Le traitement des données en entrée, un couple Effet Indésirable (EI) – Principe Actif (PA), s'appuie sur différentes ressources ; l'objectif est de construire une requête PubMed qui utilise les index MeSH (MeSH Headings et MeSH Qualifiers), afin de récupérer les résumés pertinents relatifs au cas analysé.

La stratégie de requête utilise le métathésaurus UMLS pour normaliser et mettre en correspondance le terme MedDRA de l'EI avec le ou les termes MeSH équivalents [Boden & al., 1998], le médicament est désigné directement par sa dénomination commune internationale (DCI). La requête PubMed est ensuite élaborée en combinant les termes bruts, les termes MeSH obtenus par correspondance et enfin les qualificatifs MeSH qui indiquent un rapport aux EI : "adverse effect", "chemically induced" et "drug toxicity" [Garcelon & al., 2006]. Des paramètres supplémentaires peuvent enrichir la requête à des fins de filtrage, tels que le genre ou la classe d'âge pour les patients, le type ou la période pour la publication.

D'un point de vue technique, cette stratégie de recherche est intégrée à la plateforme sous la forme d'un service web. Il utilise lui-même certains des services

web de l'UMLS et de la base de données PubMed. À partir de ce programme principal, nommé VigiPubMed, nous avons testé et évalué plusieurs méthodes d'enrichissement de la requête PubMed (Figure 2) [Delamarre & al., 2010]. La méthode 1 est une requête directe de la base PubMed, la méthode 2 utilise la normalisation par UMLS décrite ci-dessus, la méthode 3 consiste à étendre la recherche avec les sous-hiérarchies des termes MeSH. La méthode 4 s'appuie sur l'ontologie des effets indésirables OntoEIM [Jaulent & al., 2009], et un programme de recherche de cas similaires [Alecu & al., 2007]. Cette dernière permet à partir du terme initial d'élargir la recherche à un groupe de termes MedDRA. L'évaluation consiste à dénombrer le nombre d'articles pertinents de chaque méthode afin de mesurer l'efficacité de la recherche en termes de rappel et de précision.

La difficulté principale de cette étape a été de choisir l'une ou l'autre des méthodes mises en œuvre. Une méthode seule ne permet pas de répondre complètement au mode de requête du pharmacovigilant, et peut s'avérer meilleure ou moins bonne, en fonction de ses besoins et des résultats obtenus. Par ailleurs, des difficultés liées au domaine médical ont été également rencontrées. Un EI potentiel peut être en réalité une maladie pour laquelle le médicament impliqué est le traitement. C'est le cas, par exemple, de la neutropénie qui peut être l'indication ou l'EI de certains antibiotiques. Des termes complémentaires, tels que "adverse effect" ou "drug toxicity" peuvent enrichir et orienter la requête.

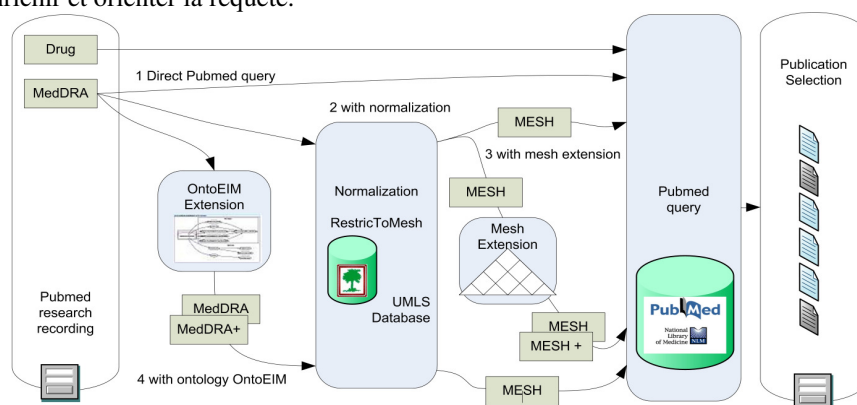


Fig. 2 – Flux d'information pour la récupération des articles PubMed

2.2 Annotation des publications de PubMed

Les publications scientifiques de PubMed fourmillent d'informations pertinentes sur les EI liés à des médicaments via des cas similaires publiés. Le processus mis en œuvre consiste à enrichir leur contenu au moyen d'annotations qui sont automatiquement envoyées à la plateforme VigiTermes pour alimenter la base de connaissance des cas de patients ayant présenté des EI "d'intérêt".

2.2.1 Le processus d'annotation

Les résumés de publications scientifiques de PubMed en réponse à la requête (étape 5 du schéma global présenté en Fig. 1) sont automatiquement analysés, à l'aide du composant logiciel dédié à l'extraction d'information Luxid[®]. Au terme du processus, les documents enrichis des entités médicales pertinentes ou encore des relations mises en évidence entre ces entités, sont transmis vers les outils de consolidation de la connaissance du référentiel ITM (étape 6 du schéma global de la Fig. 1).

Le processus d'annotation démarre en une première étape par la détection de la langue, la segmentation et l'analyse morpho-syntaxique. Les segments de phrases obtenus sont étiquetés (groupes nominaux, noms propres, verbes, adjectifs...). La seconde étape fait appel au composant linguistique dédié à la reconnaissance des entités médicales et des relations entre ces entités, la Skill Cartridge[™] ou cartouche de connaissance. Celle-ci est une modélisation de l'information à extraire, organisée hiérarchiquement en modules. Chaque module, représentant les notions du domaine, est composé de lexiques et de règles d'extraction, expressions régulières qui repèrent une séquence textuelle et l'étiquettent comme concept [Aubry & al., 2002]. Pour chaque entrée lexicale, il est possible d'accéder à leurs caractéristiques 1/ le lemme (forme canonique, comme par exemple l'infinitif d'un verbe ou le masculin singulier d'un substantif), 2/ la forme (telle qu'elle est rencontrée dans les textes), et 3/ la normalisation (convention arbitraire d'affichage). Exemple: 1/ Lemme: "hair loss" ; 2/ Forme: "hair losses" ; 3/ Normalisation: "Alopecia" (terme du MedDRA).

2.2.2 La Skill Cartridge[™] Pharmacovigilance

Dans le cadre du projet, une Skill Cartridge[™] dédiée à la Pharmacovigilance a été implémentée. Elle détecte des entités médicales (Maladies, Traitements médicaux, Méthodes de diagnostic, Symptômes, Essais Cliniques) aussi bien que des relations sémantiques entre ces entités (Diagnostic, Thérapie, Effets des Médicaments).

Nous avons enrichi les concepts initiaux pour mieux cibler le domaine d'étude et correspondre aux procédures réglementaires applicables au pharmacovigilant. Les terminologies de référence utilisées par les CRPV servent donc de ressources au système d'annotation. Nous avons ainsi intégré les lexiques issus de Thesorimed, Drug Bank et RXNorm pour enrichir le concept "Drug Class and Active Substances", ajouté le référent des effets secondaires, le MedDRA, dans le concept "Adverse Event" avec les deux derniers niveaux de sa hiérarchie (Preferred Terms et Low Level Terms), et utilisé enfin les termes MeSH pour les concepts "Disorders", "Symptoms" et "Diagnostic Methods". L'intégration de la terminologie MedDRA dans la cartouche de connaissance sous le concept "Symptoms" a permis d'augmenter la couverture de reconnaissance des concepts natifs.

Une difficulté persiste dans la distinction d'un symptôme ou d'une maladie. L'indication médicale rencontrée peut être, en fonction du contexte, soit la maladie, soit l'effet secondaire. Or les lexiques n'intègrent pas nativement cette distinction et c'est plutôt la formulation de la phrase qui permet de désambiguïser. Un développement est en cours pour tenir compte du contexte sémantique et différencier les indications des effets indésirables de celles des maladies.

2.2.3 Exemple d'annotations obtenues

Afin d'illustrer les résultats obtenus, nous prenons l'exemple d'un résumé issu de PubMed qui a déjà été validé en amont par le CRPV de l'HEGP en tant qu'article d'intérêt.

[A case of hair loss induced by carbamazepine]

Date: 05 août 2004
 Auteur(s): Ishii A., Shoji S., Ichi, Kohno Y.
 Descripteurs clés: Alopecia, Administration, Carbamazepine, Neuropsychological Disease, Hypersensitivity, Carbamazepine (Cbz), Cbz, Alopecia effluvium, Dermatitis, Man

[Supprimer le marquage](#)

[A case of hair loss induced by carbamazepine]

We report a 52 year-old man presenting with an acute considerable hair loss induced by carbamazepine (CBZ). The remarkable scalp hair loss started within a week after CBZ administration. There was no evidence of dermatitis or allergic reaction, or other cause for the hair loss. The serum concentration of CBZ was 8.6 microg/ml (therapeutic range 8-12 microg/ml). CBZ was discontinued, and the hair loss stopped within several days with new hair growth. Medication-induced hair loss is an occasional adverse effect of many drugs used for neuropsychological diseases. CBZ also induces hair loss and its frequency was reported below 2%. Only a limited number of detailed case reports describing CBZ-induced hair loss were available, and we found these cases could divide into two groups with regard to a delay in starting hair loss after administration of CBZ. In one group, the hair loss started within a week suggesting anagen effluvium and in another it started after two or three months suggesting telogen effluvium. This finding suggests the causative mechanism of CBZ-induced hair loss is not unitary.

Fig. 3 –Résumé PubMed (PMID 15293764) – [Kohno & al., 2004]

Cette publication contient la phrase suivante: *We report a 52 year-old man presenting with an acute considerable hair loss induced by carbamazepine (CBZ)* qui est représentée dans Luxid[®] sous la forme d'un graphe (Fig. 4). Celui-ci indique qu'il existe une relation d'effet secondaire entre un patient, un traitement et un symptôme. Les informations originales subissent les actions suivantes: 1) Normalisation des termes extraits : "hair loss" est normalisé sous le terme MedDRA "Alopecia" ; 2) Qualification d'entités : le patient a pour caractéristiques d'être un homme, et d'être âgé de 52 ans ; 3) Mise en relation des événements : Patient + Symptôme + Médicament, liés par un verbe d'action ("induce") = "Adverse Effect" ; 4) Direction de l'événement: la relation d'effet indésirable a pour sujet la carbamazépine, et pour objet l'alopecie.

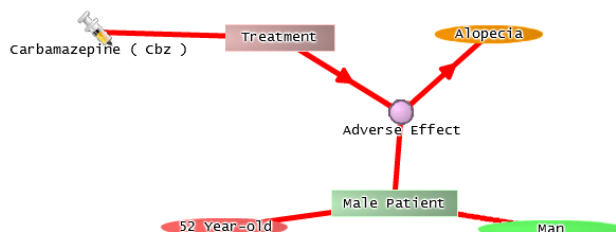


Fig. 4 – Relation d'effet indésirable impliquant la carbamazépine et l'alopecie

2.3 Consolidation des annotations produites

Avant de pouvoir consolider les informations extraites précédemment, deux étapes préliminaires au fonctionnement de la plateforme sont nécessaires [Amardeilh & al., 2007] :

- 1) Modéliser une ontologie de la pharmacovigilance et la charger dans le référentiel conjointement avec l'ensemble des terminologies médicales standard utilisées par l'application dans ses différentes phases et,
- 2) Définir et écrire l'ensemble des règles d'acquisition de connaissance qui vont permettre de transformer les annotations produites par l'outil d'extraction en instances de l'ontologie de la pharmacovigilance.

Nous nous sommes inspirés du contenu de la base de données du CRPV de l'HEGP ainsi que des pratiques des pharmacovigilants et de leurs besoins pour la modélisation de l'ontologie de la pharmacovigilance. Il s'agissait avant tout de pouvoir représenter les liens entre les traitements (décrits par un médicament ou une substance active et son contexte d'administration), les EI observés et le patient concerné. Dans la figure 5, les points représentent les classes et les arcs les relations sémantiques entre ces différentes classes de l'ontologie.

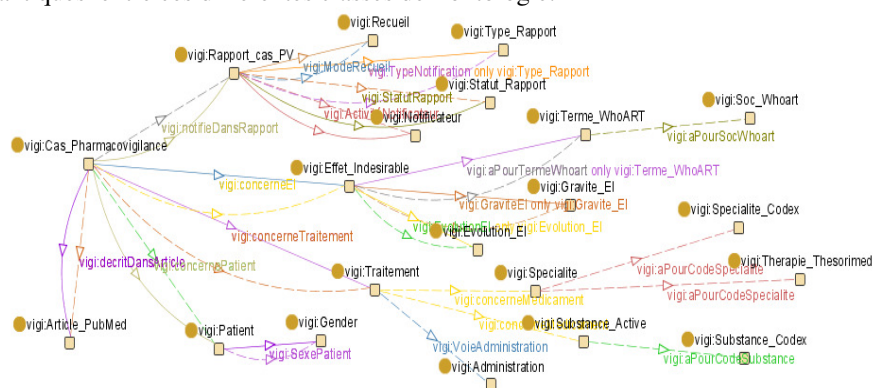


Fig. 5 – Ontologie de la pharmacovigilance pour VigiTermes

La figure 5 montre également que nous avons aligné la description des médicaments, substances actives et EI avec des terminologies médicales existantes et standard dans ce domaine : des lexiques de la base Thesorimed (symptômes et thérapies en français), le thésaurus MeSH (normalisation des symptômes et maladies en anglais), l'ontologie OntoEIM [Jaulent & al., 2009] (ontologie des EI avec alignement des terminologies WHO-ART, SnomedCT et MedDRA) et le lexique des médicaments de l'AFSSAPS (ou base Codex, liste des spécialités et substances actives officiellement répertoriées par l'AFSSAPS).

Nous avons importé toutes ces ressources dans le référentiel ITM (Intelligent Topic Manager), un outil spécialisé dans la gestion et la maintenance de ressources terminologiques et ontologiques, basé sur les langages du sémantique Web (RDF, OWL, SKOS). ITM est ici couplé avec le Content Augmentation Manager (CA Manager), un outil destiné à faciliter l'intégration entre les outils d'extraction d'information comme Luxid et les référentiels sémantiques comme ITM [Amardeilh & al., 2009]. Il joue un rôle de médiateur entre ces outils à l'aide de règles d'acquisition de connaissance (RAC) et étend l'infrastructure proposée par UIMA avec les standards du Web Sémantique (schéma d'annotation en RDF, web services, production des annotations sémantiques au format RDF par défaut, etc.).


```

NomRègle: CasPharmaco01
TypeConcept: Classe
ConceptURI: http://www.vigitermes.com/n#Cas\_Pharmacovigilance
NoeudIndicateur: AdverseEffect
IndicesContextuels:
  {Existe: [EspaceRechercheArbre: enfant]
   [NoeudIndice: Symptom]
  }
  {Existe: [EspaceRechercheArbre: enfant]
   [NoeudIndice: Treatment]
  }
  {Existe: [EspaceRechercheArbre: descendant]
   [NoeudIndice: LivingBeing]
   {Existe: [EspaceRechercheArbre : enfant]
    [NoeudIndice: MalePatient] OR [NoeudIndice:
    FemalePatient] OR [NoeudIndice: Patient] }
  }
Valeur: text()
Position: true
Confiance: élevé
finRègle

```

Fig. 6 – Exemple de RAC en langage OPAL

La deuxième étape préliminaire au fonctionnement de la chaîne d'annotation consiste donc à définir et implémenter les RAC nécessaires au couplage. Pour ce faire, nous comparons les étiquettes du graphe issu de l'analyse linguistique (cf. Fig. 4) avec les concepts, relations et attributs modélisés dans l'ontologie de la pharmacovigilance. Puis nous implémentons les règles à l'aide du langage OPAL (Ontology Population and Annotation Language) [Amardeilh & al., 2007]. Ce langage (cf. Fig. 6) se décompose en quatre parties définissant 1) l'action à mener, i.e. l'instanciation de la classe "Cas Pharmacovigilance" indiquée par son URI dans l'ontologie, 2) le nœud indicateur, i.e. l'étiquette du graphe sur lequel s'appuie l'instanciation (ex : le nœud "AdverseEffect"), 3) les indices contextuels, i.e. les conditions de réalisation de la règle sur ce nœud indicateur (ex : le nœud "AdverseEffect" doit avoir au moins un nœud fils "Symptom", un autre nœud fils "Treatment" qui lui-même possède un nœud descendant relatif à un nœud "Patient") et 4) les options d'application de cette règle, à savoir la valeur à récupérer pour créer le label de la nouvelle instance de "Cas Pharmacovigilance" (ici le texte du nœud indicateur "AdverseEffect"), les informations de positionnement (début et fin de la chaîne de caractères) et enfin le degré de confiance accordé à l'application de la règle. Les RACs sont toutes regroupées dans un fichier de mapping qui génère alors une feuille de transformation XSLT, appliquée sur chaque graphe exprimé par l'outil d'extraction d'information et produisant un réseau sémantique en RDF, selon le modèle de l'ontologie (en OWL).

Chaque nouvelle annotation ou instance générée est contrôlée dans le référentiel sémantique ITM afin de vérifier la non-redondance de la connaissance et la préservation de la cohérence et de la qualité de la base de connaissances. Un ensemble de règles de consolidation [Amardeilh & al., 2007] sont définies et appliquées sur chaque nouvelle annotation, instance de classe ou propriété. Ces règles peuvent être une simple requête sur le libellé ou bien plus complexes, multicritères. Ces règles de consolidation vont aussi vérifier la bonne conformité aux restrictions de domaine (appartenance d'une instance à la bonne classe ou à une de ses sous-classes), de range (appartenance d'une valeur de propriété à la bonne classe ou une de ses sous-classes)

dans le cas de relation, ou encore au bon type de données dans le cas d'attribut, et de cardinalité (respect du nombre autorisé de valeurs pour une propriété donnée) telles qu'elles ont été modélisées dans l'ontologie de l'application VigiTermes.

Si besoin, ces règles de consolidation peuvent être complétées par des règles de raisonnement et d'inférence permettant de vérifier des contraintes beaucoup plus pointues encore. Les informations non valides vis-à-vis du modèle ontologique ou de la base de connaissances sont mises de côté avec le statut "à valider" afin que l'application finale puisse remonter ces informations non valides au pharmacovigilant dans une interface de validation manuelle et qu'il puisse les corriger et les désambigüiser si nécessaire. L'apport des méthodes, langages et outils mis à disposition par la communauté du Web Sémantique permet donc d'aller plus loin dans l'analyse, la qualification, l'enrichissement et le contrôle de cohérence comparés aux applications uniquement basées sur un traitement automatique de la langue. Les deux approches sont donc fortement complémentaires et permettent d'offrir une solution de qualité aux pharmacovigilants. Non seulement les articles sont annotés, mais ils permettent la documentation des cas par le peuplement de la base de connaissance des cas rapportés. Ces nouvelles instances peuvent notamment être exploitées par les méthodes de raisonnement et de fouilles de données pour la détection de signaux liés à de nouveaux effets indésirables inattendus.

3 Retour d'expérience et évaluation de la plateforme

Évaluation du module de récupération des résumés PubMed

Les résultats d'évaluation de la récupération des résumés PubMed ont déjà été publiés [Delamarre & al., 2010]. Nous présenterons ici les résultats globaux obtenus.

Méthode	PubMed Requête directe	VigiPubMed		
		VigiPubMed Seul	VigiPubMed + extension MeSH	VigiPubMed + extension OntoEIM
Résumés	480	184	218	1780
Résumés pertinents	60	48	64	74
Rappel	63%	51%	67%	78%
Précision	13%	26%	29%	4%

Table 1. Évaluation des résultats obtenus par les quatre méthodes de recherche

Les mesures de rappel montrent que l'utilisation d'une recherche basée sur l'ontologie OntoEIM donne des résultats bien supérieurs à une recherche simple sur PubMed. Mais les résultats de l'évaluation conduisent finalement à retenir une

combinaison de ces différentes méthodes pour le projet VigiTermes. Dans un premier temps, une requête est envoyée via VigiPubMed, avec l'extension des termes MeSH ; elle permet d'avoir une liste de résumés réduite de 60% par rapport à une requête directe sur PubMed sans baisse du rappel (67%). Dans un second temps, si la première requête n'a apporté aucun résultat, ou encore si le pharmacovigilant souhaite élargir la revue bibliographique du cas, la requête est étendue avec les termes issus de l'ontologie OntoEIM. Ainsi, une interrogation faite sur *la thrombocytopénie* (EI) liée à la prise d'*infliximab* (PA) rapatrie deux publications. Trois publications supplémentaires peuvent être récupérées via l'extension basée sur l'ontologie, rapportant un effet hématologique proche. Ces documents additionnels sont utiles à l'analyse et à la compréhension du cas étudié. En revanche, la précision pour toutes ces méthodes doit être améliorée à l'aide d'outils complémentaires. La faible précision (4%) de la méthode utilisant l'ontologie OntoEIM est due à l'élargissement du contexte de la recherche, le résultat n'est pas surprenant mais l'objet de l'étude était bien de trouver une méthode donnant le meilleur rappel bibliographique. Pour pallier à l'imprécision, un travail en cours porte sur la présentation des résultats. En élaborant un score de pertinence, en utilisant les index, l'objectif est de fournir une interface qui permette de naviguer et d'atteindre les articles pertinents en quelques étapes.

Évaluation du module d'annotation

Pour effectuer une étude de la qualité de la Skill Cartridge™ Pharmacovigilance, nous avons choisi de nous placer dans les conditions du pharmacovigilant. Nous disposons d'un corpus de 150 résumés PubMed récupérés par le module VigiPubMed sur la base de la présence d'un EI et d'un PA dans le texte. Ces documents ont été repris dans Luxid® et annotés avec la Skill Cartridge™ Pharmacovigilance. Nous avons ensuite sélectionné uniquement les phrases qui comportaient à la fois un EI et un PA en résultat d'annotation.

Le CRPV de l'HEGP a participé à cette phase de qualité pour nous apporter leur expertise et valider les phrases annotées par la Skill Cartridge™ Pharmacovigilance. Chaque phrase a été labellisée juste ou fausse. Dans ce dernier cas, le pharmacovigilant nous a indiqué quelle était l'erreur d'annotation: silence, bruit, mauvaise catégorisation, termes incomplets, ... Nous n'avons pas convenu d'un système de priorité ou de poids sur les phrases non valides. Le pharmacovigilant décide de valider la phrase en fonction de la complétude de l'information retrouvée. Les erreurs, mêmes minimales (problèmes de formatage, redondance des annotations ou encore mauvaise catégorisation de l'information), entraînent la non-validation de la phrase. Nous envisageons pour des phases futures d'ajuster ce processus pour différencier les niveaux d'erreur de la Skill Cartridge™ Pharmacovigilance.

Les erreurs d'annotations sont redondantes et peuvent être facilement regroupées ; il s'ensuit qu'une correction mineure de la Skill Cartridge™ améliore significativement les résultats de qualité. L'évaluation finale est en cours.

Évaluation du module de consolidation des annotations

Ce travail est en cours d'évaluation aujourd'hui avec l'équipe du CRPV. Mais nous pouvons déjà signaler parmi les difficultés rencontrées la répercussion des problèmes liés à la non prise en compte du contexte entre effets indésirables et maladies dans la phase d'annotation. Les RAC ne peuvent également les distinguer car il n'existe pas aujourd'hui dans les graphes produits d'indices contextuels clairs que nous pourrions exploiter dans la partie conditions des règles.

Ceci est lié au second problème rencontré, celui de la dispersion de l'information dans le texte. En effet, les graphes ne rendent pas compte d'une certaine segmentation structurelle du résumé comme des paragraphes ou des phrases. Par conséquent, des sous-motifs arborescents adjacents dans le graphe ne peuvent être utilisés pour former des conditions car il n'y a pas d'information de localisation par rapport à une proposition ou une phrase donnée. Ceci peut être un avantage car nous pouvons très bien considérer que les résumés étant des textes courts, l'information extraite et présentée est forcément connectée.

4 Conclusion et perspectives

La littérature médicale est une source importante d'information pour évaluer la sécurité liée aux médicaments : c'est un élément essentiel pour alerter et renforcer la détection du signal. Alors que les laboratoires pharmaceutiques effectuent une surveillance systématique des effets indésirables pour les médicaments qu'ils commercialisent, les autorités régulatrices doivent suivre toutes les spécialités pharmaceutiques autorisées sur leur territoire. Dans tous les cas, le nombre d'articles scientifiques publiés au sujet d'effets indésirables liés à la prise de médicaments ne cesse de croître. Par conséquent, un outil automatisé est indispensable pour assurer cette surveillance. Des approches basées sur la connaissance ont déjà été développées pour étudier les médicaments et prédire la sécurité des médicaments lors de leur évaluation, à l'étape préclinique. Cette nouvelle plateforme de gestion de la connaissance s'intéresse à la documentation des cas rapportés spontanément après l'autorisation de mise sur le marché. Le projet de recherche VigiTermes permet d'implémenter de nouvelles méthodes pour l'analyse en pharmacovigilance et à plus long terme pour la détection du signal (détection d'effets nouveaux).

Nous venons de présenter la plateforme construite pour aider les équipes de pharmacovigilants. Les choix techniques se sont basés sur les technologies du Web Sémantique, incluant le développement d'ontologies et de services web, afin de faciliter l'intégration et l'échange de données entre les différents composants de la plateforme. Des développements additionnels sont nécessaires pour améliorer le rappel et la précision aux différentes étapes du processus : recherche sémantique étendue des résumés sur le portail PubMed, annotation des résumés trouvés et consolidation de ces annotations avant enrichissement de la base de connaissance pour la documentation des nouveaux cas rapportés. Les perspectives de développement de notre projet concernent l'élargissement des sources bibliographiques prises en compte, telles que le Martindale, le Meyler's Side effects of Drugs, et des bases de données comme le Micromedex. Ces autres sources vont nous permettre de compléter et de comparer les informations obtenues à partir de PubMed. La plateforme va aussi

intégrer d'autres outils pour la détection du signal, également développés dans le cadre du projet VigiTermes par d'autres partenaires et s'appuyant sur des méthodes statistiques. Enfin, il nous faut améliorer les interfaces utilisateurs sur le portail sémantique dédié au métier du pharmacovigilant.

Remerciements

Le projet VigiTermes est financé par l'Agence Nationale pour la Recherche (ANR) ANR-07-TECSAN-026 (<http://vigitermes.univ-rennes1.fr>)

Références

- ALECU I, BOUSQUET C, JAULENT MC (2008). A case report: using SNOMED CT for grouping Adverse Drug Reactions Terms. *BMC Med Inform Decis Mak.* 2008 Oct 27;8 Suppl 1:S4.
- ALECU I, BOUSQUET C, DEGOULET P, JAULENT MC (2007). PharmARTS: Terminology Web Services for Drug Safety Data Coding and Retrieval. *Medinfo 2007*. Ed : Kuhn KA, Warren JR, Leong TY. *Stud Health Technol Inform.* 2007; 129:699-704.
- AMARDEILH F (2007). Web Sémantique et Informatique Linguistique : propositions méthodologiques et réalisation d'une plateforme logicielle, Thèse de doctorat, Université Paris X, Nanterre.
- AMARDEILH F, DAMLIJANOVIC D (2009). Du texte à la connaissance : annotation sémantique et peuplement d'ontologie appliqués à des artefacts logiciels, *IC 2009*, Hammamet, Tunisie.
- AMARDEILH F, BOUSQUET C, GUILLEMIN-LANNE S, WISS-THEBAULT M, GUILLOT L, DELAMARRE D, LILLO-LE LOUËT A, BURGUN A (2009) A knowledge management platform for documentation of case reports in pharmacovigilance. *Stud Health Technol Inform.* 150, p517-521.
- AUBRY C, GRIVEL L, GUILLEMIN-LANNE S, LAUTIER C « Aide à la construction de composants de connaissance pour l'extraction d'information : méthodologie et environnement » CIFT 2002 Colloque International sur la Fouille de Textes, Hammamet- Tunisie, 21-23 octobre 2002.
- BPPV - BONNES PRATIQUES DE PHARMACOVIGILANCE – *Arrêté du 28/04/2005 – AFSSAPS*
- BODENREIDER O, NELSON SJ, HOLE WT, CHANG HF (1998). Beyond synonymy: exploiting the UMLS semantics in mapping vocabularies. *Proceedings / AMIA, Annual Symposium.* p 15–9.
- DELAMARRE D, LILLO-LE LOUËT A, GUILLOT L, JAMET A, SADOU E, OUAZINE T, BURGUN A, JAULENT M.C (2010). Documentation in Pharmacovigilance: using an ontology to extend and normalize Pubmed queries. *Medinfo 2010*, submitted.
- GARCELON N, MOUGIN F, BOUSQUET C, BURGUN A (2006). Evidence in pharmacovigilance: extracting Adverse Drug Reactions articles from MEDLINE to link them to case databases. *Stud Health Technol Inform.* 124, p 528-533
- JAULENT MC, ALECU I (2009). Evaluation of an ontological resource for pharmacovigilance. *Medical informatics in a united and healthy europe*. K.-P. Adlassning et al. (Eds); IOS Press, p 522-6
- KOHNO Y, ISHII A, SHOJI S.(2004). A case of hair loss induced by carbamazepine - *Rinsho Shinkeigaku.* 44(6) p 379-81.