HAL open science

Comparaison et visualisation des contre-indications des médicaments

Jean-Baptiste Lamy, Adrien Ugon, Madeleine Favre, Catherine Duclos, Alain Venot, Hélène Berthelot

To cite this version:

Jean-Baptiste Lamy, Adrien Ugon, Madeleine Favre, Catherine Duclos, Alain Venot, et al.. Comparaison et visualisation des contre-indications des médicaments. 3ème Symposium Ingénierie de l'Information Médicale (SIIM 2015), Jun 2015, Rennes, France. hal-01168293

HAL Id: hal-01168293

https://hal.science/hal-01168293v1

Submitted on 25 Jun 2015

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.

Comparaison et visualisation des contre-indications des médicaments ⋆

Jean-Baptiste Lamy[1], Adrien Ugon[1], Madeleine Favre[2], Catherine Duclos[1], Alain Venot[1], Hélène Berthelot[1]

LIMICS, Université Paris 13, Sorbonne Paris Cité, 93017 Bobigny, France, INSERM UMRS 1142, UPMC Université Paris 6, Sorbonne Universités, Paris
jean-baptiste.lamy@univ-paris13.fr, adrien.ugon@inserm.fr, catherine.duclos@avc.aphp.fr, alain.venot@univ-paris13.fr, helene.berthelot@orange.fr
Université Paris Descartes, Paris ; Société de Formation Thérapeutique du Généraliste (SFTG), Paris, France
mfavre89@gmail.com

Résumé : Les nouveaux médicaments représentent des opportunités de soin mais aussi des risques potentiels. Les médecins ont besoin d'une information indépendante et fiable leur permettant de comparer les nouveaux médicaments avec les anciens. La comparaison des propriétés entre médicament est cependant rendue difficile par l'expression de ces propriétés à des niveaux de granularité variables, par exemple "maladie hémorragique" contre "maladie hémorragique constitutionnelle".

Dans ce papier, nous proposons une approche à base d'ontologie formelle permettant la comparaison des contre-indications des médicaments et la génération automatique de tableaux comparatifs dynamiques.

Mots-clés : Ontologie, Contre-indication, Connaissances sur le médicament

1 Introduction

Les nouveaux médicaments représentent à la fois de nouvelles opportunités de soin mais aussi de nouveaux risques potentiels. Il est important que les médecins disposent d'une information indépendante et fiable sur ces nouveaux médicaments, cependant l'information actuelle provient principalement des visiteurs médicaux (Prosser *et al.*, 2003), financés par les laboratoires. Le projet VIIIP (Visualisation Intégrée de l'Information sur l'Innovation Pharmaceutique) a pour objectif de proposer une information indépendante et comparative sur les nouveaux médicaments. Nous souhaitons aussi une information qui ne soit pas l'opinion d'un expert mais plutôt la synthèse des données factuelles disponibles, car les experts sont rarement indépendants de l'industrie pharmaceutique et leur avis sont parfois divergeant.

Nous proposons ici une approche à base d'ontologie formelle pour comparer les contre-indications des médicaments. Les contre-indications sont des conditions cliniques dans lesquels un médicament ne doit pas être utilisé. Elles peuvent être consultées dans les documents de référence ou dans les bases de données sur le médicament (ou base médicament), cependant il existe peu d'outils permettant la comparaison des contre-indications de plusieurs médicaments de même indication. Les rares outils existants, tels que DOPAMINE (Wroe *et al.*, 2000), permettent la comparaison sous forme de tableau, mais

⋆. Ce travail a été financé par l'ANSM (Agence Nationale de Sécurité du Médicament et des produits de santé) au travers du projet de recherche VIIIP (AAP-2012-013).

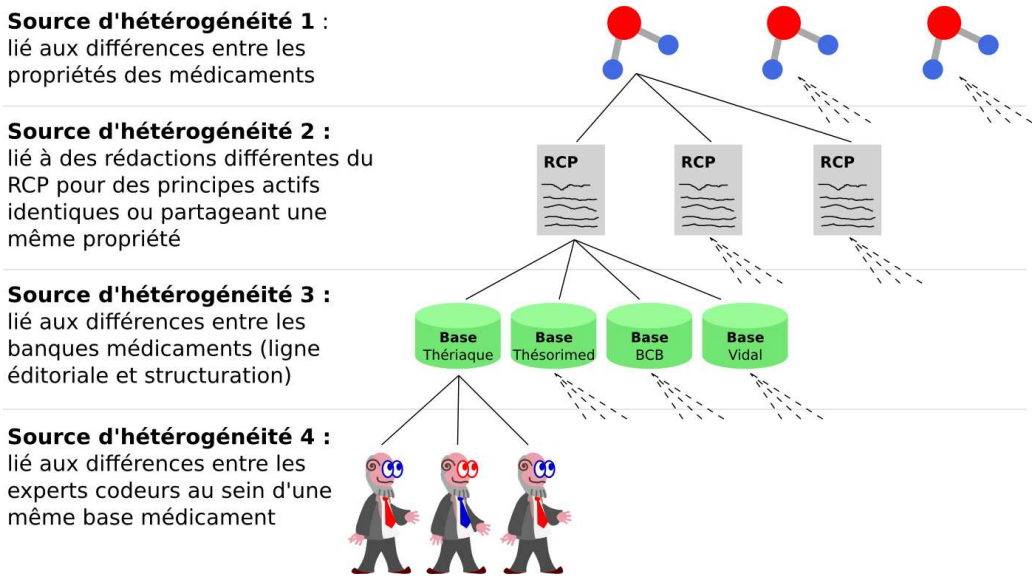

FIGURE 1 – Les quatre sources d'hétérogénéité identifiées dans les informations issues des bases médicament.

Condition clinique	ticagrélor	aspirine	héparine
maladie hémorragique	CI		
maladie hémorragique acquise		CI	
maladie hémorragique constitutionnelle		CI	CI

Condition clinique	ticagrélor	aspirine	héparine
maladie hémorragique	CI	CI	CI/ok
maladie hémorragique acquise	CI	CI	ok
maladie hémorragique constitutionnelle	CI	CI	CI

TABLE 1 – Trois contre-indications pour trois médicaments, issues de la base médicament Thériaque en haut, et telles qu'interprétées par un expert en bas (CI : contre-indiqué, ok : absence de contre-indication, CI/ok : contre-indiqué dans certaines situations seulement).

ne prennent pas en compte les différents niveaux de granularité utilisés pour décrire les contre-indications, certaines contre-indications étant plus générales et pouvant en regrouper d'autres plus spécifiques.

Nous commencerons par présenter la comparaison des contre-indications des médicaments. Nous présenterons ensuite la construction d'une ontologie, reliée à une base médicament, puis l'interface de visualisation permettant la comparaison des médicaments. Pour finir, nous discuterons les difficultés rencontrées et les perspectives.

2 Le problème de la comparaison des contre-indications des médicaments

Les médicaments peuvent être décrits par un ensemble de propriétés cliniques, telles que les indications, contre-indications, effets indésirables, précautions d'emploi, et interactions

médicamenteuses. La plupart des propriétés cliniques concernent la sécurité du médicament et se composent de listes souvent longues d'items, par exemple de contre-indications. Ces propriétés sont décrites dans le Résumé des Caractéristiques Produits (RCP) qui est le document de référence, rédigé par le laboratoire puis validé par les agences de santé (au niveau national ou européen) et remis à jour tout au long de la commercialisation du médicament. Les RCP sont des documents qui suivent un plan standard mais restent des textes non structurés ni codés. Les RCP sont ensuite repris par les éditeurs de bases médicament qui se chargent de structurer les connaissances présentes dans ces textes. En France, quatre bases médicament sont disponibles : Thériaque, Thésorimed, Banque Claude Bernard (BCB) et Vidal, chaque base ayant elle-même plusieurs experts à son service pour le codage de l'information.

Ce processus complexe de rédaction, de structuration et de codage conduit à une hétérogénéité importante dans les connaissances disponibles. La Figure 1 montre les quatre sources d'hétérogénéités que nous avons identifiées : (1) les différences réelles entre médicaments, (2) la rédaction du RCP : par exemple pour une même molécule, l'aspirine, certains RCP font apparaître des contre-indications supplémentaires, comme la goutte, (3) le choix de la base médicament, chaque base ayant une approche éditoriale et une structure différente, (4) l'existence de plusieurs experts pour chaque base ayant chacun des habitudes de codage différentes. La première source correspond aux différences que nous souhaitons mettre en évidence, et il est facile d'éviter la troisième en se limitant à une seule base. Cependant les deux sources restantes compliquent la lecture des propriétés d'un médicament, et encore plus la comparaison des propriétés de plusieurs médicaments.

La Table 1 (haut) montre trois exemples de situations de contre-indications pour trois médicaments différents, extraits de la base Thériaque. Cependant ce simple tableau ne permet pas une comparaison aisée. En effet, bien que cela n'apparaisse pas dans le tableau, le ticagrélor est *a fortiori* contre-indiqué avec les maladies hémorragiques acquises, et constitutionnelles, puisque contre-indiqué dans l'ensemble des maladies hémorragique (raisonnement par héritage). Et l'aspirine est contre-indiqué dans toutes les maladies hémorragiques puisque contre-indiqué à la fois dans les maladies acquises et constitutionnelles (raisonnement par partition). La Table 1 (bas) montre l'interprétation que ferait un expert à partir des connaissances de la Table 1 (haut). Enfin, il est possible de déduire les situations dans lesquelles un médicament *n'est pas* contre-indiqué, par exemple les maladies hémorragiques acquises pour l'héparine ; ces absences de contre-indication ("ok" dans la Table 1 en bas) ne doivent pas être confondues avec l'absence de mention de la contre-indication dans la base médicament (cases vides dans la Table 1 en haut).

C'est ce raisonnement que nous souhaitons automatiser, afin de faciliter la comparaison des contre-indications entre médicaments.

3 Construction d'une ontologie des contre-indications

Nous avons structuré les contre-indications avec une ontologie formelle (Figure 2). Dans la base Thériaque, les contre-indications sont décrites par deux termes issus du thésaurus CDF (CoDiFication) de Thériaque : un terme de niveau 1 et un terme de niveau 2, et un niveau de contre-indication. Les termes de niveau 1 restent généraux et sont organisés en arbre. Les termes de niveau 2 sont plus spécifiques et ne sont pas organisés (liste plate).

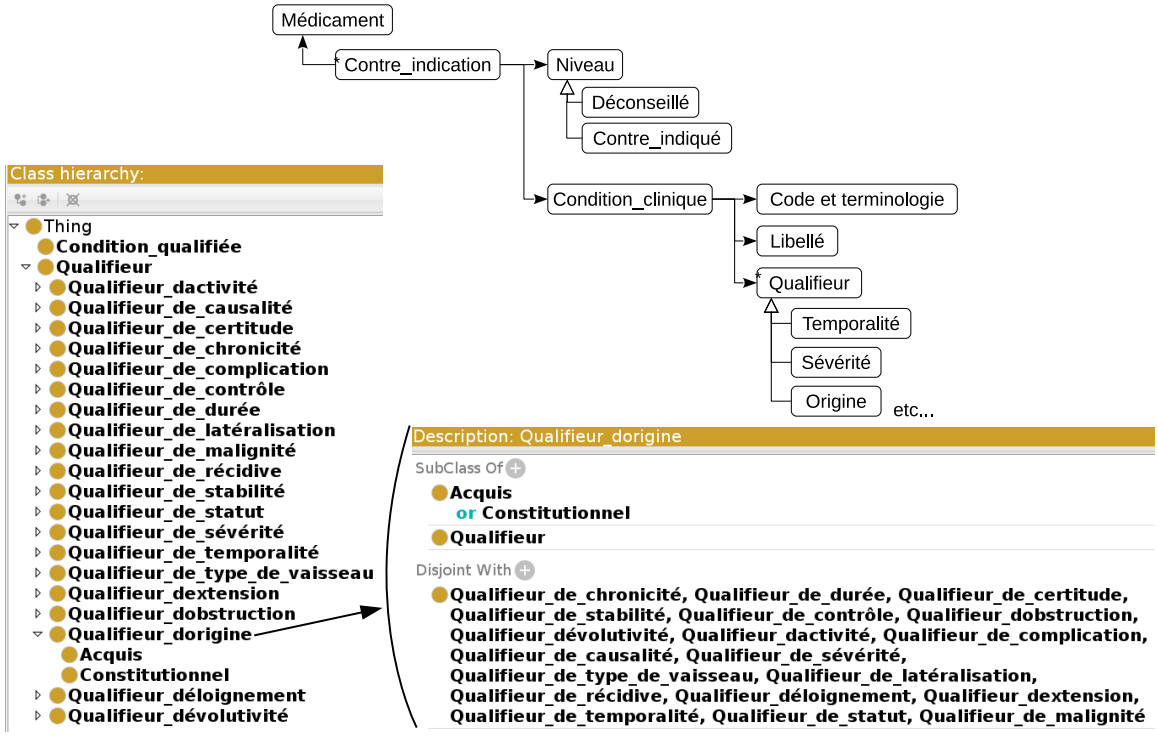

FIGURE 2 – Schéma de l'ontologie des contre-indications (en haut) et copie d'écran dans Protégé montrant les qualifieurs (en bas).

Souvent, le terme de niveau 2 reprend le niveau 1 en ajoutant une précision, par exemple "insuffisance hépatique" (niveau 1) et "insuffisance hépatique sévère" (niveau 2).

Cette structuration n'est pas suffisante pour une ontologie, c'est pourquoi nous avons (1) enrichi les relations existantes entre termes et (2) décomposé les termes de niveau 2 lorsqu'il s'agit de "compléments" au niveau 1, et structuré ces compléments en qualifieurs.

Premièrement, nous avons utilisé les liens existant entre CDF et la CIM10 pour établir de nouvelles relations d'héritage et d'équivalence entre termes d'un même niveau. Si deux termes ont les mêmes correspondances CIM10, ils ont été proposé comme équivalents ; si un terme a des correspondances CIM10 plus spécifiques qu'un autre, alors le premier terme est proposé comme fils du second. PyMedTermino (Lamy *et al.*, 2015) a été utilisé pour gérer le thésaurus CDF et les correspondances CIM10. Les relations proposées (305 équivalences et 1731 héritages) ont été revues par un expert médical (HB). 97 relations d'équivalence ont été conservées, et 1233 relations d'héritage. Ces relations conduisent à une structure multiaxiale plus riche que la structure arborescente initiale ; par exemple "hémorragie cérébrale" est à la fois une "hémorragie" et une "maladie cérébrovasculaire".

Deuxièmement, nous avons cherché à décomposer les termes de niveau 2 en effectuant la "différence" terme de niveau 2 - terme de niveau 1 sur l'ensemble des contre-indications pour lesquelles le terme de niveau 2 reprenait le terme de niveau 1 (soit 1076 contre-indications sur 2325). Après avoir retiré les doublons, nous avons obtenu une liste de 249 compléments de libellé. Cette liste a ensuite été traitée manuellement pour (a) découper les compléments comprenant plusieurs qualifieurs distincts (par exemple "antécédent sé-

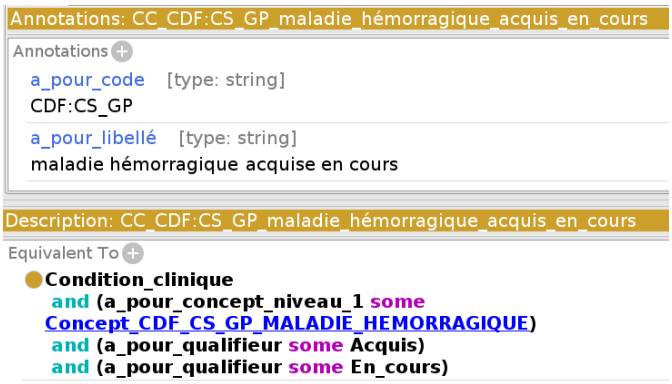

FIGURE 3 – Exemple de classe définie de condition clinique, extraite de Thériaque puis visualisée dans Protégé. Cette condition est décrite par un terme de niveau 1 et deux qualifieurs.

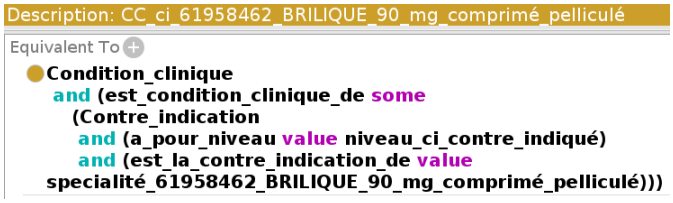

FIGURE 4 – Exemple de classe définie permettant d'obtenir après classification la liste des conditions cliniques dans lesquelles le ticagrélor est contre-indiqué (Brilique est le nom commercial du ticagrélor).

vère"), (b) retirer les doublons lexicaux ("acquis" *vs* "acquise") et les synonymes ("non corrigé" et "non contrôlé") et (c) regrouper les qualifieurs en catégorie, selon la qualité qu'ils renseignent (par exemple "léger", "modéré" et "sévère" expriment la sévérité).

20 catégories ont été trouvées ; elles ont ensuite été modélisées dans une ontologie formelle en langage OWL 2 avec Protégé 4 (Figure 2 en bas à gauche). Des disjonctions et des relations de partitions ont été intégrées dans l'ontologie (Figure 2 en bas à droite). Pour chaque qualifieur, nous avons défini sous forme d'annotations le ou les morceaux de libellés CDF correspondants (par exemple "acquis" et "acquise" pour le qualifieur Acquis).

Une ontologie des contre-indications a ensuite été modélisée dans Protégé 4 (Figure 2 en haut). Dans cet ontologie, un médicament peut avoir plusieurs contre-indications. Une contre-indication associe un niveau (seuls les niveaux contre-indiqué et déconseillé ont été retenus pour le projet) et une condition clinique. La condition clinique comprend un libellé textuel, un terme de niveau 1, un éventuel terme de niveau 2 et un ou plusieurs qualifieurs.

4 Peuplement de l'ontologie des contre-indications et classification

Le peuplement de l'ontologie a été réalisé à partir de la base Thériaque. Dans un premier temps, nous avons extraits les médicaments et les contre-indications (c'est-à-dire

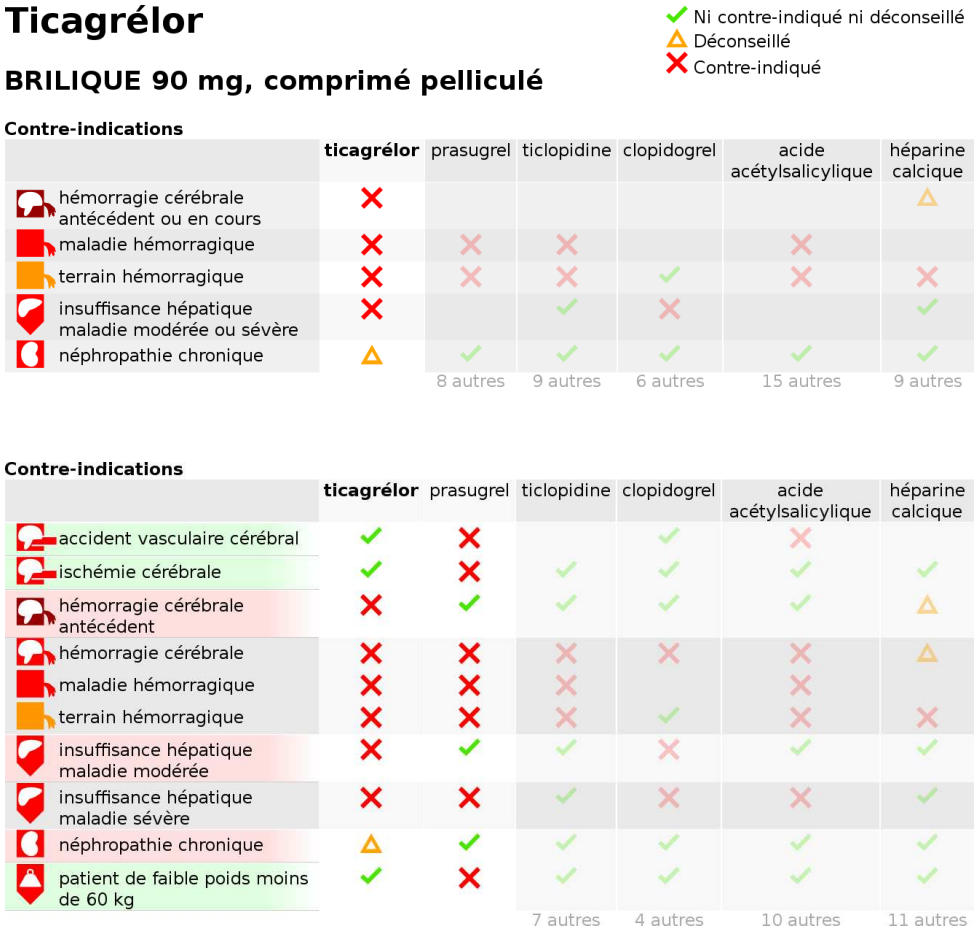

FIGURE 5 – Copie d'écran de la visualisation de la comparaison des contre-indications du ticagrélor avec 5 comparateurs (en haut : contre-indications du ticagrélor, en bas après sélection du comparateur prasugrel).

l'équivalent de la Table 1). Les conditions cliniques des contre-indications sont représentées par des classes définies (Figure 3) et non des instances, afin de pouvoir prendre en compte les relations *est-un* existant entre conditions cliniques (par exemple maladie hémorragique acquise *est une* maladie hémorragique). Lors de l'extraction, les termes de niveau 2 de Thériaque sont décomposés en qualifieurs lorsque cela est possible, en recherchant les libellés des qualifieurs dans le terme de niveau 2. Lorsque le terme de niveau 2 n'apporte rien de plus que la somme du terme de niveau 1 et des qualifieurs, celui-ci est supprimé. Dans un second temps, nous avons généré pour chaque médicament 3 classes, correspondant respectivement aux conditions cliniques contre-indiquées, déconseillées, ou "ni contre-indiquées ni déconseillées" pour ce médicament (Figure 4).

L'utilisation d'un raisonneur permet alors (a) de classer les conditions cliniques entre elles, par exemple de déduire que maladie hémorragique acquise *est une* maladie hémorragique, et (b) de classer les conditions cliniques dans les classes de conditions contre-indiquées, déconseillées, ou "ni contre-indiquées ni déconseillées" de chaque médicament.

5 Visualisation

Une fois l'ontologie classifiée, nous pouvons générer un tableau dynamique en HTML semblable à la Table 1 (bas). Trois symboles sont utilisés pour indiquer les trois niveaux de contre-indication retenus dans le projet (contre-indiqué, déconseillé, ni contre-indiqué ni déconseillé). À cause du nombre important de contre-indications, toutes ne sont pas affichées. Par défaut, seules les contre-indications du nouveau médicament sont visibles (Figure 5 haut). Pour les comparateurs, les informations correspondantes à ces contre-indications sont affichés lorsque cela est possible : par exemple le ticagrélor est contre-indiqué dans les terrains hémorragiques et en cas d'insuffisance hépatique modérée ou sévère ; le prasugrel est aussi contre-indiqué dans les terrains hémorragiques, mais la case de l'insuffisance hépatique est laissée vide car il est contre-indiqué dans seulement certaines insuffisance hépatique. L'interface intègre aussi des icônes VCM (*Visualisation des Connaissances Médicales*) (Lamy *et al.*, 2009).

En cliquant sur l'un des comparateurs, celui-ci est sélectionné pour une comparaison deux à deux (Figure 5 bas). L'ensemble des contre-indications des deux médicaments apparaît alors. Dans l'exemple précédent, l'insuffisance hépatique modérée ou sévère est alors séparée en deux ligne, insuffisance hépatique modérée et insuffisance hépatique sévère. Le prasugrel n'est contre-indiqué que dans la seconde. Les lignes où les deux médicaments diffèrent sont mises en surbrillance, avec une couleur rouge, ou vert, pâle sur la gauche selon que le nouveau médicament est davantage, ou moins, contre-indiqué que le comparateur.

6 Implémentation

Le peuplement de l'ontologie nécessite de requêter une base médicament (en SQL), et l'affichage des résultats de la classification en tableaux dans un site Web nécessite la génération de pages HTML. Ces deux opérations requièrent l'utilisation d'un langage de programmation, lequel doit pouvoir accéder à l'ontologie en lecture et en écriture, et effectuer la classification. L'implémentation du système a été réalisée en langage Python, avec le module Ontopy [1] (Lamy & Berthelot, 2015) qui permet la programmation orientée ontologie (Goldman NM, 2003) : Ontopy permet de créer et manipuler les classes et les instances OWL comme des objets Python, et de classifier automatiquement des classes et des instances *via* le raisonneur Hermit (Motik *et al.*, 2009). Nous avons aussi utilisé le module PyMedTermino pour accéder aux terminologies médicales (Lamy *et al.*, 2015).

7 Discussion et conclusion

Dans cet article, nous avons présenté le problème de la comparaison des contre-indications des médicaments, ainsi qu'une approche à base d'ontologie formelle pour automatiser cette comparaison et générer des tableaux dynamiques facilitant la comparaison des médicaments par un humain.

Notre modèle des contre-indications est similaire à celui proposé dans la littérature (Liu *et al.*, 1998) ; il utilise cependant un formalisme d'ontologie plus riche que le modèle

1. Disponible en ligne sous licence libre GNU LGPL v3 : `https://bitbucket.org/jibalamy/ontopy`

objet proposé par Liu *et al.* Un problème de représentation des connaissances a été rencontré lors de la modélisation de l'insuffisance hépatique. En effet, après discussion avec les experts médicaux, il apparaît que :

l'insuffisance hépatique ***est une*** atteinte hépatique
l'insuffisance hépatique modérée ***est une*** insuffisance hépatique
l'atteinte hépatique sévère ***est une*** atteinte hépatique
l'insuffisance hépatique **modérée** ***est une*** atteinte hépatique **sévère**

Le modèle devra donc être revu car reporter par héritage les relations existant entre deux conditions cliniques (par exemple A et B) sur ces deux mêmes conditions cliniques à un stade de sévérité précis (par exemple A modéré et B modéré) n'est pas suffisant.

Notre approche a été appliquée aux contre-indications, cependant la plupart des propriétés des médicaments (interactions médicamenteuses, effets indésirables, précautions d'emploi, allergies croisés, présences d'excipients à effet notoire...) suivent un schéma similaire. Les perspectives de ce travail incluent donc l'extension aux autres propriétés, ainsi que l'évaluation de l'extraction automatique des données à partir de la base médicament (en comparant à une extraction manuelle par un expert) et de la visualisation auprès de médecins (en comparant à la consultation des documents papier existant).

Une dernière perspective consisterait à combiner et à recouper les informations de plusieurs bases médicaments. Une telle approche est cependant complexe à cause des terminologies différentes utilisées dans chaque base, ainsi que méthodes de codages différentes (un seul code par contre-indication ou plusieurs, sur un ou deux niveaux, codage dague + étoile en CIM10, *etc*). Nous expérimentons actuellement une variante utilisant un codage en CIM10, qui pourrait être utilisée comme terminologie pivot entre les différentes bases.

Références

Goldman NM (2003). Ontology-oriented programming : static typing for the inconsistent programmer. In *Lecture notes in computer science : the SemanticWeb, ISWC*, volume 2870, p. 850–865.

Lamy J. B. & Berthelot H. (2015). Ontopy : programmation orientée ontologie en Python. In *Actes du congrès d'Ingénierie des Connaissances (IC2015)*, Rennes, France.

Lamy J. B., Duclos C. & Venot A. (2009). De l'analyse d'un corpus de texte à la conception d'une interface graphique facilitant l'accès aux connaissances sur le médicament. In *Actes des 20es Journées Francophones d'Ingénierie des Connaissances*, volume 265-276, Hammamet, Tunisie : PUG.

Lamy J. B., Venot A. & Duclos C. (2015). PyMedTermino : an open-source generic API for advanced terminology services. *Stud Health Technol Inform.*

Liu J. H., Milstein C., Séné B. & Venot A. (1998). Object-oriented modeling and terminologies for drug contraindications. *Methods Inf Med*, **37**, 45–52.

Motik B., Shearer R. & Horrocks I. (2009). Hypertableau reasoning for description logics. *Journal of Artificial Intelligence Research*, **36**, 165–228.

Prosser H., Almond S. & Walley T. (2003). Influences on GPs'decision to prescribe new drugs-the importance of who says what. *Fam Pract*, **20**(1), 61–68.

Wroe C., Solomon W., Rector A. & Rogers J. (2000). DOPAMINE : a tool for visualizing clinical properties of generic drugs. In *Proceedings of the Fifth Workshop on Intelligent Data Analysis in Medicine and Pharmacology (IDAMAP)*, volume 61-65.