

# Extraction d'association d'EIM à partir de dossiers patients : expérimentation avec les structures de patrons et les ontologies

Gabin Personeni, Marie-Dominique Devignes, Michel Dumontier, Malika Smaïl-Tabbone, Adrien Coulet

## ► To cite this version:

Gabin Personeni, Marie-Dominique Devignes, Michel Dumontier, Malika Smaïl-Tabbone, Adrien Coulet. Extraction d'association d'EIM à partir de dossiers patients : expérimentation avec les structures de patrons et les ontologies. Deuxième Atelier sur l'Intelligence Artificielle et la Santé, Jun 2016, Montpellier, France. 2016, Atelier IA & Santé. <<https://ia-sante-2016.sciencesconf.org/>>. <hal-01391172>

**HAL Id: hal-01391172**

**<https://hal.inria.fr/hal-01391172>**

Submitted on 2 Nov 2016

**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.

# **Extraction d'association d'EIM à partir de dossiers patients : expérimentation avec les structures de patrons et les ontologies**

Gabin Personeni<sup>1</sup>, Marie-Dominique Devignes<sup>1</sup>, Michel Dumontier<sup>2</sup>, Malika  
Smail-Tabbone<sup>1</sup> et Adrien Coulet<sup>1</sup>

<sup>1</sup>LORIA (CNRS, Inria NGE, Université de Lorraine), Vandœuvre-lès-Nancy, F-54506, France

<sup>2</sup>Stanford Center for Biomedical Informatics Research, Stanford University, Stanford, CA, Etats-Unis

**Résumé :** Les Dossiers Médicaux Electroniques (DME) constituent une ressource de grand intérêt pour étudier les Evènements Indésirables Médicamenteux (EIM). Nous proposons ici de fouiller les DME pour identifier des EIM fréquemment associés dans des sous-groupes de patients. Les EIM ayant des manifestations complexes, nous utilisons l'analyse formelle de concepts et ses structures de patrons, un cadre mathématique permettant la généralisation, en exploitant les connaissances du domaine médical formalisées dans des ontologies. Les résultats obtenus dans trois expériences montrent que cette approche est flexible et permet d'extraire des règles d'association à divers niveaux de généralisation.

## **1 Introduction**

Les Evènements Indésirables Médicamenteux (EIM) apparaissent inégalement dans différents sous-groupes de patients. Leurs causes sont multiples: génétiques, métaboliques, interactions de médicaments. Des études précédentes ont montré que les EIM pouvaient être détectés et étudiés en fouillant des Dossiers Médicaux Electroniques (DME) (LePendou, et al., 2013). Nous souhaitons explorer les DME pour révéler qu'un groupe de patients sensible aux effets secondaires d'un médicament est également sensible aux effets secondaires d'un autre. Pour cela, nous proposons une méthode pour identifier des EIM fréquemment associés dans des sous-groupes de patients. Les manifestations des EIM étant variables et complexes, nous utilisons une extension de l'Analyse Formelle de Concepts (AFC) : les structures de patrons (Ganter & Kuznetsov, 2001), associée à des ontologies pour permettre la généralisation des EIM extraits des DME. Nous avons utilisé un jeu de DME de patients diagnostiqués avec le Lupus Erythémateux Disséminé (LED), une maladie auto-immune. Ces patients sont souvent sujets aux EIM, de par les traitements pour le LED et les maladies opportunistes qui l'accompagnent (Vasudevan & Ginzler, 2009). Ces DME ont été extraits de STRIDE, l'entrepôt de données clinique de l'Hôpital universitaire de Stanford (Lowe, et al., 2009).

## **2 Matériel et méthodes**

### **2.1 Corpus de données**

Notre corpus de données est un ensemble de 6 869 DME, de patients anonymisés, diagnostiqués avec le LED. Ce corpus documente 451 000 visites à l'hôpital, avec leur dates relatives, diagnostics encodés en codes ICD9-CM (International Classification of Diseases, Clinical Modification) et prescriptions sous la forme de listes d'ingrédients, représentés par leurs identifiants RxNorm. Aussi, nous employons le terme « médicament » pour désigner un ingrédient actif plutôt que le médicament sous sa forme commerciale.

Afin d'identifier des EIM fréquemment co-occurents, nous devons d'abord extraire ces EIM des DME, puis sélectionner les patients qui en présentent au moins deux. Nous établissons d'abord, pour chaque patient, une liste d'EIM candidats. Pour deux visites consécutives dans le DME, nous extrayons l'ensemble des médicaments prescrits (une prescription) dans la première visite  $D_i$ , et les phénotypes  $P_i$  diagnostiqués durant la seconde. L'intervalle entre deux visites doit être d'au plus 14 jours : il est raisonnable de supposer

qu'un effet secondaire se manifeste dans un court délai après la prescription. De plus, nous avons pu observer qu'augmenter cet intervalle maximal n'augmente pas le nombre de patients retenus dans notre corpus, comme illustré en Table 1.

Un EIM candidat  $C_i$  est donc un couple d'ensembles  $C_i = (D_i, P_i)$ . Nous ne conservons dans  $P_i$  que les phénotypes listés dans SIDER 4.1 comme effets secondaires d'un médicament de  $D_i$ . SIDER est une base de données d'indications et d'effet secondaires de médicaments (Kuhn, et al., 2016). Nous excluons ensuite les candidats où  $P_i$  est vide. Nous excluons aussi un EIM candidats  $(D_1, P_1)$  si il existe un autre EIM candidat  $(D_2, P_2)$  pour le même patient tel que  $D_1 \subseteq D_2$  : en effet, si une prescription est répétée pour un patient, cela indique qu'elle n'a pas été jugée dangereuse pour lui.

TABLE 1 – Nombre de patients avec au moins 2 EIM sélectionnés, et nombre d'EIM pour ces patients, pour différents intervalles maximum entre deux visites.

Intervalle (jours)	1	2	6	10	14	18	22	26	30
Patients	434	461	498	526	548	555	558	564	576
EIM	2 396	2 587	2 902	3 110	3 286	3 388	3 454	3 501	3 621

Par ce processus de sélection, nous obtenons un corpus de 3 286 EIM provenant de 548 patients présentant au moins 2 EIM. La Table 2 présente des exemples d'EIM qui pourraient être extraits de DME, qui serviront à illustrer les expériences présentées dans cet article.

TABLE 2 – Exemple de corpus avec 3 patients ayant présenté 2 EIM chacun

Patients	EIM
P1	({prednisone}, {ICD 599.8}); ({acetaminophen}, {ICD 599.9})
P2	({prednisone}, {ICD 599.8}); ({prednisone}, {ICD 719.4})
P3	({prednisone, acetaminophen}, {ICD 599.9}); ({acetaminophen}, {ICD 719.4})

## 2.2 Analyse Formelle de Concepts et structures de patrons

L'AFC (Ganter & Wille, 1999) est un cadre mathématique pour l'organisation d'un ensemble de données dans un treillis de concepts formels, *i.e.*, une structure hiérarchique où un concept représente un ensemble d'objets partageant des propriétés. Il permet l'extraction de règles d'associations.

En AFC, les données sont un ensemble d'objets, chacun décrit par un ensemble d'attributs binaires. Les structures de patrons généralisent l'AFC pour l'appliquer à des objets munis d'une description de nature complexe (pas forcément binaire), par exemple : des ensembles, graphes, intervalles, annotations par des classes d'une ontologie (Ganter & Kuznetsov, 2001; Coulet, et al., 2013). Une structure de patron est un triplet  $(G, (\mathcal{D}, \sqcap), \delta)$  où (i)  $G$  est un ensemble d'objets, (ii)  $\mathcal{D}$  est un ensemble de descriptions, (iii)  $\delta$  est une fonction qui associe un objet à sa description, (iv)  $\sqcap$  est un opérateur définissant un ordre partiel  $\leq_{\sqcap}$  sur les éléments de  $\mathcal{D}$ , appelé *INF*, tel que  $X \sqcap Y$  est la description la plus spécifique qui est plus générale que  $X$  et  $Y$ . Ainsi  $X \leq_{\sqcap} Y$ , signifiant que  $Y$  est plus spécifique que  $X$  est équivalent à  $X \sqcap Y = X$ . Cet opérateur permet la généralisation des descriptions d'objets. Ceci sera illustré pour les opérateurs que nous définirons en Section 3.

Dans les structures de patrons, l'opérateur  $\cdot^{\square}$  définit une connexion de Galois entre des ensembles d'objets et des descriptions, tel que :

$$A^{\square} = \sqcap_{g \in A} \delta(g) \text{ pour tout ensemble d'objets } A \subseteq G$$

$$d^{\square} = \{g \in G \mid d \leq_{\sqcap} \delta(g)\} \text{ pour toute description } d \in \mathcal{D}$$

Un concept de patrons est une paire  $(A, d)$  vérifiant  $A^{\square} = d$  et  $d^{\square} = A$ . Dans notre étude,  $G$  est un ensemble de patients, à qui on associe par  $\delta$  la description de leurs EIM dans  $\mathcal{D}$ . La Section 3 décrit trois expériences utilisant toutes les structures de patrons, mais avec une représentation différente des EIM et ainsi une structure de patrons  $(G, (\mathcal{D}, \sqcap), \delta)$  différente.

### 2.3 Ontologies médicales

Nous utilisons deux ontologies biomédicales : ICD9-CM qui décrit des classes de phénotypes et l'Anatomical Therapeutic Chemical Classification System (ATC) qui décrit des classes de médicaments. Nous considérons seulement la seule hiérarchie de classes de ces ontologies afin de généraliser la description des phénotypes et des prescriptions extraits des DME. De plus, nous utilisons uniquement les trois niveaux les plus spécifiques d'ATC : sous-groupes pharmacologiques, sous-groupes chimiques, substances chimiques.

## 3 Expériences

Dans cette Section, nous décrivons trois expériences utilisant les structures de patrons pour extraire des règles d'association entre des EIM. Chaque expérience utilise une représentation différente des EIM de chaque patient, chacune faisant un plus grand usage des ontologies.

### 3.1 Première expérience avec les structures de patrons

On définit ici la structure de patrons  $(G, (\mathcal{D}_1, \Pi_1), \delta_1)$  où les objets de  $G$  sont des patients, et les descriptions dans  $\mathcal{D}_1$  sont des vecteurs de *sous-descriptions*. Chaque sous-description est un ensemble d'ensembles de médicaments, *i.e.*, un ensemble de prescriptions, associé à une classe de phénotypes du premier niveau d'ICD.

TABLE 3 – Exemple de descriptions de patients pour la structure de patrons  $(G, (\mathcal{D}_1, \Pi_1), \delta_1)$ , avec deux classes ICD de premier niveau : maladies du système urogénital (580-629), et maladies du système musculo-squelettal system et des tissus conjonctifs (710-739).

	ICD 580-629	ICD 710-739
Patient P1	{{prednisone}, {acetaminophen}}	$\emptyset$
Patient P2	{{prednisone}}	{{prednisone}}
Patient P3	{{prednisone, acetaminophen}}	{{acetaminophen}}

Par exemple, en considérant uniquement les deux classes ICD de la Table 3, les deux sous-descriptions associées au patient P1 sont :

$$\delta_{1, \text{ICD } 580-629}(\text{P1}) = \{\{\text{prednisone}\}, \{\text{acetaminophen}\}\} \quad \text{et} \quad \delta_{1, \text{ICD } 710-739}(\text{P1}) = \emptyset$$

Les sous-descriptions sont associées à un des classes de premier niveau d'ICD pour représenter des EIM : le patient présente un phénotype de cette classe après avoir été prescrit l'un des ensembles de médicaments dans cette sous-description. Nous définissons les sous-descriptions comme des ensembles de prescriptions, où aucune prescription n'est comparable à une autre par l'ordre partiel  $\subseteq$ . Nous définissons alors l'opérateur  $\Pi_1$  tel que, pour deux descriptions  $X$  et  $Y$  de  $\mathcal{D}_1$  quelconques :

$$X \Pi_1 Y = \max(\subseteq, \{x \cap y \mid (x, y) \in X \times Y\})$$

où  $\max(\subseteq, S)$  est l'unique sous-ensemble des éléments maximaux de l'ensemble  $S$  pour l'ordre partiel  $\subseteq$ . Formellement, on définit  $\max(\subseteq, S) = \{s \mid \nexists x. (s \subseteq x)\}$ . Dans le cas présent,  $\max$  ne conserve dans la description que les prescriptions les plus spécifiques. Par exemple, en considérant 4 médicaments  $d_1, d_2, d_3$ , et  $d_4$  :

$$\begin{aligned} & \{\{d_1, d_2, d_3\}\} \Pi_1 \{\{d_1, d_2\}, \{d_2, d_4\}\} \\ &= \max(\subseteq, \{\{d_1, d_2, d_3\} \cap \{d_1, d_2\}, \{d_1, d_2, d_3\} \cap \{d_2, d_4\}\}) \\ &= \max(\subseteq, \{\{d_1, d_2\}, \{d_2\}\}) \\ &= \{\{d_1, d_2\}\} \end{aligned}$$

On ne conserve que  $\{d_1, d_2\}$  puisque  $\{d_2\} \subseteq \{d_1, d_2\}$  et  $\{d_1, d_2\}$  est l'unique élément maximal par  $\subseteq$ . En effet la sémantique de  $\{d_2\}$  – une prescription qui contient le médicament

$d_2$  – est plus générale que la sémantique de  $\{d_1, d_2\}$  – une prescription qui contient les médicaments  $d_1$  et  $d_2$ .

A chaque patient est associée une sous-description pour chaque classe de premier niveau d'ICD, l'opérateur défini pour une seule sous-description peut être généralisé à un vecteur de sous-descriptions :

$$\begin{aligned} \delta_1(P1) \sqcap_1 \delta_1(P2) &= \langle \delta_{1,1}(P1), \dots, \delta_{1,n}(P1) \rangle \sqcap_1 \langle \delta_{1,1}(P2), \dots, \delta_{1,n}(P2) \rangle \\ &= \langle \delta_{1,1}(P1) \sqcap_1 \delta_{1,1}(P2), \dots, \delta_{1,n}(P1) \sqcap_1 \delta_{1,n}(P2) \rangle \end{aligned}$$

La Figure 1 présente le semi-treillis associé à la structure de patron et les données de la Table 3. Cet exemple illustre qu'avec cette structure de patron les informations sur les EIM sont perdues lors de la généralisation.

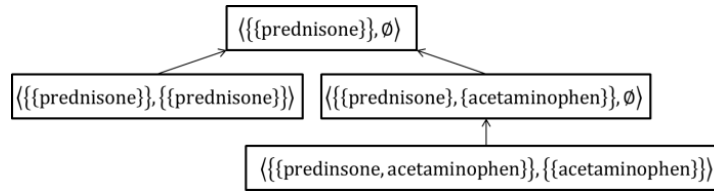


FIGURE 1 – Représentation des données de la Table 3 dans un semi-treillis construit à partir de la structure de patrons  $(G, (\mathcal{D}_1, \sqcap_1), \delta_1)$ , où les flèches représentent l'ordre partiel  $\leq_{\sqcap_1}$  défini par  $\sqcap_1$ .

### 3.2 Extension de la structure de patrons avec une ontologie de médicaments

Nous proposons de représenter les médicaments avec les termes de l'ontologie ATC afin de trouver des associations entre des EIM concernant des classes de médicaments plutôt qu'un simple médicament. Pour cela, nous devons étendre la structure de patrons décrite précédemment pour considérer cette ontologie. Chaque médicament est alors identifié par sa ou ses classes ATC, comme illustré en Table 4.

TABLE 4 – Exemple de descriptions de patients pour la structure de patrons  $(G, (\mathcal{D}_2, \sqcap_2), \delta_2)$ . H02AA03 : desoxycortone, H02AB07 : prednisone, N02BE01 : acetaminophen.

	ICD 580-629	ICD 710-739
Patient P1	{{H02AB07}, {N02BE01}}	∅
Patient P2	{{H02AB07}}	{{H02AB07}}
Patient P3	{{H02AB07}, {N02BE01}}	{{N02BE01}}
Patient P4	{{H02AA03}}	∅

On définit la seconde structure de patrons  $(G, (\mathcal{D}_2, \sqcap_2), \delta_2)$ , où les descriptions de  $\mathcal{D}_2$  sont des ensembles de prescriptions dont les médicaments sont représentés par des classes ATC. Pour définir  $\sqcap_2$ , il nous faut comparer des ensembles de classes d'une ontologie  $\mathcal{O}$ , et donc définir un opérateur intermédiaire  $\sqcap_{\mathcal{O}}$ , tel que, pour  $x$  et  $y$  deux ensembles de classes de  $\mathcal{O}$  :

$$x \sqcap_{\mathcal{O}} y = \max(\sqsubseteq, \{LCA(c_x, c_y) \mid (c_x, c_y) \in x \times y\})$$

où  $LCA(c_x, c_y)$  est l'ancêtre commun le plus spécifique de  $c_x$  et  $c_y$  dans  $\mathcal{O}$ , et  $\sqsubseteq$  est l'ordre partiel défini par la hiérarchie de classes de  $\mathcal{O}$ . Ainsi,  $\max(\sqsubseteq, S)$  est le sous-ensemble des classes de  $S$  les plus spécifiques, et  $x \sqcap_{\mathcal{O}} y$  est l'ensemble des ancêtres communs les plus spécifiques des classes dans  $x$  et  $y$ . A partir de  $\sqcap_{\mathcal{O}}$  nous définissons l'ordre partiel  $\leq_{\mathcal{O}}$  comparant deux ensembles de classes de  $\mathcal{O}$ , tel que  $x \leq_{\mathcal{O}} y$  signifie que  $y$  est un ensemble de classes plus spécifique que  $x$ . Par ailleurs,  $x \leq_{\mathcal{O}} y \Leftrightarrow x \sqcap_{\mathcal{O}} y = x$ . Nous définissons ensuite l'opérateur  $\sqcap_2$  tel que, pour deux descriptions  $X$  et  $Y$  de  $\mathcal{D}_2$  quelconques :

$$X \sqcap_2 Y = \max(\leq_{\mathcal{O}}, \{x \sqcap_{\mathcal{O}} y \mid (x, y) \in X \times Y\})$$

Cette structure de patrons permet la généralisation d'EIM déclenchés par des médicaments différents qui partagent un sous-groupe pharmacologique ou chimique. Par exemple :

$$\begin{aligned}\delta(P1) \sqcap_2 \delta(P4) &= \langle \{\{H02AB07\}, \{N02BE01\}\} \sqcap_2 \{\{H02AA03\}\}, \emptyset \rangle \\ &= \langle \{H02A\}, \emptyset \rangle\end{aligned}$$

Ce vecteur représente la généralisation la plus spécifique des descriptions des EIM des patients P1 et P4. Il exprime qu'au moins un médicament de la classe H02A (*corticosteroids for systemic use, plain*) est associé à un phénotype dans la classe ICD 580-629, et qu'aucun médicament n'est associé à la classe ICD 710-739.

### 3.3 Extension de la structure de patrons avec une ontologie de phénotypes

Nous définissons une troisième structure de patrons permettant d'utiliser à la fois ATC et ICD pour généraliser également les phénotypes. Ici, nous n'utilisons que les deux niveaux les plus spécifiques d'ICD, l'expérience précédente couvrant le niveau le plus général. La Table 5 illustre la représentation des données pour cette structure de patrons. Ici, les EIM sont représentés comme des vecteurs à deux dimensions  $\langle D_i, P_i \rangle$ , qui associe à un ensemble de médicaments  $D_i$  un ensemble de phénotypes  $P_i$ . Un patient est décrit par un ensemble d'EIM.

TABLE 5 – Exemple de descriptions de patients pour la structure de patrons  $(G, (\mathcal{D}_3, \sqcap_3), \delta_3)$ .

	Description
P1	$\langle \{\{H02AB07\}, \{ICD 599.8\}\}, \{\{N02BE01\}, \{ICD 599.9\}\} \rangle$
P2	$\langle \{\{H02AB07\}, \{ICD 599.9\}\}, \{\{H02AB07\}, \{ICD 719.4\}\} \rangle$
P3	$\langle \{\{H02AB07, N02BE01\}, \{ICD 599.9\}\}, \{\{N02BE01\}, \{ICD 719.4\}\} \rangle$

On définit ainsi la structure de patrons  $(G, (\mathcal{D}_3, \sqcap_3), \delta_3)$ , où une description de  $\mathcal{D}_3$  est un ensemble d'EIM représentés comme des vecteurs. On définit d'abord un opérateur  $\sqcap_{ADE}$  sur ces vecteurs, tel que, pour deux EIM  $v_x$  et  $v_y$  :

$$\begin{aligned}v_x \sqcap_{EIM} v_y &= \langle x_{ATC}, x_{ICD} \rangle \sqcap_{EIM} \langle y_{ATC}, y_{ICD} \rangle \\ &= \begin{cases} \langle x_{ATC} \sqcap_{\emptyset} y_{ATC}, x_{ICD} \sqcap_{\emptyset} y_{ICD} \rangle & \text{si les deux dimensions} \\ & \text{contiennent une classe non-racine,} \\ \langle \emptyset, \emptyset \rangle & \text{sinon.} \end{cases}\end{aligned}$$

L'opérateur  $\sqcap_{ADE}$  applique l'opérateur  $\sqcap_{\emptyset}$  sur chaque dimension du vecteur, en utilisant les ontologies ATC et ICD pour généraliser, respectivement, les médicaments et les phénotypes. Les deux dimensions du vecteur doivent contenir chacune au moins une classe non-racine de leur ontologies respectives. Si ce n'est pas le cas, on lui donne pour valeur  $\langle \emptyset, \emptyset \rangle$  afin de l'ignorer : cela permet de s'assurer que l'on ne généralise que sur des EIM avec au moins un médicament et un phénotype.

On définit l'opérateur  $\sqcap_3$  tel que, pour deux descriptions  $X$  et  $Y$  de  $\mathcal{D}_3$  quelconques :

$$X \sqcap_3 Y = \max(\leq_{ADE}, \{v_x \sqcap_{EIM} v_y \mid (v_x, v_y) \in X \times Y\})$$

Par rapport à  $\sqcap_2$ ,  $\sqcap_3$  ajoute une opération intermédiaire sur les EIM avec l'opérateur  $\sqcap_{EIM}$ , et permet d'utiliser  $\sqcap_{\emptyset}$  avec à la fois ATC et ICD.

## 4 Résultats et discussion

Les trois expériences décrites dans cet article produisent trois treillis de concepts, desquels on extrait des Règles d'Association (RA). Une RA est identifiée dans le treillis entre deux concepts parents dans le treillis, de descriptions  $\delta(l)$  et  $\delta(r)$ , avec  $\delta(r)$  plus spécifique que  $\delta(l)$ . Une RA comporte une partie gauche  $L = \delta(l)$  et une partie droite  $R = \delta(r) - \delta(l)$ , et est notée  $L \rightarrow R$ . Empiriquement, on n'extrait seulement les RA avec un support (nombre de patients vérifiant la règle) d'au moins 5 et une confiance (ratio de patients vérifiant  $L$  qui

vérifient également  $R$ ) d'au moins 0,75. La Table 6 présente quelques statistiques sur les treillis et RA obtenus pour les trois expériences.

TABLE 6 – Statistiques sur le processus d'extraction de RA pour les trois expériences, implémentées en Java.

Expérience	1	2	3
Taille du treillis (en millions de concepts)	1,9	2,3	2,5
RAs extraites (en millions)	5	7	9
RAs sélectionnées	772	1 907	913

L'extraction des RA produit un grand nombre de règles, parmi lesquelles les RA répondant à notre problème doivent être identifiées. Nous définissons donc un critère de sélection pour ces RA par la conjonction des deux conditions suivantes. (i) La partie droite  $R$  d'une RA contient au moins un EIM, noté  $\langle D_R, P_R \rangle$ , pour lequel il n'existe pas d'EIM dans la partie gauche  $L$ , noté  $\langle D_L, P_L \rangle$ , tel que  $D_L$  et  $D_R$  (resp.  $P_L$  et  $P_R$ ) sont comparable par  $\leq_0$ . Cette condition permet de s'assurer que la partie droite de la règle introduit des médicaments et phénotypes non liés à ceux de la partie gauche, *i.e.*, l'association entre les EIM n'est pas triviale. (ii) Puisque les patients du corpus sont diagnostiqués avec le LED, les règles ne doivent pas contenir de phénotype associé au LED (classe ICD 710 et ses descendants).

Nous présentons un exemple de RA obtenu dans la troisième expérience, avec un support de 10 et une confiance de 0,77 :

$$\{\{\{C08DB01\}, \{ICD 428.0\}\}\} \rightarrow \{\{\{A02B\}, \{ICD 427.31\}\}\}$$

Cette règle signifie que 77% des patients qui présentent *congestive heart failure* (ICD 428.0) après prescription de diltiazem (C08DB01), présentent également *atrial fibrillation* (ICD 427.31) après prescription d'un médicament pour l'ulcère gastro-duodénal et le reflux gastro-œsophagien (A02B). Cette RA est vérifiée pour 10 patients dans notre corpus. L'ensemble complet des RA sélectionnées est disponible en ligne à l'adresse suivante : <http://www.loria.fr/~gpersonne/eim-assoc/>.

Un grand nombre de RA peut être extrait de nos treillis de concepts. Nous avons automatiquement sélectionné un ensemble de ces RA en excluant les règles en dehors de la portée de notre étude. Nous devons maintenant classer ces RA par rapport à leur importance en termes de risques et de coûts des phénotypes présents dans leur partie droite. Une limite de cette approche est l'absence de relation temporelle entre les EIM. Nous n'avons pas considéré cet aspect, car l'ordre d'apparition d'EIM associés peut varier entre les patients. Cependant, cet ordre peut être vérifié dans les DME, puisque les concepts de patrons conservent les identifiants des patients.

Nous explorons dans cet article une approche fondée sur les structures de patrons pour extraire des EIM fréquemment associés depuis des DME. Cette approche est flexible et permet manipuler et fouiller les objets complexes que sont les EIM. De plus elle permet une généralisation sur les différents composants des EIM grâce aux ontologies médicales. La représentation des EIM pourrait être étendue en incluant d'autres composants, par exemple les cibles des médicaments annotées par des classes Gene Ontology.

## Références

- COULET, A., *ET AL.* (2013). Using pattern structures for analyzing ontology-based annotations of biomedical data. ICFCFA. Springer.
- GANTER, B., & KUZNETSOV, S. O. (2001). Pattern Structures and Their Projections. 9th International Conference on Conceptual Structures.
- GANTER, B., & WILLE, R. (1999). Formal Concept Analysis: Mathematical Foundations. Springer.
- KUHN, M., *ET AL.* (2016). The SIDER database of drugs and side effects. Nucleic Acids Research.
- LEPendu, P., *ET AL.* (2013). Pharmacovigilance Using Clinical Notes. Clinical Pharmacology & Therapeutics.
- LOWE, H. J., FERRIS, T. A., *ET AL.* (2009). STRIDE—An Integrated Standards-Based Translational Research Informatics Platform. AMIA Annual Symposium Proceedings.
- VASUDEVAN, A. R., & GINZLER, E. M. (2009). Established and Novel Treatments for Lupus. The Journal of Musculoskeletal Medicine.