

# Construction automatique d'ontologie à partir de bases de données relationnelles : application au médicament dans le domaine de la pharmacovigilance

Sonia Krivine, Jérôme Nobécourt, Lina Soualmia, Farid Cerbah, Catherine Duclos

► **To cite this version:**

Sonia Krivine, Jérôme Nobécourt, Lina Soualmia, Farid Cerbah, Catherine Duclos. Construction automatique d'ontologie à partir de bases de données relationnelles : application au médicament dans le domaine de la pharmacovigilance. IC 2009, May 2009, Tunisie. pp -, 2009. <hal-00377455>

**HAL Id: hal-00377455**

**<https://hal.archives-ouvertes.fr/hal-00377455>**

Submitted on 21 Apr 2009

**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.

# Construction automatique d'ontologie à partir de bases de données relationnelles: application au médicament dans le domaine de la pharmacovigilance

Sonia Krivine<sup>1</sup>, Jérôme Nobécourt<sup>1</sup>, Lina Soualmia<sup>1</sup>, Farid Cerbah<sup>2</sup>, Catherine Duclos<sup>1</sup>

<sup>1</sup>Laboratoire d'informatique médicale & Bioinformatique, Université Paris 13  
UFR de Santé Médecine et Biologie humaine, 74, rue Marcel Cachin 93017 Bobigny France  
sonia.krivine@free.fr, j.nobecourt@smbh.univ-paris13.fr, lina.soualmia@univ-paris13.fr,  
catherine.duclos@avc.aphp.fr

<sup>2</sup>Dassault aviation, Département des Etudes Scientifiques  
78, quai Marcel Dassault 92552 Saint-Cloud, France  
farid.cerbah@dassault-aviation.fr

**Résumé** : Afin de construire une ontologie des médicaments en OWL dans le cadre d'un projet sur la pharmacovigilance, nous envisageons la possibilité de réutiliser les connaissances contenues dans une base de données. L'étude des outils permettant cette transition montre qu'ils ne sont pas entièrement satisfaisants. En effet, ils ne considèrent pas chaque enregistrement comme étant un concept et de ce fait les hiérarchies implicites contenues dans la base ne sont pas restituées. Nous proposons une extension de l'outil RDBToOnto afin d'empêcher l'instanciation des enregistrements de certaines tables de la base et de reproduire les hiérarchies de concepts.

**Mots-clés** : Ontologies, bases de données relationnelles, pharmacovigilance, médicament.

## 1. Introduction

Les médicaments sont responsables d'effets indésirables. Ces effets peuvent être connus et observés pendant les phases expérimentales de leur développement, ou encore apparaître lors de leur commercialisation et de leur usage à grande échelle. Les agences sanitaires ont mis en place un processus de pharmacovigilance qui permet de surveiller à l'échelle régionale, nationale et supra-nationale l'apparition d'effets indésirables, et d'avoir un système d'alerte identifiant les médicaments dont le rapport bénéfice/risque est trop défavorable. Le processus vise à colliger, documenter et imputer tout cas d'effet indésirable lié à l'utilisation de médicaments déclaré par un médecin notificateur. Les informations sur ces cas sont enregistrées dans des bases de données de pharmacovigilance sur lesquelles des algorithmes de détection du signal identifient une présence statistiquement atypique de couples {médicament, effet

indésirable}(Hauben, 2003). Cela permet par la suite de mener des enquêtes spécifiques sur ces médicaments afin d'en réévaluer le bénéfice/risque.

Henegar et al (Henegar, 2006) ont montré que la détection du signal était améliorée si certains cas étaient regroupés sur la base de la proximité conceptuelle des effets indésirables. Pour cela, ils ont développé une ontologie des effets indésirables qui permet de reclasser les concepts de la nomenclature MedDRA<sup>1</sup> utilisée notamment pour coder les effets indésirables dans les cas de pharmacovigilance. Par exemple, si le médicament M1 est déclaré dans un cas comme donnant une élévation des transaminases (marqueur d'inflammation hépatique) et dans un autre cas comme donnant une hépatite (atteinte du foie par un processus inflammatoire), les deux cas peuvent être regroupés comme associant M1 et une inflammation du foie.

Faisant suite à ces travaux, il a été émis, dans le cadre du projet VigiTermes<sup>2</sup>, l'hypothèse que les médicaments pourraient bénéficier de modalités de regroupement. Par exemple, un cas décrivant un effet indésirable de type rhabdomyolyse après la prise de Zocor<sup>3</sup> et un autre décrivant le même effet indésirable après la prise de Lipanthyl<sup>4</sup> seraient regroupés car ils impliquent tous les deux des médicaments hypolipémiants. Cela permettrait ensuite de déduire que l'effet indésirable de rhabdomyolyse est lié à la classe des molécules hypolipémiantes.

Pour tester cette hypothèse, il faut disposer d'une ressource ontologique décrivant les médicaments à l'aide des propriétés susceptibles d'être à l'origine d'effets indésirables pour permettre un processus de classification des médicaments selon leur risque iatrogène<sup>5</sup>. En effet, un certain nombre de propriétés du médicament peuvent être identifiées comme ayant un lien potentiel avec la survenue d'effets indésirables (par exemple le fait qu'un médicament est fortement métabolisé par une enzyme hépatique peut conduire à un surdosage et donc à un effet indésirable chez un patient qui est déficitaire en cette enzyme, ou chez qui cette enzyme est inhibée par un autre médicament).

Plusieurs ontologies traitant du médicament sont disponibles comme Drug Ontology (Solomon, 1999), National Drug File – Reference Terminology (NDF-RT) (Chute, 2003), SNOMED-CT (Kim, 2001). Leurs limites sont soit un point de vue classificatoire orienté vers la prescription (description de la composition en ingrédients, appartenance à des classes pharmaco-thérapeutiques), ou lorsque des propriétés supplémentaires sont décrites, un abandon dans leur maintenance conduisant à leur obsolescence et à la non exhaustivité de leur contenu.

Les propriétés du médicament sont par ailleurs complètement décrites dans les banques de données commerciales dans un but de diffusion de l'information sur le médicament ou d'opérationnalisation dans les systèmes d'aide à la prescription. Ces

---

<sup>1</sup> Medical Dictionary for Regulatory Activities (<http://meddramsso.com/MSSOWeb/index.htm>)

<sup>2</sup> VigiTermes (ANR-07-TECSAN-026-04)

<sup>3</sup> Zocor est une simvastatine, molécule hypolipémiante de la famille des statines

<sup>4</sup> Lipanthyl est un fénofibrate, molécule hypolipémiante de la famille des fibrates

<sup>5</sup> Iatrogène: "se dit d'un trouble ou d'une maladie provoquée par un acte médical ou par les médicaments, même en l'absence d'erreur du médecin" (Petit Larousse)

banques ont l'intérêt d'avoir un contenu exhaustif pour décrire des propriétés liées à la iatrogénie, extrêmement structuré et maintenu à jour.

Notre objectif, dans le cadre du projet VigiTermes, est de construire de façon la plus automatisée possible une ressource ontologique permettant de classer les médicaments selon leur risque iatrogène. Il semble intéressant de partir d'une base de données structurée sur le médicament et d'outiller sa transition vers un formalisme ontologique (Gomes-Perez, 2004).

Nous présentons dans cet article tout d'abord le périmètre des propriétés liées au risque iatrogène du médicament et la sélection d'un sous-ensemble de tables de la base de données sur le médicament Thésorimed<sup>6</sup>. Nous décrivons ensuite les outils existants permettant de faire la transition entre le formalisme base de données relationnelle et le formalisme OWL et les utilisons sur l'extrait de la base Thésorimed. Confrontant les résultats à nos attentes, nous proposons une méthode pour enrichir ces outils de l'aspect important de la conservation des taxonomies. L'analyse des résultats de ce travail exploratoire est abordée en discussion.

## **2. Description du domaine**

### **2.1. Modèle de la iatrogénie du médicament**

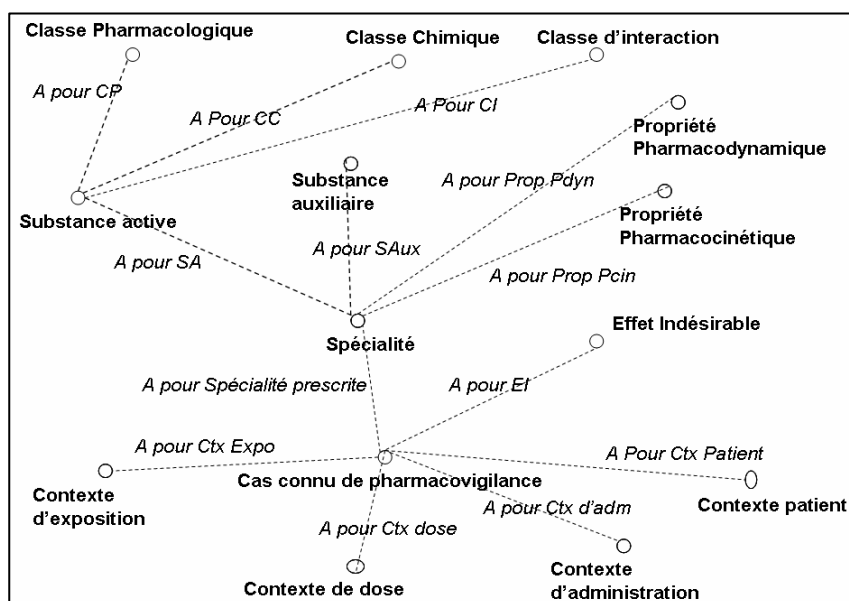
Les effets indésirables peuvent se comprendre grâce à certaines propriétés qu'ont les médicaments. Un effet indésirable peut (a) être spécifique d'une substance active (*l'amoxicilline peut donner une dyschromie de l'email dentaire*) ou d'une substance auxiliaire entrant dans la composition du médicament mais n'ayant pas de propriétés thérapeutiques (*l'aspartam contenu dans une suspension buvable d'amoxicilline peut être dangereux chez les sujets atteints de phénylcétonurie*); (b) être un effet de classe se produisant avec tous les médicaments appartenant à une classe chimique ou pharmacologique (*les amidinopénicillines peuvent donner des diarrhées ; les antiulcéreux inhibiteurs de la pompe à proton peuvent donner des douleurs abdominales*); (c) être la manifestation d'une interaction générique avec un organe ou d'une interaction expliquée au niveau moléculaire avec un site effecteur recherché ou secondaire (*le captopril dont le mécanisme d'action est d'inhiber de l'enzyme de conversion rénine angiotensine pour diminuer la pression artérielle risque d'entraîner une hypotension artérielle directement liée à son mécanisme d'action mais aussi une toux dont le mécanisme de survenue est inconnu*) ; (d) survenir à dose thérapeutique ; (e) être la manifestation d'un surdosage se traduisant par une majoration de l'ensemble des effets du médicament et pouvant conduire à une toxicité du médicament.

L'effet indésirable qui survient à dose thérapeutique peut apparaître dans certains contextes (a) patients (physiologique, pathologique, génétique ou allergique) ; (b) de doses (faible, forte) ; (c) d'administration (forme, voie, débit) ; (d) d'exposition

---

<sup>6</sup> Thésorimed est une base de données sur le médicament développée par le GIE SIPS (Système d'Information sur les Produits de Santé), (<http://www.giesips.org/>)

(administration prolongée). Le surdosage peut être lié à (a) une posologie erronée (dose par prise, dose par jour, fréquence d'administration, durée) ;(b) une altération des paramètres pharmacocinétiques<sup>7</sup> en raison d'interactions médicamenteuses, d'un contexte patient particulier ; (c) une altération de la pharmacodynamie<sup>8</sup> en raison d'une interaction médicamenteuse, d'un contexte patient particulier. La figure 1 illustre le modèle de la iatrogénie du médicament.



**Fig. 1** - Modèle de la iatrogénie du médicament pour documenter un cas de pharmacovigilance (les lignes en pointillés représentent des relations entre concepts).

## 2.2 – Etude de Thésorimed et constitution de la base de données sur la iatrogénie du médicament

La base de données sur le médicament Thésorimed contient les informations illustrant les propriétés précédemment décrites (figure 1). Ces propriétés sont représentées soit en utilisant une terminologie contrôlée (termes d'indexation de Thésorimed), soit sous forme de texte. Nous nous intéresserons ici aux éléments décrits par une terminologie contrôlée (figure 2), à savoir : (i) les substances actives, auxiliaires, classes chimiques, classes pharmacologiques donnant des effets indésirables à dose thérapeutique ou en cas de surdosage, (ii) les hiérarchies de classes chimiques, pharmacologiques et d'effets indésirables.

<sup>7</sup> La pharmacocinétique étudie le devenir d'un médicament dans l'organisme

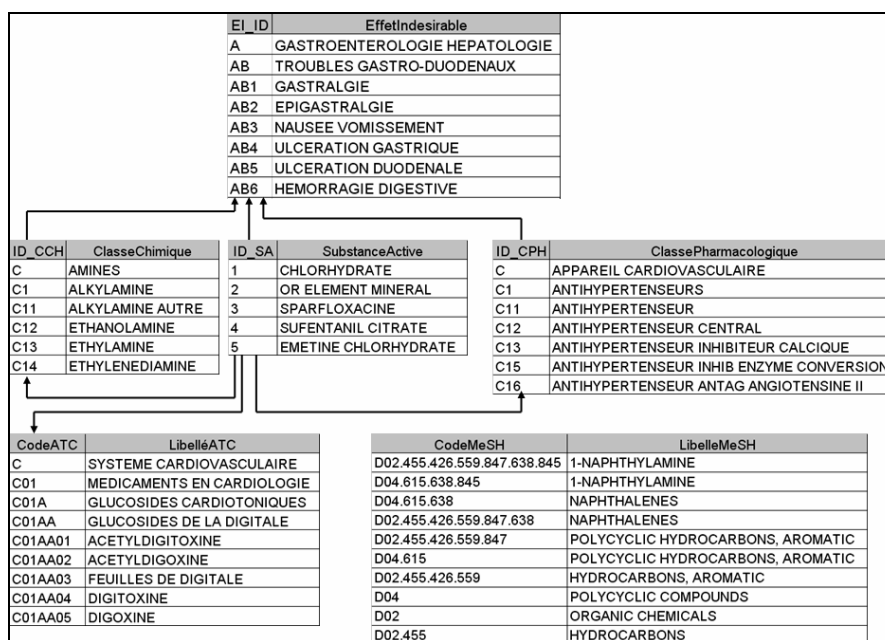
<sup>8</sup> La pharmacodynamie décrit ce que le médicament fait à l'organisme

A cela il nous semble intéressant d'ajouter des hiérarchies comme le MeSH<sup>9</sup> pour enrichir la description chimique ou comme l'ATC<sup>10</sup> pour la description pharmacothérapeutique des médicaments.

Les hiérarchies dans ces tables sont exprimées de trois façons différentes. La hiérarchie de type ATC est représentée à l'aide d'un code alphanumérique sans séparateur qui s'étend de père en fils et pour lequel un fils a un seul père. Chaque caractère du code représente un niveau de la hiérarchie. Ce codage est retrouvé dans Thésorimed pour les classes chimiques, pharmacologiques et d'effets indésirables.

La hiérarchie de type MeSH reprend le principe d'un code alphanumérique avec séparateur qui s'étend de père en fils. Elle a la caractéristique particulière qu'un fils peut avoir plusieurs pères selon le principe de l'héritage multiple.

Dans la hiérarchie de type Thésorimed le lien de subsomption est exprimé grâce la relation entre la table des substances et celle des classes chimiques ou des classes pharmacologiques.



**Fig. 2** - Tables et champs retenus dans la banque Thésorimed pour représenter les éléments initiaux du modèle de la iatrogénie des médicaments et présentation d'extraits de leur contenu, les tables reliées à la table effet indésirable servent à la construction de cas connus de pharmacovigilance

<sup>9</sup> Medical Subject Headings (<http://www.nlm.nih.gov/mesh/>)

<sup>10</sup> Classification Anatomique, Thérapeutique et Chimique (<http://www.whocc.no/atcddd/>)

### 3. Outils de transition entre bases de données et OWL

DataMaster<sup>11</sup>, KAON2<sup>12</sup>, et RDBToOnto<sup>13</sup> sont trois outils permettant de faire la transition entre les formalismes base de données et OWL. Pour chacun de ces outils, nous avons étudié comment se fait le reverse engineering pour créer un concept, une relation, un héritage, une propriété, une instance.

#### 3.1 Description des outils

##### 3.1.1 DataMaster

DataMaster est un plug-in de Protégé spécialisé dans l'import des structures et des données de bases de données relationnelles (Nyulas, 2007) qui propose une méthode d'import des tables de la base de données relationnelle comme des concepts OWL.

Avec DataMaster, le paradigme relationnel est conservé puisque les données extraites de la base sont simplement inscrites dans l'ontologie générique *Relational.OWL* [de Laborda *et al.* 2005]. Cet outil n'est pas dimensionné pour effectuer une restructuration profonde du modèle et des données sources, il permet néanmoins de migrer les données vers une description dans une ontologie.

En utilisant DataMaster avec notre base de données, chaque table a été convertie en un concept et chaque ligne de la table a été convertie en une instance du concept correspondant. Les valeurs des attributs ont été instanciées avec les valeurs des champs correspondants de la table.

##### 3.2.2 KAON2

KAON2 est une plateforme de construction d'ontologies développée à l'université de Karlsruhe.

Cette plateforme est en fait équipée d'une fonctionnalité de mise en correspondance entre bases de données et ontologies, dans la lignée des outils et formalismes de mapping, tels que D2RQ [Bizer 2003] et R2O [Barrasa *et al.* 2004]. Dans cette perspective, il ne s'agit pas d'automatiser la construction du modèle de classes et de propriétés. L'objectif visé est d'offrir des moyens déclaratifs pour décrire des procédés d'instanciation à partir de bases relationnelles d'ontologies prédéfinies manuellement. KAON2 permet de fournir une « vue » sous forme d'ontologie (ce que ses concepteurs appellent une « ontologie virtuelle »), alimentée à la volée par des instances extraites d'une base de données.

La documentation très succincte fournie sur le site Web de la plateforme<sup>14</sup> et dans les exemples présents dans la distribution du logiciel laisse prévoir un type de résultat très proche de celui de DataMaster.

---

<sup>11</sup> <http://protegewiki.stanford.edu/index.php/DataMaster>

<sup>12</sup> <http://kaon2.semanticweb.org/>

<sup>13</sup> <http://www.tao-project.eu/researchanddevelopment/demosanddownloads/RDBToOnto.html>

<sup>14</sup> <http://kaon2.semanticweb.org/#documentation>

Les procédés de mapping proposés présupposent l'utilisation de structures ontologiques simples, sans axiomes, quasi-calquées sur les schémas relationnels sources.

### **3.2.3 RDBToOnto**

RDBToOnto [Cerbah 2008a] est un outil dédié à la conversion de bases de données en ontologies, développé dans le cadre du projet européen TAO (Transitioning Applications to Ontologies).<sup>11</sup>

RDBToOnto permet à l'utilisateur un paramétrage du projet de conversion à savoir :

- la définition interactive de contraintes locales, attachées aux tables d'entrée, pour orienter le processus de transformation. En particulier, ces contraintes permettent de spécifier des motifs de nommage des classes et des instances (en combinant des valeurs d'attributs pour former les noms) ou encore d'exclure des tables et surtout des colonnes lors de la conversion,
- Le choix d'un convertisseur (implémentation d'une méthode de transformation), parmi ceux prédéfinis, ou le recours à une nouvelle méthode obtenue en spécialisant un convertisseur déjà intégré dans la plateforme.

Le développement de RDBToOnto est orienté vers la reconnaissance de structures de catégorisation en analysant conjointement le schéma de la base et les données stockées. Ainsi, le convertisseur RTAXON, central dans cet outil, implémente une méthode générique de reconnaissance des attributs de catégorisation, se basant d'une part sur le nom de l'attribut, d'autre part sur la redondance dans les extensions des attributs à l'aide d'une méthode basée sur l'entropie (Cerbah, 2008b).

Le paramétrage de l'outil en « Catégorisation forcée » permet déjà d'obtenir un second niveau de profondeur hiérarchique, qu'aucun des autres outils examinés ne propose et témoigne d'une orientation dans la conception de l'outil vers l'extraction de relations de subsomption, particulièrement intéressante pour notre projet.

Chaque table est convertie en une classe et chaque ligne de la table donne lieu à la création d'une sous-classe.

## **3.2 Limites des outils existants**

Notre étude a révélé que les outils existants partagent certaines limitations majeures.

**Une conception orientée vers le peuplement d'ontologies** c'est-à-dire l'instanciation de concepts représentés par les tables de la base de données plutôt que vers l'apprentissage d'une structure conceptuelle (avec hiérarchies multi niveaux, propriétés, héritage multiple...etc.) à partir de bases de données. Ceci à l'exception de RDBToOnto, dont la méthode de conversion RTAXON est orientée vers l'extraction de relations d'héritage contenues dans certains champs de la base de données.

**La quasi-absence de paramétrage de l'importation par l'utilisateur** : à l'exception de RDBToOnto, ces outils s'intègrent au sein d'un logiciel (ou d'une



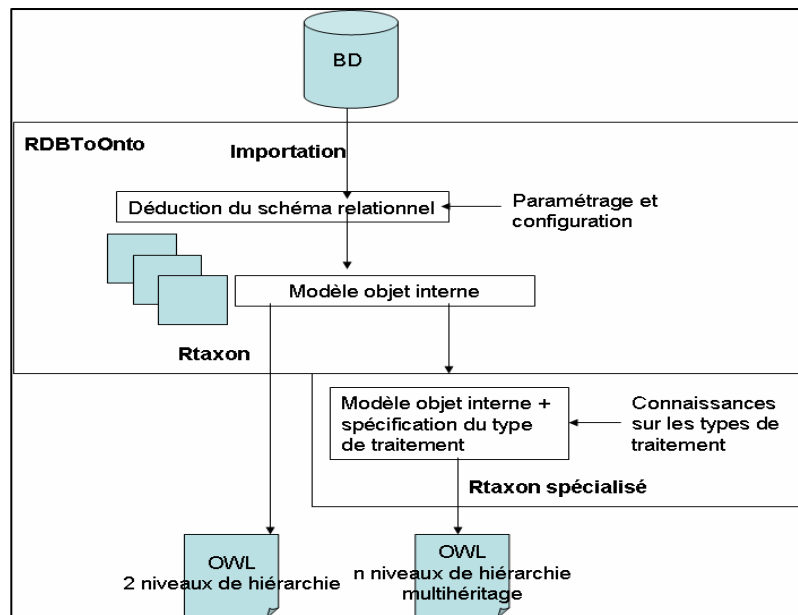
suite logicielle) d'édition d'ontologies, sous la forme d'une fonctionnalité qui laisse peu de possibilités de paramétrage de la conversion à l'utilisateur comme par exemple l'exclusion de certaines tables, des traitements spécifiques à certaines tables, le choix de conventions sur le nommage des concepts et propriétés extraits.

**L'héritage multiple est exclu.**

**L'extraction d'une structure ontologique « plate » :** la structure obtenue reste très proche du schéma de la base de données. Toutes les tables subissent un traitement identique et rudimentaire : une table de la base source est convertie en un concept et chaque ligne de cette table est convertie en instance de ce même concept, les valeurs des colonnes étant reportées sous forme de valeurs d'attributs. Les structures hiérarchiques extraites sont de ce fait peu profondes : deux niveaux de profondeur maximum obtenus avec l'option de catégorisation de RDBToOnto.

**4. Enrichissement de RDBToOnto**

Nous avons choisi de procéder à une adaptation de RDBToOnto en spécialisant le convertisseur RTAXON de manière à en contourner les limitations exposées précédemment (figure 3).



**Fig. 3 – Méthode d'enrichissement de RDBToOnto**

Le développement de quelques composants Java supplémentaires est requis afin que cet outil :

- génère des hiérarchies de concepts d'une profondeur supérieure à 2 dans le cas de hiérarchies de type ATC

- autorise le multihéritage dans le cas des tables de type MeSH,
- prévoit l'import spécifique de tables contenant des relations entre concepts, par exemple « SubstanceActive (aPourClasseChimique) Classe\_Chimique ».

#### 4.1 – Traitement de type ATC pour la construction des taxonomies de profondeur fixe ou variable, sans multihéritage

L'utilisateur identifie une table comme étant hiérarchique et l'attribut qui contient l'information sur les niveaux hiérarchiques. Cette information doit être décodée pour en extraire la taxinomie. Dans notre cas, l'information sur ces niveaux hiérarchiques est représentée par un code alphanumérique qui s'étend de père en fils. La hiérarchie peut se déduire automatiquement par parcours récursif du code.

Dans ces tables, les libellés ne représentent pas le terme préféré pour nommer un concept (par exemple le libellé « associations » est utilisé quelque que soit le type d'association (plusieurs antibiotiques ou plusieurs antidiabétiques)), le sens de l'association est explicité lorsque la hiérarchie est déployée (on déduit que « association » a le sens « association d'antibiotiques » parce qu'elle est associée au code J01CA20, qui est dans la hiérarchie des antibiotiques). Pour nommer un concept, nous avons donc choisi d'associer le code et le libellé du code (figure 4).

CodeATC	LibelléATC
C	SYSTEME CARDIOVASCULAIRE
C01	MEDICAMENTS EN CARDIOLOGIE
C01A	GLUCOSIDES CARDIOTONIQUES
C01AA	GLUCOSIDES DE LA DIGITALE
C01AA01	ACETYLDIGITOXINE
C01AA02	ACETYLDIGOXINE
C01AA03	FEUILLES DE DIGITALE
C01AA04	DIGITOXINE
C01AA05	DIGOXINE

```

graph TD
    C[C-SYSTEME_CARDIOVASCULAIRE] --> C01[C01-MEDICAMENTS_EN_CARDIOLOGIE]
    C01 --> C01A[C01A-GLUCOSIDES_CARDIOTONIQUES]
    C01A --> C01AA[C01AA-GLUCOSIDES_DE_LA_DIGITALE]
    C01AA --> C01AA01[C01AA01-ACETYLDIGITOXINE]
    C01AA --> C01AA02[C01AA02-ACETYLDIGOXINE]
    C01AA --> C01AA03[C01AA03-FEUILLES_DE_DIGITALE]
    C01AA --> C01AA04[C01AA04-DIGITOXINE]
    C01AA --> C01AA05[C01AA05-DIGOXINE]
    
```

Fig. 4 – Exemple de transformation de type ATC à partir de la table de « classification ATC » et visualisation de la taxinomie résultante dans Protégé

#### 4.2 Traitement de type MeSH pour la construction des taxonomies avec multihéritage

De la même manière qu'en 4.1, la table doit être identifiée comme étant hiérarchique, un attribut doit contenir une information sur un des niveaux hiérarchiques (code) et un autre attribut doit comporter le terme. Il existe donc plusieurs lignes ayant le même terme mais des codes différents. Le terme est pris comme label du concept, ses pères sont calculés en fonction de chacun de ses codes (figure 5).

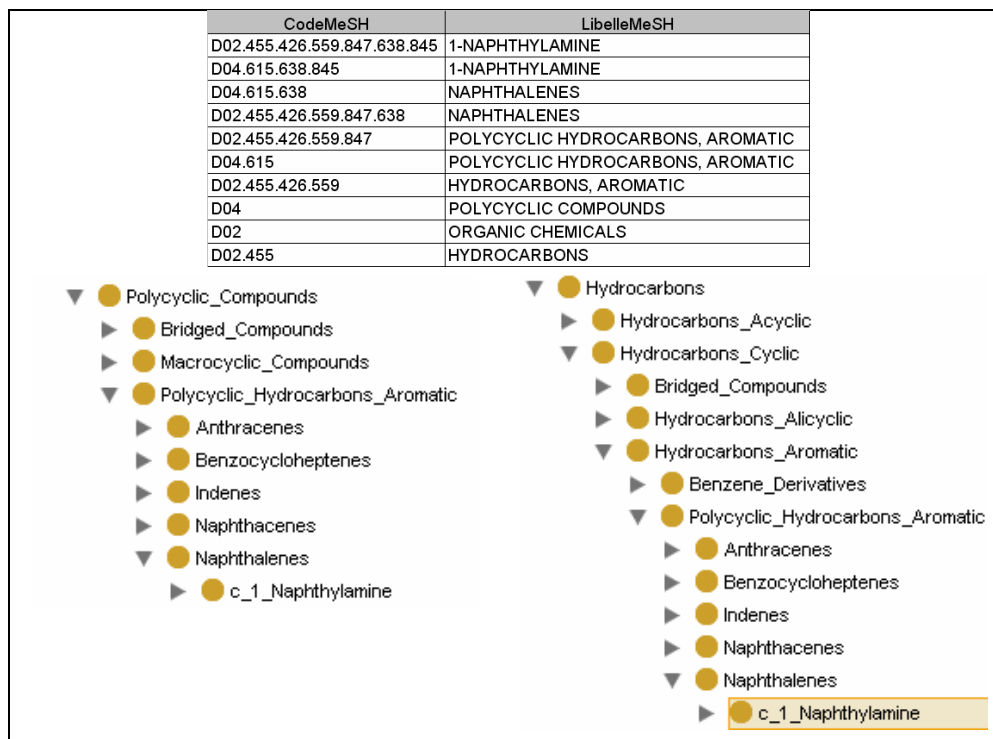


Fig. 5 – Exemple de transformation de type MeSH à partir de la table de « classification MeSH » et visualisation du multihéritage dans Protégé

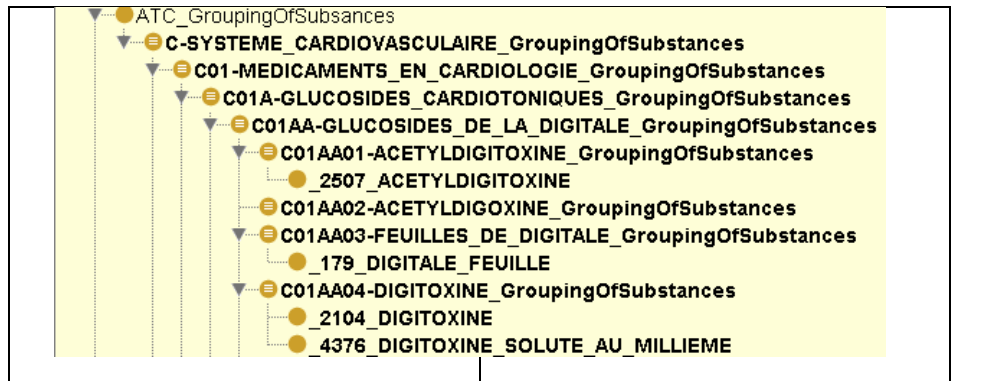
#### 4.4 Traitement de type Thesorimed pour la construction des relations et des hiérarchies contenues dans des tables de relation binaire entre substance et classification

L'utilisateur identifie trois tables : la table de classification, la table de substance et la table de jointure entre substance et classification. A partir de la table de classification le traitement décrit en 4.1 est réalisé, puis on décrit des concepts de regroupement de substances. Ces concepts sont définis en utilisant un rôle dont le domaine est la substance et le co-domaine le code de classification.

Pour chaque concept de substance, nous définissons une propriété « aPourClassification » grâce à la table de jointure.

A l'issue de ces descriptions, nous utilisons le classifieur Fact++ pour reclasser les substances dans les regroupements de substances.

Nous avons pour l'instant réalisée cette classification sur 5234 concepts (concepts ATC et concepts substances), et 60 concepts de regroupements. La figure 6 illustre ce traitement.



**Fig. 6** – Exemple de transformation de type Thesorimed à partir des tables de « classification ATC », « Substances » et « SubstanceApourATC » et visualisation de la hiérarchie inférée par Fact++ dans Protégé

## 5. Discussion et conclusion

Les outils disponibles pour faire la transition entre les bases de données relationnelles et OWL supportent principalement le postulat que les enregistrements trouvés dans les bases sont des instances du concept constitué par la table. Dans le cas des bases de données sur le médicament, il s'avère qu'un certain nombre de tables contiennent des termes normalisés servant à la description des propriétés du médicament et qu'ils peuvent être considérés non comme des instances de concepts mais comme des concepts. Ces concepts entretiennent en particulier des relations de subsumption qu'il faut pouvoir rendre dans l'ontologie.

La création de notre ontologie est guidée à la fois par la connaissance sur les données mais également par l'usage envisagé pour cette ontologie. Il existe d'autres approches de reverse engineering qui intègrent l'usage de la base de données pour augmenter la spécificité des traitements (Benslimane 2007).

Nous avons distingué trois types de traitement à appliquer à notre base de données sur le médicament pour : (a) reproduire une hiérarchie contenue dans un code par parcours récursif du code, (b) créer des liens de hiérarchie multiple, (c) créer des associations décrites par la relation entre deux entités.

Cela nous a conduit, dans une première approche exploratoire, à spécialiser l'outil RDBToOnto afin de proposer à l'utilisateur un nouveau convertisseur lui permettant d'indiquer des propriétés des tables pour réaliser des traitements spécifiques aux différents types de hiérarchies.

Cette spécialisation de l'outil remet en cause le principe de généralité de RDBToOnto et nous amène à reconsidérer le problème général de la construction de nouveaux points de vue sur l'utilisation des connaissances sur le médicament tout en réutilisant le contenu de bases de données commerciales sur le médicament. Cela

devrait nous orienter vers le développement d'outils dédiés spécifiquement à la réutilisation d'une base médicaments donnée.

Le fait de s'adosser à une base de données mise à jour par des éditeurs nous permet d'envisager une maintenance plus aisée de l'ontologie grâce à des traitements les plus automatisés possibles réalisés sur cette base.

## 6. Références

- BARRASA J, CORCHO O, GÓMEZ-PÉREZ A. (2004) "R2O, an Extensible and Semantically Based Database-to-Ontology Mapping Language". *Second Workshop on Semantic Web and Databases (SWDB2004)*. Toronto, Canada.
- BENSLIMANE S.M, BENSLIMANE D, MALKI M, AMGHAR Y, GARGOURI F.(2006) Construction d'une ontologie à partir d'une base de données relationnelle: approche dirigée par l'analyse des formulaires HTML. *INFORSID'06, Hammamet, Tunisie* pp. 991-1010.
- BIZER, C.(2003) "D2R MAP - A Database to RDF Mapping Language", WWW03, Budapest.
- CERBAH, F. (2008a) Learning Highly Structured Semantic Repositories from Relational Databases - RDBtoOnto Tool, in *Proceedings of the 5th European Semantic Web Conference (ESWC 2008)*, Tenerife, Spain, June, 2008.
- CERBAH, F. (2008b) Mining the Content of Relational Databases to Learn Ontologies with deeper Taxonomies. *Proceeding of IEEE/WIC/ACM International Joint Conference on Web Intelligence (WI'08) and Intelligent Agent Technology (IAT'08)*, Sydney, Australia, 9-12.
- CHUTE ET AL (2003) Integrating pharmacokinetics knowledge into a drug ontology: as an extension to support pharmacogenomics. *AMIA Annu Symp Proc.* 170-4.
- DE LABORDA, C.P., CONRAD, S. (2005) Relational.OWL: a data and schema representation format based on OWL. *In: APCCM '05: Proc. of the 2nd Asia-Pacific conference on Conceptual modelling, Darlinghurst*, Australian Computer Society, Inc.
- GOMEZ-PEREZ A., FERNANDEZ-LOPEZ M., CORCHO O. (2004) *Ontological engineering, with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web*, Springer-Verlag London.
- HAUBEN M, ZHOU X.(2003) Quantitative methods in pharmacovigilance: focus on signal detection. *Drug Saf.* ;26(3):159-86.
- HENEGAR C, BOUSQUET C, LILLO-LE LOUËT A, DEGOULET P, JAULENT MC (2006). Building an ontology of adverse drug reactions for automated signal generation in pharmacovigilance. *Comput Biol Med.* 36(7-8):748-67.
- KIM JM, FROSDICK P (2001).Description of a drug hierarchy in a concept-based reference terminology. *Proc AMIA Symp*:314-8.
- NYULAS C. O'CONNOR, M. TU, S. (2007), DataMaster – a Plug-in for Importing Schemas and Data from Relational Databases into Protégé, ([http://protege.stanford.edu/conference/2007/presentations/10.01\\_Nyulas.pdf](http://protege.stanford.edu/conference/2007/presentations/10.01_Nyulas.pdf)).
- SOLOMON ET AL (1999).A reference terminology for drugs. *Proc AMIA Symp.* 152-5.