



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



ELSEVIER  
MASSON

Disponible en ligne sur  
 ScienceDirect  
 www.sciencedirect.com

Elsevier Masson France  
  
 www.em-consulte.com

IRBM

IRBM 29 (2008) 350–356

Article original

## Recherche d'information multiterminologique : application à un portail d'information sur le médicament en Europe

### Multiterminology information retrieval: Application to a drug information portal in Europe

C. Letord<sup>a</sup>, S. Sakji<sup>a</sup>, S. Pereira<sup>a,b</sup>, B. Dahamna<sup>a</sup>, I. Kergoulay<sup>a</sup>, S. Darmoni<sup>a,\*</sup>

<sup>a</sup> CISMef, CHU de Rouen, GCSIS, LITIS EA 4108, institut de recherche biomédicale, université de Rouen, Rouen, France

<sup>b</sup> Vidal, Issy-les-Moulineaux, France

Reçu le 12 septembre 2008 ; accepté le 1<sup>er</sup> octobre 2008

Disponible sur Internet le 20 novembre 2008

#### Résumé

**Objectifs.** – L'objectif de ce travail est la création d'un portail d'information bilingue (français/anglais) sur le médicament, dans un contexte multiterminologique.

**Matériel et méthodes.** – La construction du portail d'information sur le médicament (PIM) s'est appuyée sur le portail CISMef en adaptant certaines fonctionnalités spécifiques : création du concept de substance, regroupant des descripteurs MeSH et des *supplementary concepts* traitant du médicament. Nous avons créé également une arborescence spécifique de type de ressources, pour traiter du contenant de l'information (par exemple : notice médicamenteuse). Enfin, nous avons créé un métaterme spécifique « médicament » englobant sémantiquement tous les concepts précédents.

**Résultats.** – PIM est le résultat d'une collaboration entre l'équipe CISMef et la société privée Vidal, spécialiste de l'information sur les médicaments. L'URL de ce site est : <http://doccismef.chu-rouen.fr/servlets/PIM>. Les terminologies et classifications interrogeables par le PIM sont : le MeSH, l'ATC, les codes CAS et codes EINCS, les codes français CIS, CIP et UCD, ainsi que les noms commerciaux et les dénominations communes internationales des médicaments.

**Discussion.** – En 2008, la National Library of Medicine (NLM) a mis en place le « Drug Information Portal ». Ce portail permet d'accéder aux informations concernant plus de 12 000 médicaments. La recherche peut s'y effectuer à partir du nom générique ou du nom commercial. À notre connaissance, la recherche par code CAS ou ATC n'est pas possible dans le US Drug Information Portal.

**Conclusion.** – La réalisation du PIM permet de faciliter la recherche d'information de l'utilisateur en se restreignant au domaine médicamenteux, mais en utilisant un contexte multiterminologique.

© 2008 Publié par Elsevier Masson SAS.

#### Abstract

**Objectives.** – The objective of this work is the creation of a bilingual drug information portal (French/English), in a multiterminological context.

**Material and methods.** – The construction of the drug information portal (PIM) was leaned on the CISMef portal by adapting certain specific functionalities: creation of the concept “substance”, which gather the MeSH descriptors and the Supplementary Concepts referring to medicinal drugs. We also created a specific tree structure of resource types, to describe the container of information (for example: medicinal note). Lastly, we created a specific meta-term “medicine” including semantically all the preceding concepts.

**Results.** – PIM is the result of a collaboration between the CISMef team and the private company VIDAL, specialized in drug information. The URL of this site is: <http://doccismef.chu-rouen.fr/servlets/PIM>. The terminologies and classifications that may be tested by PIM are: MeSH, the ATC, the CAS codes and the EINECS/ ELINCS codes, the French codes CIS, CIP and UCD, as well as trade names and the international generic names of the drugs.

\* Auteur correspondant. Rouen University Hospital Research, 1, rue de Germont, 76031 Rouen cedex, France.

Adresse e-mail : [stefan.darmoni@chu-rouen.fr](mailto:stefan.darmoni@chu-rouen.fr) (S. Darmoni).

*Discussion.* – In 2008, the National Library of Medicine (NLM), set up the “Drug Information Portal”. This portal gives access to information concerning more than 12 000 drugs. Research can be carried out by the generic name or by trade name. To our knowledge, a research by CAS or ATC codes is not possible in the US Drug Information Portal.

*Conclusion.* – The realization of the PIM facilitates the search of information by the user while being restricted with the medicinal field, but by using a multiterminological context.

© 2008 Publié par Elsevier Masson SAS.

*Mots clés :* Services d’information sur le médicament ; Internet ; Europe ; stockage et recherche information ; Terminologie comme sujet

*Keywords:* Drug information services; Internet; Europe; Storage and research information; Terminology like subject

## 1. Introduction

Quelle que soit leur expérience du Web et leur compétence en recherche d’information, les utilisateurs finaux rencontrent des difficultés à rechercher une information de santé sur l’Internet [1]. Vu le développement croissant de l’information, le recours aux portails d’information représente une bonne solution pour faciliter l’accès à une information disparate, qu’elle soit localisée sur des sites Web statiques détectables facilement par les moteurs de recherches généralistes, comme Google ou aux sites Web dynamiques et, en particulier, les bases de données disposant d’une information structurée ; ces sites dynamiques sont beaucoup plus difficilement repérables par les outils de recherche : c’est le « Web invisible ». Un portail est un site Web, cataloguant les principales ressources disponibles dans un domaine donné, ici, le médicament. Un portail inclut le plus souvent un moteur de recherche. Un portail de santé a plusieurs spécificités : il doit respecter des critères de qualités précis ; dans le domaine de la santé, la référence des grilles de critères de qualité est Health on the Net (HON) [2]. Cette grille a été choisie en 2007 en France par la Haute Autorité de santé pour certifier les sites e-santé de ce pays, et ce, sur un mode volontaire.

Dans le cadre du projet européen Patient Safety Through Intelligent Procedures in Medication (PSIP) [3], nous nous intéressons à mettre au point un portail d’information sur le médicament (PIM) permettant d’améliorer la recherche d’information de l’utilisateur sur ce sujet précis. PIM est un portail d’information bilingue (français/anglais) sur le médicament, dans un contexte multiterminologique. PIM se restreint par un choix éditorial de ces auteurs aux informations concernant le médicament et émanant d’institutions ou de sociétés savantes. Aucune information commerciale ne sera recensé dans PIM. PIM est le résultat d’une collaboration entre l’équipe CISMef et la société privée Vidal, spécialiste de l’information sur les médicaments. En revanche, PIM pourra faire des liens contextuels vers des sites commerciaux, comme Vidal Recos par exemple.

## 2. Matériel et méthodes

### 2.1. CISMef et sa terminologie

CISMef (URL : <http://www.chu-rouen.fr/cismef> ou <http://www.cismef.org>) est l’acronyme de Catalogue et Index des Sites Médicaux Francophones sur l’Internet [4]. Il s’agit d’un portail de santé [5] qui a été conçu pour cataloguer et indexer les sources d’information institutionnelle de santé

françaises les plus importantes ( $n \approx 54\,000$ ) et ce, afin de permettre une recherche plus pertinente pour les professionnels de santé, les étudiants mais aussi les patients, leur familles et, d’une façon encore plus large, le cyber-citoyen. CISMef adhère aux principes de qualité de l’information de santé sur l’Internet défini par la HON Foundation depuis dix ans maintenant [2].

CISMef utilise deux outils standards pour organiser l’information : le thésaurus MeSH (utilisé généralement pour indexer les articles scientifiques de la base de données Medline) et un ensemble de métadonnées extraites du Dublin Core [6] (URL : <http://www.dublincore.org>). Les métadonnées se réfèrent aux informations descriptives des ressources Web et dont les plus importantes sont : le titre, l’identifiant, la date, le contenu, le descripteur et le type de ressource. Pour plus de précision, huit métadonnées spécifiques à CISMef ont été ajoutées, tels que pays, institution, niveau de preuve pour les données actuelles de la science.

La terminologie CISMef encapsule la version française du thésaurus MeSH dans la mesure où, d’une part, elle représente une extension des concepts déjà existants dans le MeSH et, d’autre part, elle emploie de nouveaux concepts. Dans le but de généraliser le MeSH aux ressources de santé sur l’Internet, des améliorations ont été réalisées depuis l’année 2000. En plus des descripteurs MeSH (termes qui permettent l’indexation des ressources) et des qualificatifs MeSH (qui permettent de préciser le sens d’un descripteur et d’en souligner un aspect particulier), deux nouveaux concepts ont été créés par l’équipe CISMef : des types de ressources qui sont une extension des types de publications de Medline et des métatermes (MT). La paire (descripteur/qualificatif) décrit le centre d’intérêt de la ressource : par exemple, asthme/thérapie, où asthme est le descripteur et thérapie le qualificatif. Les types de ressources sont utilisés afin de catégoriser la nature de la ressource. Les MT sont, en général, des spécialités médicales ou des sciences biologiques, qui ont un lien sémantique avec un ou plusieurs termes MeSH, qualificatifs et types de ressources (Fig. 1). En septembre 2008, nous avons créé 123 MT et 289 types de ressources dans la terminologie de CISMef. Les liste complètes des MT et des types de ressources sont disponibles respectivement aux URL suivantes : [http://doccismef.chu-rouen.fr/liste\\_des\\_meta\\_termes\\_anglais.html](http://doccismef.chu-rouen.fr/liste_des_meta_termes_anglais.html) et <http://www.chu-rouen.fr/documed/typeeng.html>

### 2.2. Contexte multiterminologique

Pour s’adapter à l’information sur le médicament, CISMef a amélioré son serveur de terminologie (URL :

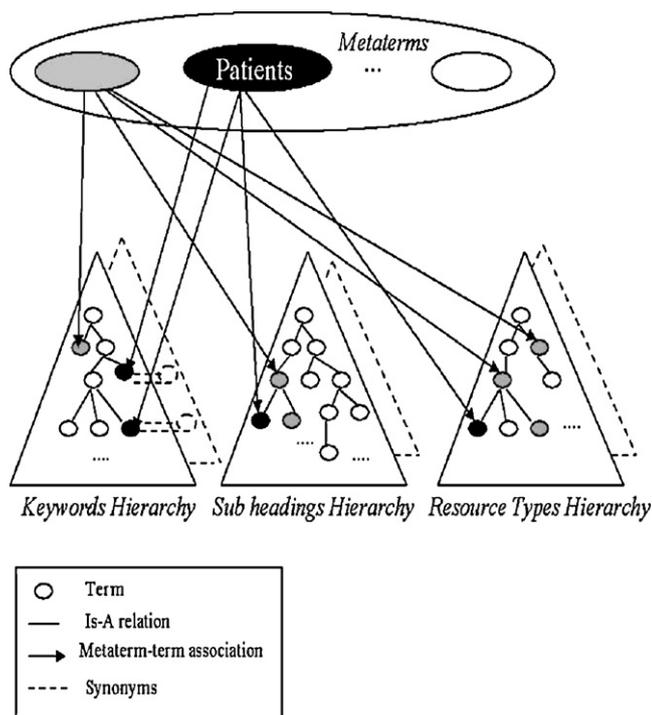


Fig. 1. Liens sémantiques entre les métatermes et les Types de Ressources, qualificatifs et descripteurs CISMef.

rouen.fr/terminologiecismef/), de façon à ce que les utilisateurs du PIM puissent accéder à toutes les substances chimiques (y compris médicamenteuses), aux actions pharmacologiques, ainsi qu'aux types de ressources liés aux médicaments.

Au sein du thésaurus MeSH, les noms des substances chimiques (y compris les substances médicamenteuses) peuvent correspondre soit à des descripteurs ou mots clés hiérarchisés (MC), soit à des *supplementary concept records* (SC), soit à des synonymes de ces termes. Les SC du thésaurus MeSH (environ 175 000) sont des termes de référence non hiérarchisés, permettant de décrire les substances chimiques. Depuis leur intégration dans notre serveur de terminologie, plus de 3000 ont été traduits en français et plus de 3200 synonymes français ont été créés. Sans être considérés comme des descripteurs du thésaurus MeSH, les SC possèdent des relations sémantiques avec ces derniers : d'une part, pour chaque SC, le MeSH recommande une projection vers des descripteurs ; d'autre part, il mentionne le ou les descripteur(s) correspondant à (aux) action(s) pharmacologique(s) de la substance décrite. Par exemple, pour le SC « cétuximab », le MeSH recommande de « mapper » vers le descripteur « anticorps monoclonal » et nous précise que l'action pharmacologique correspondante est « antinéoplasiques ».

En ce qui concerne les médicaments, le plus important désormais est de retenir la notion de substance et non plus seulement la notion de *supplementary concept* ou descripteur MeSH. C'est pourquoi, pour les besoins du PIM, nous avons créé le concept « Substance » (SB) qui permet de regrouper l'ensemble des substances chimiques, qu'il s'agisse des descripteurs ou des *supplementary concepts records*.

Au sein du thésaurus MeSH, comme au sein du serveur de terminologie CISMef, la plupart des termes correspondant à des

Tableau 1

L'arborescence spécifique des types de ressources sur les médicaments.

Information sur le médicament
Avis de vigilance sanitaire
Évaluation médicament
Avis de la commission de transparence
Monographie pharmacie
Notice médicamenteuse
Recommandation de bon usage du médicament
Résumé des caractéristiques du produit

substances sont reliés à des actions pharmacologiques. Selon la National Library of Medicine (NLM), une action pharmacologique est une « catégorie d'actions chimiques et d'utilisations qui ont comme conséquence la prévention, le traitement ou le diagnostic de la maladie. Sont inclus les produits chimiques qui agissent en changeant des fonctions normales du corps et les effets des produits chimiques sur l'environnement ». Une action pharmacologique peut correspondre à un descripteur ou à un concept particulier qui permet de regrouper l'ensemble des substances (qu'il s'agisse de descripteurs [MC] ou SC) ayant une action pharmacologique commune. Par exemple, l'action pharmacologique « antianémiques » permet de regrouper les substances suivantes : acide folique (MC), composés de fer III (MC), darbépoétine alfa (SC), dextriferron (MC), époétine alfa (MC), extraits hépatiques (MC), ferric oxide, saccharated (SC), gluconate ferreux (SC), gluconate ferrique (SC), hexaméthylène bisacétamide (SC), hydroxocobalamine (MC), *iron protein succinylate* (SC) et le téferrol (SC). À ce jour, 374 actions pharmacologiques ont été intégrées au serveur de terminologie CISMef.

La terminologie CISMef a dû aussi être adaptée à la nature des informations sur le médicament, et ce, grâce à l'ajout de types de ressources spécifiques du médicament. Une définition de chacun de ces types de ressources a été fournie soit par l'équipe CISMef, soit par une institution (le plus souvent l'Agence française de sécurité sanitaire des produits de santé (Afssaps)). Une arborescence spécifique sur les médicaments a été créée, avec en tête d'arborescence le type de ressources le plus général « information sur le médicament » (Tableau 1).

La mise en place d'une terminologie adaptée aux médicaments se fait aussi par l'intégration des noms commerciaux, des Dénominations Communes Internationales (DCI) et des différents codes nationaux et internationaux liés aux médicaments et aux substances chimiques tels que le code identifiant de présentation (CIP) [7], code identifiant de spécialité (CIS) [8] et unité commune de dispensation (UCD) [9] pour les codes nationaux, et les codes classification anatomique, thérapeutique et chimique (ATC) [10], Chemical Abstract Service (CAS) [11], European Inventory of Existing Commercial Substances ou encore Inventaire Européen des Substances Commerciales Existantes (EINECS/ELINCS) pour les codes internationaux. Ces fichiers nous ont été fournis par la société Vidal.

Le système de l'ATC [10] est utilisé pour classer les médicaments. C'est le Collaborating Centre for Drug Statistics Methodology de l'Organisation mondiale de la santé (OMS) qui le contrôle. Les médicaments sont divisés en différents groupes

selon l'organe ou le système sur lequel ils agissent et/ou leurs caractéristiques thérapeutiques et chimiques. Le code ATC a la forme générale LCCLLCC où (L représente une lettre et C un chiffre). Dans ce système, les médicaments sont classés en groupes à cinq niveaux différents :

- le premier niveau : groupe anatomique (un caractère alphabétique) ;
- le deuxième niveau : groupe thérapeutique principal (deux caractères numériques) ;
- le troisième niveau : sous-groupe thérapeutique/pharmacologique (un caractère alphabétique) ;
- le quatrième niveau : sous-groupe chimique/thérapeutique/pharmacologique (un caractère alphabétique) Le cinquième niveau : sous-groupe pour la substance chimique : le principe actif individuel ou l'association de principes actifs (deux caractères numériques).

Exemple :

N Le système nerveux  
 N05 Psycholeptiques  
 N05B Anxiolytiques  
 N05BA Benzodiazépines  
 N05BA01 Diazépam

À chaque stade de cette classification correspond un code ATC et un libellé ATC. Le libellé du cinquième niveau correspond à la DCI e la substance, quand elle existe.

Ce code est attribué en fonction de son indication principale. Or, cette dernière peut varier d'un pays à l'autre, ce qui explique qu'il peut exister plusieurs codes ATC pour un même médicament en fonction du pays concerné. C'est le cas pour environ 10 % des médicaments qui n'ont pas le même code ATC entre la France et le Danemark (étude interne réalisée par la société Vidal pour le projet PSIP). Il a donc fallu s'adapter au contexte français et au contexte danois pour pallier au problème des ATC « variants ». Cette adaptation a été rendue possible grâce à la participation de la société Vidal qui a fourni les fichiers adéquats. Le code CAS (*CAS number* ou *CAS registry number* en anglais) d'un produit chimique polymère, d'une séquence biologique ou d'un alliage correspond à son numéro d'enregistrement unique auprès de la banque de données de CAS, une division de l'American Chemical Society (ACS). Le CAS assigne ces numéros, identifiables par un algorithme. L'ACS maintient et commercialise une base de données de ces substances, le CAS Registry. Chaque produit chimique possède un numéro CAS, permettant de l'identifier de manière unique au niveau de la base de données CAS Registry.

Au sein du thésaurus MeSH, les codes CAS sont reliés aux descripteurs et aux SC qui correspondent à des substances et ont ainsi été intégrés au serveur de terminologie CISMef.

Les codes EINECS/ELINCS sont représentés par un inventaire qui définit la liste définitive de toutes les substances chimiques censées se trouver sur le marché communautaire entre le 1<sup>er</sup> janvier 1971 et le 18 septembre 1981. Quant aux numé-

ros European List of Notified Chemical Substances (ELINCS), ils sont décrits par une liste qui complète la liste EINECS et qui attribue un numéro aux nouvelles substances mises sur le marché européen après le 18 septembre 1981. Les nouvelles substances sont incluses au fur et à mesure de leur notification et paraissent lors des mises à jour de l'ELINCS.

En France, le code CIS est à huit chiffres et est attribué aux spécialités pharmaceutiques faisant ou ayant fait l'objet d'une autorisation de mise sur le marché (AMM). Il est administré par l'Afssaps. Le code CIP (code à sept et prochainement à 13 chiffres) identifie la présentation d'une spécialité pharmaceutique. Il est aussi administré par l'Afssaps. Un médicament peut être identifié par plusieurs numéros CIS qui font référence à un dosage et/ou une forme galénique différents pour un médicament spécifique. Pour un même code CIS, on peut avoir plusieurs codes CIP selon les différentes présentations existantes (la taille et/ou le conditionnement).

Le code unité UCD caractérise la plus petite unité intègre utilisée pour la dispensation des médicaments dans les établissements de soins. Il s'agit d'un code à sept chiffres établie par le Club interpharmaceutique, et publié au *Journal officiel*.

### 3. Résultats : création du portail PIM

La construction d'un PIM s'est appuyée sur le portail CISMef, dont il est une sous-partie, mais avec des fonctionnalités spécifiques. L'URL du site PIM est : <http://doccismef.chu-rouen.fr/servlets/PIM>. Pour des raisons de protection industrielle, ce site est protégé par un identifiant et un mot de passe.

PIM a été développé en quatre étapes :

- la première étape a été de créer un MT médicament permettant de regrouper les ressources ayant trait à celui-ci ;
- la deuxième étape a consisté à créer le site portail et d'y associer des formulaires de recherche simple et avancée bilingue ;
- la troisième étape permet la mise en place des liens contextuels vers des banques de données anglophones, en particulier Drug Information Portal [12], de la NLM of United States, dépendant du National Institutes of Health (NIH) et Entrez [13], outil de recherche du National Center for Biotechnology Information (NCBI) dans les sciences de la santé, qui englobe, notamment, PubMed/Medline et les bases données chimiques du NIH ;
- quant à la quatrième étape, elle se traduit par la mise au point d'un « Google sélection PIM » permettant d'effectuer des recherches Google limitées à une sélection de site de qualité concernant le médicament.

Ce portail bilingue a été créé afin de faciliter l'accès à l'information concernant le médicament. Il est destiné, non seulement, aux professionnels de santé mais aussi au grand public. Le portail permet d'accéder à de nombreuses ressources, émanant pour la plupart, d'institutions ou de sociétés savantes.

Tableau 2

Description booléenne du métaterme médicament.

Requête : médicament.mt

*Reformulation de la requête* : traitement médicamenteux.**mc** ou coût médicament.**mc** ou médicament orphelin.**mc** ou pharmacologie.**mc** ou toxicité des médicaments.**mc** ou utilisation médicament.**mc** ou phénomènes chimiques et pharmacologiques.**mc** ou préparations pharmaceutiques.**mc** ou évaluation préclinique médicament.**mc** ou voies d'administrations des médicaments.**mc** ou actions pharmacologiques.**mc** ou malformations dues aux médicaments et aux drogues.**mc** ou autorisation de mise sur le marché d'un médicament.**mc** ou évaluation médicament.**mc** ou hypersensibilité médicamenteuse.**mc** ou produits biopharmaceutiques.**mc** ou système distribution médicaments.**mc** ou technologie pharmaceutique.**mc** ou contamination de médicaments.**mc** ou rythme administration médicament.**mc** ou surveillance médicament.**mc** ou stents à élution de médicament.**mc** ou biomarqueurs pharmacologiques.**mc** ou information sur le médicament.**tr** ou action des médicaments et substances chimiques.**qu** ou traitement médicamenteux.**qu** ou administration et posologie.**qu** ou pharmacocinétique.**qu**

**mc** : code booléen correspondant à un mot clef ; **qu** : code booléen correspondant à un qualificatif ; **tr** : code booléen correspondant à un type de ressource ; caractères gras pour mettre en relief les codes booléens ; le caractère italique (« ou ») pour signaler l'opérateur booléen.

### 3.1. Étape 1

Les MT sont utilisés pour regrouper les descripteurs, qualificatifs et types de ressources (Fig. 1) qui correspondent à une même thématique et permettent ainsi de retrouver l'ensemble des documents ayant trait à cette thématique. De fait de l'organisation hiérarchique de la terminologie CISMef (descripteurs, qualificatifs et types de ressources), tous les termes hiérarchiquement inférieurs à l'ensemble des termes sélectionnés sont ainsi rattachés implicitement à ce métaterme. Pour ce faire, nous avons rattaché manuellement au MT « médicament » tous les descripteurs MeSH (MC) en rapport avec le médicament, tels que : actions pharmacologiques, AMM d'un médicament, contamination de médicaments. . .

Ensuite, nous avons sélectionné les qualificatifs (QU) qui sont utilisés pour l'indexation des documents relatifs à des médicaments, à savoir : action des médicaments et substances chimiques, pharmacocinétique, traitement médicamenteux et administration et posologie. Enfin, nous avons relié à ce MT tous les documents qui avaient été décrits avec le type de ressource « information sur le médicament » et implicitement tous les types de ressources hiérarchiquement inférieurs (Tableau 1).

Le regroupement de l'ensemble de ces termes au niveau du MT « médicament » permet d'établir la requête suivante (Tableau 2) et, par conséquent, de réunir tous les documents relatifs au médicament. La création du MT « Médicament » a permis de regrouper plus de 14 400 ressources.

### 3.2. Étape 2

PIM contient essentiellement un outil de recherche inspiré grandement de Doc'CISMef, mais avec quelques spécificités centrées sur le médicament. L'outil de recherche de PIM contient une recherche simple et une recherche avancée, les deux

Fig. 2. Page de garde du portail d'information sur le médicament (PIM).

sous forme bilingue (français et anglais). Ces deux modes de recherche sont spécifiques d'un pays (France ou Danemark, par exemple), du fait des codes ATC variant d'un pays à l'autre pour un même médicament.

La recherche simple peut se faire sur le nom commercial ou la DCI, ou sur n'importe quel code que nous avons vu au-dessus (Fig. 2).

La recherche avancée permet une recherche spécifique sur l'ensemble des éléments décrivant les médicaments :

- nom commercial/*brand name* ;
- DCI/*International Nonproprietary Name* (INN) ;
- action pharmacologique/*pharmacologic action* ;
- code ATC et code CAS ;
- code CIP, code CIS et code UCD (formulaire français uniquement) ;
- codes EINECS/ELINCS.

Ainsi, la recherche peut se faire de différentes manières et par le biais de plusieurs codes. Concernant les différents codes proposés, certains sont exclusivement français (codes CIP, CIS, UCD), d'autres sont européens (codes EINECS/ELINCS) et d'autres internationaux (codes ATC, codes CAS), c'est pour cette raison qu'il a fallu adapter les formulaires de recherche selon les différentes variations et besoins de l'utilisateur.

En plus, comme on l'a mentionné au premier paragraphe : un même principe actif de médicament peut avoir plusieurs codes ATC en fonction des indications reconnues, il a donc fallu s'adapter au contexte français et au contexte danois pour pallier au problème des ATC « variants ». Cette adaptation a été rendue possible grâce à la participation de la société Vidal qui a fourni les fichiers adéquats.

### 3.3. Étape 3

La troisième étape permet la mise en place des liens contextuels vers des banques de données anglophones (Drug

Information Portal de la NLM of United States [12] et Entrez, outil de recherche du NCBI dans les sciences de la santé, qui englobe notamment PubMed/Medline et PubChem Substance). Ces liens contextuels sont fondés sur la traduction de termes MeSH entre le français et l'anglais, comme nous l'avons déjà fait pour le moteur de recherche Doc'UMVF [4].

#### 3.4. Étape 4

Enfin, nous avons réalisé un Google spécifique au médicament nommé « Google sélection PIM ». Nous avons utilisé « Google™ Custom Search Engine » (Google CSE), en utilisant la plateforme « Google Co-op™ » [14]. Il est possible de définir une version customisée de Google, fondée sur la base du moteur de recherche Google (URL: [www.google.com](http://www.google.com)). Nous avons fourni à Google une liste de sites éditeurs de qualité, déjà recensée par CISMef mais restreint au domaine du médicament ( $n=79$ ), pour créer le Google-PIM. Ces sites éditeurs sont les principes sites institutionnels francophones (comme la Haute Autorité de santé ou l'Afssaps). Du fait que le crawler de Google récupère au moins toutes les pages statiques d'un site, le corpus de Google PIM inclut toutes les ressources de PIM, mais aussi d'autres pages qui n'ont pas été sélectionnées manuellement par l'équipe CISMef. Pour la version danoise du PIM, une liste spécifique devra être définies par les partenaires danois du projet PSIP.

Les résultats d'une requête dans le PIM sont présentés sous forme de notices descriptives permettant d'accéder directement à des informations émanant des différentes institutions francophones ou européennes (comme l'European Medicines Evaluation Agency [EMA]), mais aussi, de sociétés savantes. Au sein de chaque notice, un champ nommé « substance » a été mis en place (équivalent au « Substance Name » de PubMed), permettant de regrouper et de repérer les substances impliquées, ainsi que leurs actions pharmacologiques contextuelles.

#### 4. Discussion et perspectives

En 2008, la NLM a mis en place le « Drug Information Portal » [12]. Ce portail permet d'accéder aux informations concernant plus de 12 000 médicaments. La recherche peut s'y effectuer à partir du nom générique ou du nom commercial. À notre connaissance, la recherche par code CAS ou ATC n'est pas possible dans le US Drug Information Portal, alors qu'elle l'est pour le CAS dans PubMed. Une fois la recherche effectuée, le système propose de consulter les ressources de différentes banques de données. Il s'agit de banques de données de la NLM (Medline, MedlinePlus, Toxnet, Clinicaltrials.gov, DailyMed, AIDSInfo, ToxSeek. . .), et plus généralement du NIH, mais aussi de différentes agences gouvernementales américaines comme la Food and Drug Administration (FDA), le Centers for Disease Control and Prevention (CDC), etc. À la différence du PIM, l'interrogation n'est possible qu'en langue anglaise et seuls les noms génériques ou les noms commerciaux états-uniens sont reconnus.

Par ailleurs, au sein du PIM, l'interrogation peut se faire dans un contexte multiterminologique (la recherche est possible à

partir de la DCI, du nom commercial, des codes ATC, codes CIP, codes CIS, codes UCD, codes CAS, codes EINECS/ELINCS, et des termes du thésaurus MeSH correspondant à des noms de substances). Les spécificités nationales sont aussi prises en compte.

La mise en place d'un formulaire de recherche avancé permet aussi de mieux cibler ses recherches, en croisant, par exemple, une DCI avec une action pharmacologique.

Enfin, une fois l'interrogation réalisée, le système permet d'effectuer les mêmes recherches sur le « Drug Information Portal » de la NLM, permettant ainsi d'accéder aux médicaments états-uniens et ce, grâce à un algorithme de recherche adapté à ce contexte.

La création du Drug Information Button (DIB ; bouton d'information sur le médicament), accessible à partir des logiciels de prescriptions informatisées, permettra aux médecins prescripteurs d'accéder à des connaissances médicamenteuses contextuelles via le PIM.

#### 5. Conflits d'intérêts

Néant

#### Remerciements

Les auteurs remercient Mme Ioana Robu pour son aide à la traduction en anglais du résumé.

#### Références

- [1] Keselman A, Browne AC, Kaufman DR. Consumer health information seeking as hypothesis testing. *J Am Med Inform Assoc* 2008;15:484–95. Epub 2008 Apr 24.
- [2] Boyer C, Gaudinat A, Baujard V, Geissbühler A. Health on the Net Foundation: assessing the quality of health web pages all over the world. *Medinfo* 2007;12:1017–21.
- [3] PSIP Web site; URL: <https://www.psip-project.eu/>. [accessed 2008 Aug 08].
- [4] Douyère M, Soualmia LF, Névéol A, Rogozan A, Dahamna B, Leroy JP, et al. Enhancing the MeSH thesaurus to retrieve French online health resources in a quality-controlled gateway. *Health Info Libr J* 2004;21:253–61.
- [5] Koch T. Quality-controlled subject gateways: definitions, typologies, empirical overview. *Subject gateways. Online Information Review* 2000;24:24–34.
- [6] Dekkers M, Weibel S. State of the Dublin core metadata initiative. *D-Lib Mag* 2003;9 (<http://webdoc.sub.gwdg.de/edoc/aw/d-lib/dlib/april03/weibel/04weibel.html>) date de consultation juillet 2008.
- [7] Standardisation de la codification et du marquage des produits de santé. URL: <http://www.cipclub.org/institutionnel/fr/real/pdf/CIP%20Recommandation%20Codification%20V1-2.pdf>. Consulté en juillet 2008.
- [8] Health Administration systems 2008. URL: <http://www.has-sante.fr/portail/jcms/c.671889/de-la-certification-des-lap>. Consulté en juillet 2008.
- [9] L'assurance maladie en ligne URL: <http://www.ameli.fr/professionnels-de-sante/%20directeurs-d-etablissements-de-sante/codage/medicaments/medicaments-a-code-ucd.php>. [accessed 2008].
- [10] Skrbo A, Begović B, Skrbo S. Classification of drugs using the ATC system (Anatomic, Therapeutic, Chemical Classification) and the latest changes. *Med Arh* 2004;58(Suppl 2):138–41 [Bosnian. PMID: 5137231].

- [11] Dittmar PG, Stobaugh RE, Watson CE. The chemical abstracts service chemical registry system I. General Design. *J Chem Inf Comput Sci* 1976;16:111.
- [12] US NLM Drug Information Portal; URL: <http://druginfo.nlm.nih.gov/drugportal/drugportal.jsp>. [accessed 2008 Feb 26].
- [13] Schuler GD, Epstein JA, Ohkawa H, Kans JA. Entrez: molecular biology database and retrieval system. *Methods Enzymol* 1996;266:141–62.
- [14] Google Coop; URL: <http://www.google.com/coop/>. [accessed 2008 Feb 26].