

Recherche des sources d'information institutionnelle de santé françaises : le « Catalogue et index des sites médicaux francophones » (CISMeF)

Saoussen Sakji, Benoît Thirion, Badisse Dahamna, Stéphan Jacques Darmoni

CISMeF, Centre hospitalier universitaire de Rouen, GCSIS, LITIS EA 4108,
Institut de recherche biomédicale, Université de Rouen, F-76031 Rouen, France

Correspondance :

Stéphan Jacques Darmoni, CISMeF, Centre hospitalier universitaire de Rouen,
1 rue de Germont, F-76031 Rouen, France.
stefan.darmoni@chu-rouen.fr

Disponible sur internet le :

■ Key points

Searching French institutional health information sources: Catalogue and index of French-language medical sites (CISMeF)

The Catalogue and index of French-language medical sites (CISMeF) is a medical portal that provides users with results as pertinent as possible according to their requirements, expectations, and context of use.

Indexing and single-term research are based on the Medical subject headings (MeSH) thesaurus.

The integration of new medical terminology for indexing the catalogue's resources is intended to minimize false-negatives during searches and to contextualize the users' needs.

The creation of a drug information portal makes more targeted research possible, with numerous entries according to user (physicians, pharmacists, chemists, and pharmacologists).

For simplicity's sake, the catalogue's index of resources by different nomenclatures is not entirely displayed. The choice of display is left to the user, with MeSH only as the default.

These multi-nomenclature tools should be applicable as well to electronic patient records. In this case, the objective is to improve patient care by better searches and identification of the information required during consultations and hospitalization.

■ Points essentiels

Le Catalogue et index des sites médicaux francophones (CISMeF), est un portail dédié au domaine médical permettant de fournir à l'utilisateur un résultat aussi pertinent que possible selon ses besoins, ses attentes et son contexte d'utilisation.

L'indexation et la recherche monoterminologiques sont fondées sur le thésaurus *Medical subject headings* (MeSH).

L'intégration de nouvelles terminologies médicales pour l'indexation des ressources du catalogue a pour but de minimiser les faux-négatifs lors de la recherche d'information et de contextualiser les besoins des utilisateurs.

La création du Portail d'information sur les médicaments (PIM) permet une recherche plus ciblée avec de nombreux accès selon l'utilisateur (médecin, pharmacien, chimiste, pharmacologiste). Par souci de simplicité pour l'utilisateur, l'indexation des ressources du catalogue par les différentes terminologies n'est pas affichée en totalité. Nous laissons à l'utilisateur le choix d'affichage, avec par défaut seulement le MeSH.

L'application de l'univers multiterminologique devrait être applicable sur les dossiers électroniques des patients (DEP). Dans ce cas, l'objectif serait de mieux suivre les patients par une meilleure recherche d'information et bien cerner l'information requise lors des consultations et hospitalisations.

Avec le développement des nouvelles technologies de l'information, de la communication et de l'informatique, le volume d'information disponible sur l'Internet est en augmentation. Ce développement a modifié les modes d'accès à l'information. L'Internet diffuse des informations hétérogènes, parfois peu validées ou mal repérées.

S'intéressant au domaine médical, quelles que soient leurs expériences du Web et leurs compétences en recherche d'information, les utilisateurs rencontrent des difficultés à chercher une information de santé sur l'Internet [1]. La plupart des utilisateurs entament leurs recherches *via* les moteurs de recherche généralistes (tel que *Google*), plutôt que les bases de données médicales spécialisées [2]. Cependant, avoir recours à un moteur de recherche spécialisé peut révéler de meilleurs résultats répondant aux besoins et aux attentes des utilisateurs.

À travers cet article nous présentons le portail CISMeF comme exemple de recherche d'informations sur l'Internet. Un portail inclut le plus souvent un moteur de recherche et se définit comme un site *Web* recensant les principales ressources disponibles dans un domaine donné.

Catalogue CISMeF et sa monoterminologie

CISMeF (URL : <http://www.chu-rouen.fr/cismef>) acronyme de Catalogue et index des sites médicaux francophones, est un portail de santé qui a été conçu pour cataloguer et indexer les sources d'information institutionnelle de santé françaises les plus importantes (*encadrés 1 et 2*).

CISMeF applique une politique éditoriale stricte : recenser des ressources institutionnelles ; ne pas recenser de ressources commerciales. CISMeF adhère aux principes de qualité de l'information de santé sur l'Internet défini par *Health on net foundation* depuis dix ans maintenant [3]. Dans ce contexte, le portail CISMeF peut être défini comme un portail de santé de qualité. Comme le définit Koch [4], un portail de qualité dédié est un ensemble de services de l'Internet qui respectent et intègrent plusieurs mesures de qualité afin de détecter d'une manière automatique les nouvelles ressources de qualité qui font référence au domaine d'application. Intégrant un moteur de recherche spécialisé DocCISMeF, le portail CISMeF permet une recherche plus pertinente non seulement pour les professionnels de santé et les étudiants en médecine mais aussi pour les patients et le grand public.

Depuis 1995, CISMeF utilise deux outils standards pour organiser l'information : le thésaurus *Medical subject headings* (MeSH) (utilisé le plus souvent pour indexer les articles scientifiques de la base de données *Medline*) et un ensemble de métadonnées extraites du *Dublin Core* [5]. Les métadonnées se réfèrent aux informations descriptives des ressources *Web* et dont les plus importantes sont : site éditeur, auteurs, titre de la ressource, date de publication, résumé, indexation avec ou sans vocabulaire contrôlé, type de la ressource, format. Pour plus de précision, huit métadonnées spécifiques à CISMeF ont été ajoutées, telles que pays, institution, niveau de preuve. La terminologie CISMeF encapsule la version française du thésaurus MeSH dans la mesure où, d'une part, elle comprend des concepts déjà existants dans le MeSH (descripteurs,

Glossaire

| | |
|---------------|---|
| ATC | Anatomique, thérapeutique et chimique |
| CAS | <i>Chemical abstract service</i> |
| CCAM | Classification commune des actes médicaux |
| CIF | Classification internationale du fonctionnement, du handicap et de la santé |
| CIM-10 | Classification internationale des maladies, version10 |
| CIP | Code d'identification de la présentation |
| CIS | Code d'identification de la spécialité |
| CISMeF | Catalogue et index des sites médicaux francophones |
| CISP2 | Classification internationale des soins primaires, deuxième édition |
| DCI | dénomination commune internationale |
| DEP | dossiers électroniques des patients |
| DRC | Dictionnaire des résultats de consultation |
| JD | <i>journal descriptors</i> |
| MeSH | <i>Medical subject headings</i> |
| NLM | <i>US National library of medicine</i> |
| PIM | portail d'information sur les médicaments |
| PSIP | <i>Patient safety through intelligent procedures in medication</i> |
| SNOMED | <i>Systematized nomenclature of medicine</i> |
| SR | stratégies de recherche |
| TALN | traitement automatique du langage naturel |
| UCD | Unités communes de dispensation |

ENCADRÉ 1

CISMeF : financement

Le Catalogue et index des sites médicaux francophones (CISMeF) est depuis son origine en février 1995 accessible gratuitement. Néanmoins, la part de financement du CHU de Rouen a diminué au cours du temps pour atteindre en 2009 environ 30 % des coûts directs. Le CISMeF a obtenu un grand nombre de financements publics et plus récemment privés. Tous ces financements ne sont pas pérennes pour la plupart. Un changement stratégique a été entrepris en 2005: nous avons déjà développé un logiciel de connaissance contextuelle valorisé à l'Université de Rouen ; une entreprise privée à partir des différentes réalisations de l'équipe CISMeF est en cours d'incubation en Haute-Normandie.

ENCADRÉ 2

CISMeF : statistiques de consultation

Le Catalogue et index des sites médicaux francophones (CISMeF) est utilisé par environ 40 000 personnes par jour ouvré (sans savoir la proportion respective des professionnels de santé et des patients, essentiellement pour des raisons éthiques). Ces chiffres nous classait en 5^e position des sites de santé en France en 2008 (source: *Le Monde*) derrière par exemple *Doctissimo*. Sur le plan institutionnel, CISMeF se classe derrière *ameli.fr* mais bien devant la majorité des sites institutionnels producteurs d'information, comme la Haute autorité de santé (HAS) ou l'Agence française de sécurité sanitaire des produits de santé (AFSSAPS). Sur le plan hospitalier, CISMeF et le CHU de Rouen ont le premier trafic des CHU français.

74 qualificatifs et *supplementary concepts*) et, d'autre part, elle
75 utilise de nouveaux concepts (métatermes, types de
76 ressources et stratégies de recherche). En effet, dans le but
77 d'adapter le MeSH aux ressources de santé sur l'Internet plutôt
78 qu'au signalement des références bibliographiques, des amé-
79 liorations ont été réalisées depuis l'année 2000. En plus des
80 *supplementary concepts* (termes MeSH non hiérarchiques
81 faisant référence aux substances chimiques), des descripteurs
82 MeSH (termes qui permettent l'indexation des ressources) et
83 des qualificatifs MeSH (termes qui permettent de préciser le
84 sens d'un descripteur), deux nouveaux concepts ont été créés
85 par l'équipe CISMeF : les types de ressources et les méta-
86 termes.

87 **Types de ressources**

88 La paire (descripteur/qualificatif) décrit le centre d'intérêt du
89 document. Par exemple, une ressource décrite par le terme
90 composé asthme/thérapie, où asthme est le descripteur et
91 thérapie est le qualificatif, indique que le contenu du document
92 traite des effets thérapeutiques de l'asthme. Comme les
93 descripteurs MeSH et les qualificatifs MeSH, les types de
94 ressources sont organisés hiérarchiquement depuis leur créa-
95 tion en 1997. Ils sont une extension des types de publications
96 du MeSH et sont utilisés afin de catégoriser la nature de la
97 ressource, son contenant tels que : information patient et grand
98 public, dictionnaire médical, ouvrage d'orientation... À ce jour,
99 nous avons créé 293 types de ressources CISMeF, dont la liste
100 est disponible : [http://www.chu-rouen.fr/documed/typeres-
102 source.html](http://www.chu-rouen.fr/documed/typeres-
101 source.html). Par exemple, si une ressource de l'Internet est une
103 recommandation sur le traitement de l'asthme, le concept de
104 « recommandation » est le type de ressources puisqu'il s'agit du
105 contenant, le concept d'« asthme » est le descripteur MeSH et le
106 concept de « traitement » est un qualificatif (ces deux derniers
traitant du contenu de ce document).

Métatermes

Les métatermes CISMeF (tels que cardiologie, oncologie) sont des spécialités médicales ou des sciences biologiques ayant un lien sémantique avec un ou plusieurs descripteurs MeSH, qualificatifs MeSH et types de ressources. Cette liste a été établie en partie à l'aide des « *Nomenclatures des professions de santé : formations, qualifications, postes hospitaliers, activités exercées* » [Bulletin officiel ; 93-7 bis. Paris : Journal officiel] et des sections et sous-sections médicales du Conseil national des universités. Ces liens sémantiques ont été établis manuellement par le conservateur de la bibliothèque médicale et codirecteur de l'équipe CISMeF (Benoit Thirion). À ce jour, nous avons créé 116 métatermes CISMeF, dont la liste est disponible : [http://doccismef.chu-rouen.fr/liste_des_meta-
termes.html](http://doccismef.chu-rouen.fr/liste_des_meta-
termes.html)

L'ajout d'un tel concept permet d'avoir une vision générale du MeSH concernant une spécialité. Les travaux de Gehanno *et al.* [6] ont montré que l'emploi des métatermes permet d'optimiser la recherche en augmentant le rappel. Par exemple, utiliser le métaterme « psychiatrie » permet une stratégie efficace pour couvrir tout le champ de la psychiatrie dans le MeSH : au lieu d'explorer une seule hiérarchie MeSH, on peut obtenir une expansion automatique de la requête en explorant d'autres relations MeSH. Par exemple, « hôpital psychiatrique », un mot clef MeSH, et « un dispensaire de santé mentale », un type de ressource, vont être exploités grâce aux liens sémantiques existants entre le métaterme et les autres concepts de la terminologie de CISMeF (descripteur, qualificatif et type de ressources) (figure 1). Les métatermes sont également utilisés pour catégoriser les ressources CISMeF, ce qui permet d'avoir une vision par discipline plus générale que celle indiquée par l'indexation par mots-clés. Dans une première étude, cette catégorisation par les métatermes a montré une précision de 81 % et un rappel de 93 % sur 123 ressources [7]. Grâce à une collaboration importante avec la *US National library of medicine* (NLM), une seconde étude a comparé les métatermes de CISMeF et les *journal descriptors* (JD) de la NLM pour catégoriser 100 articles scientifiques indexés dans *Medline* : aucune différence significative entre les deux approches n'a été trouvée (les métatermes fondés sur des liens sémantiques alors que les JDs sont calculés statistiquement) [8].

Les stratégies de recherche (SR) permettent d'introduire des concepts non précis dans le MeSH en construisant des requêtes complexes maniant plus d'un descripteur MeSH ou d'un qualificatif. Par exemple, la notion de fonction de la conscience n'existe pas dans le MeSH. Nous avons créé la stratégie de recherche suivante SR (fonction de la conscience) = conscience/physiologie, où conscience est un descripteur MeSH et physiologie un qualificatif. À ce jour, nous avons créé dans CISMeF plus 200 stratégies de recherche.

107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157

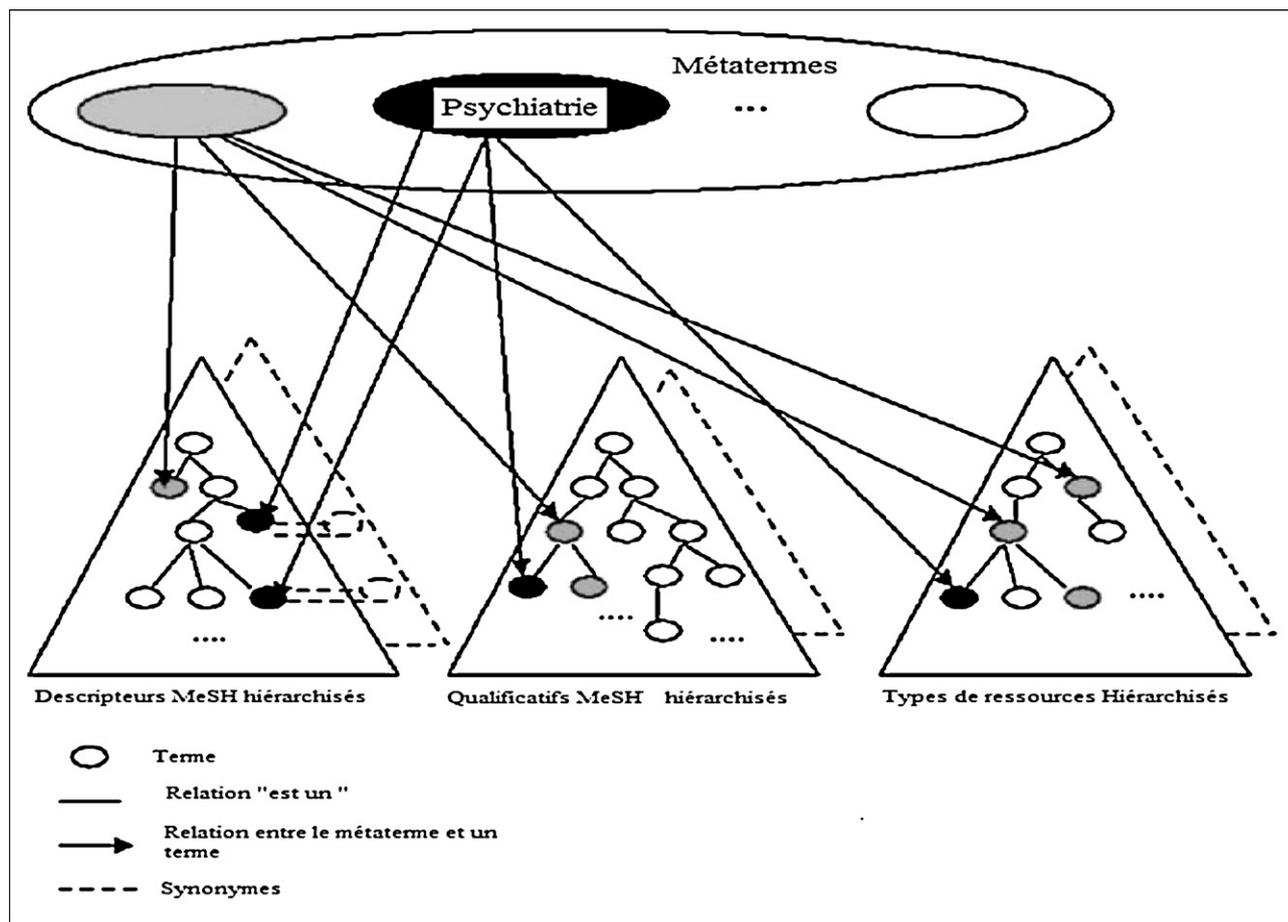


FIGURE 1

Les relations entre les métatermes et les descripteurs MeSH, les qualificatifs MeSH et les types de ressources CISMef

Stratégie d'indexation des ressources du catalogue CISMef

Selon le degré d'importance des ressources recensées dans le catalogue CISMef, trois niveaux d'indexation sont appliqués : le premier niveau (N1) pour une indexation manuelle, le deuxième niveau (N2) pour une indexation supervisée et le troisième niveau (N3) pour une indexation automatique. L'indexation manuelle concerne les ressources jugées importantes et prioritaires telles que : les recommandations nationales, les lectures critiques d'articles, les ressources issues des sites institutionnels (ministériels ou gouvernementaux), les cours de campus nationaux. L'indexation supervisée concerne les ressources qui sont moins importantes que celles du premier niveau, cependant jugées assez importantes pour qu'elles ne soient pas indexées qu'automatiquement. Ainsi, l'indexation supervisée est, d'abord, automatique, puis revue

manuellement dans un second temps par les indexeurs de l'équipe CISMef. On retrouve les rapports techniques, les études d'évaluation, les articles de périodiques concernant les formations médicales continues et les ressources sur l'information sur les médicaments. L'indexation automatique concerne les ressources qui ont une importance mineure dans notre contexte, telles que : les rapports sur la politique de santé et de santé publique, les cours facultaires, certaines ressources sur l'information sur les médicaments. À ce jour, le catalogue contient 19 164 ressources indexées manuellement (N1), 22 542 ressources du niveau N2 (dont 9 141 ressources en attente de supervision) et 24 695 ressources indexées automatiquement (N3)¹.

¹ Statistiques datant du 2 avril 2009

187 **Veille documentaire**
188 La veille documentaire se traduit par la recherche sur l'Internet
189 et la sélection de ressources électroniques susceptibles d'être
190 intégrées au catalogue CISMeF.
191 Pour ce faire, une liste de sites publiant ou référant régu-
192 lièrement des documents relatifs à la santé a été construite,
193 enrichie et mise à jour, depuis la création du catalogue en 1995.
194 Ces sites sont classés en fonction de la fréquence à laquelle ils
195 sont actualisés (quotidienne, hebdomadaire, mensuelle, an-
196 nuelle). Ils représentent soit des catalogues généralistes (ex :
197 *Nomade, Yahoo*) soit des sites spécifiques au domaine de la
198 santé (ex : *Santé Canada, ANAES*). Quel que soit leur domaine
199 d'application, ils ont en commun la particularité d'annoncer, sur
200 une page spécifique « nouveauté », les pages *Web* ou les
201 documents nouvellement publiés ou référencés, avec les liens
202 correspondants. Les ressources ainsi recensées sont soumises à
203 une sélection minutieuse afin de vérifier que le document n'est
204 pas déjà répertorié dans la base CISMeF, et par la suite s'assurer
205 de sa pertinence pour le catalogue.
206 Cette tâche est réalisée par un documentaliste qui se charge de la
207 surveillance quotidienne de 854 pages *Web* et flux RSS. Une
208 veille automatique sur ces mêmes sites éditeurs permet de
209 retrouver d'autres ressources. Depuis septembre 2008, une
210 moyenne de 23 ressources par jour est recensée. Cette liste
211 est soumise, par la suite, au conservateur de la bibliothèque
212 médicale qui valide le degré de pertinence des documents. Ainsi,
213 les ressources sont répertoriées comme suit : cinq ressources (en
214 moyenne) au niveau d'indexation manuelle, dix ressources (en
215 moyenne) au niveau d'indexation supervisée, quatre ressources
216 (en moyenne) au niveau d'indexation automatique et
217 quatre ressources (en moyenne) jugées non pertinentes pour
218 qu'elles soient intégrées dans le catalogue CISMeF. Cependant,
219 ces chiffres ne correspondent pas tout à fait à l'accroissement du
220 catalogue puisqu'il faut tenir compte des suppressions et des
221 ajouts « en masse ». Par exemple, le documentaliste a récem-
222 ment rajouté 600 ressources en N2 du site <http://campus.neurochirurgie.fr/>. Les suppressions sont le plus souvent dues à des
223 ressources retirées du Web par leur auteur, aboutissant à la
224 fameuse erreur 404 (document non retrouvé).
225

226 Recherche d'information au début des 227 années 2000

228 La recherche d'information au sein du catalogue CISMeF peut être
229 réalisée de plusieurs façons : recherche simple (texte libre),
230 recherche avancée (selon plusieurs critères de sélection), re-
231 cherche booléenne (grâce aux opérateurs booléens)... Lors
232 d'une requête simple, le moteur de recherche fait correspondre,
233 tant que possible, les termes de la requête avec les termes MeSH
234 d'indexation du document. Les ressources affichées sont celles
235 qui sont indexées par les termes de la requête et leurs équiva-
236 lents (les synonymes, les termes hiérarchiquement inférieurs).

Le processus de la recherche d'information se déroule sur
plusieurs étapes. En effet, la requête de l'utilisateur est traitée
par certaines techniques du TALN (traitement automatique du
langage naturel) pour obtenir un ensemble de termes re-
présentant la requête initiale. Par exemple, la requête *q* :
l'asthme de l'enfant est représentée par les termes *asthme*
et *enfant* qui sont des descripteurs MeSH. D'autre part, les
ressources de CISMeF sont indexées par le thésaurus MeSH.
Chaque document est représenté par un ensemble de termes
(descripteurs MeSH) décrivant au mieux son contenu. Disposant
de ces deux représentations (celle du document et celle de la
requête) la mise en correspondance est réalisée permettant de
retrouver les ressources répondant aux besoins de l'utilisateur.

À ses débuts, la recherche d'information s'est déroulée en cinq étapes [9]

- Étape 1 : rechercher une correspondance possible entre la
représentation de la requête (et les termes d'indexation du
document. Par exemple, pour la requête *q* on cherche
enfant.mr et asthme.mr ; (*mr* (mot réservé) : fait référence au
terme d'indexation) ;
- Étape 2 : rechercher les termes représentatifs de la requête
dans le titre de la ressource. Par exemple, pour la requête *q*, on
cherche *enfant*.ti et asthme*.ti* ; (*ti* : fait référence au titre) ;
- Étape 3 : rechercher les termes représentatifs de la requête,
aussi bien dans le titre qu'au niveau des termes d'indexation
du document. Par exemple, pour la requête *q'* : *épigastralgie*
incidence, on cherche *incidence.mr et épigastralgie.ti* puisque
le terme *épigastralgie* n'est pas un descripteur MeSH ;
- Étape 4 : rechercher une correspondance entre les termes
représentatifs de la requête et les termes d'indexation du
document. Si ce n'est pas le cas, rechercher les termes
représentatifs de la requête au niveau des métadonnées du
document (le titre de la ressource, la date de publication, la
description du contenu, le type de la ressource, le format, le
pays...). Le troisième niveau de recherche dans cette étape
est de rechercher la requête initiale au niveau des
métadonnées avec une mesure d'adjacence égale à *n* où
 $n = 5 \times (\text{nombre de termes de la requête} - 1)$. Par exemple, la
requête *les problèmes respiratoires des enfants* est traduite
par (*enfant.mr et problemes.tc et respiratoires.tc*) ou
(*problemes respiratoires enfants.at*) ; (*tc* (tout champ) : fait
référence aux métadonnées, et *at* : fait référence à la mesure
d'adjacence). Dans cet exemple, on cherche le terme *enfant*
en tant que descripteur MeSH et les termes *problemes* et
respiratoires au niveau de toutes les métadonnées (non
seulement au niveau du titre de la ressource) où on cherche
tous les termes de la requête (*problemes respiratoires*
enfants) au niveau de toutes les métadonnées avec une
mesure d'adjacence égale à 5 ;
- Étape 5 : rechercher les termes représentatifs de la requête en
tant que termes d'indexation du document, ou au niveau de

289 toutes les métadonnées ou en plein texte avec une mesure
290 d'adjacence égale à n où $n = 10 \times$ (nombre de termes de la
291 requête - 1). Par exemple, la requête *bronchite asthmatifor-*
292 *me* est traduite par (*bronchite.mr et asthmatiforme.tc*) ou
293 (*bronchite asthmatiforme.ap*) ; ap : fait référence à la mesure
294 d'adjacence en plein texte.

Recherche d'information en 2007

295 En 2007, afin d'améliorer le rappel et la qualité des ressources
296 trouvées lors de la recherche d'information, l'algorithme a été
297 simplifié en trois étapes, notamment pour considérer au même
298 niveau la recherche d'information sur les termes MeSH et sur le
299 titre de la ressource. Par comparaison, l'algorithme de PubMed
300 par défaut met au même niveau la recherche sur les termes
301 MeSH d'une part et la recherche sur le titre et le résumé d'une
302 référence bibliographique d'autre part. Par conséquent, la
303 recherche d'information dans CISMeF se déroulait en :

- 304 • Étape 1 : rechercher les termes représentatifs de la requête dans
305 le titre de la ressource et au niveau des termes d'indexation ;
- 306 • Étape 2 : rechercher les termes représentatifs de la requête
307 dans toutes les métadonnées de la ressource avec une mesure
308 d'adjacence entre les mots ;
- 309 • Étape 3 : rechercher les termes représentatifs de la requête en
310 plein texte avec une mesure d'adjacence entre les mots.
311

Passage d'un monde monoterminologique vers un univers multiterminologique

312 Dans un objectif d'une recherche d'information plus exhaustive
313 et plus efficace en termes de rappel², l'équipe CISMeF a pris la
314 décision stratégique de passer d'un monde monoterminologique
315 à un univers multiterminologique depuis 2005. Cet objectif a dû
316 prendre en compte la disponibilité de plusieurs terminologies,
317 classifications, thésaurus et nomenclatures médicaux disponibles
318 en français, et l'alignement entre celles-ci, notamment grâce au
319 projet InterSTIS³ dont l'objectif est l'interopérabilité sémantique
320 inter et intraterminologies [10].

321 Grâce aux relations terminologiques on peut améliorer la
322 recherche d'information et mieux répondre aux requêtes de
323 l'utilisateur et ce *via* l'expansion ou la reformulation de la
324 requête. On peut chercher toutes les liaisons possibles entre
325 les termes de la requête appartenant à une terminologie
326 donnée et tous les termes des autres terminologies qui sont
327 en relation. Cette procédure permet d'élargir le champ de
328 la recherche de l'utilisateur selon son contexte, sans néanmoins
329 mettre en cause la pertinence thématique de l'information ni le
330 degré de précision du système. Par exemple, grâce à la

333 correspondance entre le terme MeSH « appareil correction
334 auditive » et le terme *Systematized nomenclature of medicine*
335 (SNOMED) « prothèse auditive », on peut enrichir notre résultat
336 et retrouver toutes les ressources indexées non seulement par
337 le terme « appareil correction auditive » mais aussi par le terme
338 « prothèses auditives » et les termes qui les subsument.

339 Le passage à un univers multiterminologique se traduit par la
340 mise en place d'une structure intégrant les terminologies de
341 santé en français telles que : le thésaurus MeSH, la SNOMED
342 [11], la Classification internationale des maladies, version 10
343 (CIM-10), la Classification commune des actes médicaux
344 (CCAM) [12], la Classification internationale des soins primaires,
345 deuxième édition (CISP2) [13], le Dictionnaire des résultats de
346 consultation (DRC) [14], la Classification internationale du
347 fonctionnement, du handicap et de la santé (CIF)⁴ [15], la
348 classification Anatomique, thérapeutique et chimique (ATC)
349 [16], très utilisée pour les médicaments. Au sein de cette
350 structure, il a fallu respecter les spécificités particulières et la
351 structure d'origine de chaque terminologie, et assurer par la
352 suite, les relations inter et intraterminologiques garantissant la
353 navigation entre les terminologies.

Recherche d'information multiterminologique en 2009

354 Disposant de cette structure multiterminologique, la *back-*
355 *office* (base de données) de CISMeF a été mis à jour à son
356 tour pour y inclure les nouvelles techniques et technologiques
357 de l'univers multiterminologique. Par ailleurs, le moteur de
358 recherche DocCISMeF a été modifié permettant une recherche
359 plus large sur toutes les terminologies médicales disponibles
360 dans la nouvelle structure.

361 Désormais, l'algorithme de recherche est décrit en :

- 362 • Étape 1 : rechercher les termes représentatifs de la requête
363 dans le titre de la ressource et au niveau des termes
364 d'indexation. Cependant, respectant la stratégie éditoriale de
365 l'équipe CISMeF, une condition *sine qua non* est de privilégier
366 l'indexation MeSH par rapport aux autres terminologies
367 médicales. Par exemple, soit la requête de l'utilisateur
368 « endoscopie du système nerveux » ; même si, dans un
369 premier temps, le processus de recherche trouve un terme
370 d'indexation CCAM avec la représentation la plus fine de la
371 requête (soit « endoscopie du système nerveux »), il poursuit
372 l'appariement avec des représentations moins fines (soit par
373 exemple « endoscopie » et « système nerveux »). Ainsi, pour
374 cet exemple, le résultat est non seulement les ressources
375 indexées par le terme CCAM *endoscopie du système nerveux*,
376 (étant donné que ce dernier est un terme d'indexation CCAM)
377 mais aussi les ressources indexées par les descripteurs MeSH
378
379

² Terme de science de l'information qui indique que plus le rappel est élevé, plus le taux de faux-négatifs est faible

³ Interopérabilité sémantique des terminologies dans les systèmes d'informations de santé français. URL : <http://www.interstis.org>.

⁴ Un mappage partiel manuel a été réalisé entre la CIF et le MeSH lors du projet PIH (Portail d'information sur le handicap)

380 endoscopie et système nerveux (étant donné que n'a pas un
381 descripteur MeSH *endoscopie du système nerveux* mais plutôt
382 deux descripteurs MeSH *endoscopie* et *système nerveux* ;
383 • Étape 2 : rechercher tous les termes représentatifs de la
384 requête (des plus fines au moins fines) dans toutes les
385 métadonnées de la ressource avec une mesure d'adjacence
386 entre les mots ;
387 • Étape 3 : rechercher tous les termes représentatifs de la
388 requête (des plus fines au moins fines) en plein texte avec une
389 mesure d'adjacence entre les mots.
390 Par ailleurs, grâce à l'option du filtrage, un utilisateur a toujours
391 la possibilité de restreindre sa recherche sur une seule ou
392 quelques terminologies médicales selon son contexte d'utilisa-
393 tion et ses besoins en information. Ainsi, dans le précédent
394 exemple, si l'utilisateur limite sa recherche à la CCAM, par
395 exemple, il n'a que les ressources indexées par le terme
396 *endoscopie du système nerveux*.

397 Exemple de recherche d'information 398 multiterminologique

399 Un exemple déjà opérationnel en recherche documentaire
400 multiterminologique est la réalisation d'un PIM [17], en colla-
401 boration avec la société Vidal, dans le cadre du projet Européen
402 *Patient safety through intelligent procedures in medication*
403 (PSIP) [18]. Ce portail représente une sous-partie du portail
404 CISMeF avec des fonctionnalités spécifiques restreintes aux
405 informations médicamenteuses. Le PIM sera intégré prochaine-
406 ment aux sites Web du Vidal. Il ne sera donc accessible qu'aux
407 clients du Vidal. Ce portail bilingue (français/anglais) a été créé
408 afin de faciliter l'accès à l'information concernant le médica-
409 ment. Il est destiné non seulement aux professionnels de santé,
410 mais aussi au grand public.
411 Ce portail permet un accès par nom commercial du médicament
412 ou encore par sa dénomination commune internationale (DCI).
413 L'information médicamenteuse peut être indexée par la clas-
414 sification internationale ATC [1] et représentée à travers les
415 termes MeSH (y compris les *supplementary concepts*), les
416 différents codes sur les médicaments utilisés en France [Code
417 d'identification de la présentation (CIP) [19], Code d'identifica-
418 tion de la spécialité (CIS) [20], Unités communes de dispensa-
419 tion (UCD) [21]] et le code international *Chemical abstract*
420 *service* (CAS) [22]. Notre portail constitue ainsi un service
421 dédié aux utilisateurs du domaine médical pour accéder aux
422 données médicamenteuses disponibles sur l'Internet.
423 Pour des raisons pédagogiques, le résultat de l'indexation des
424 ressources par la classification ATC est affiché hiérarchiquement
425 permettant une meilleure compréhension du système de clas-
426 sification ATC et une meilleure connaissance des médicaments.
427 Ainsi, les utilisateurs, notamment les étudiants en médecine,
428 peuvent obtenir l'information concernant les caractéristiques
429 chimiques et thérapeutiques des médicaments ainsi que l'or-
430 gane ou le système sur lequel ils agissent.

Discussion

Nous avons présenté dans cet article le portail CISMeF et les principes du passage d'une recherche d'information médicale monoterminologique grâce au thésaurus MeSH vers une recherche d'information multiterminologique plus large et assez complémentaire selon les différents contextes de l'utilisateur. Par défaut, nous n'affichons pas les différentes indexations par la dizaine de terminologies décrites pour ne perdre l'utilisateur dans un portail qui est déjà décrit par certains médecins généralistes comme déjà trop complexe. Nous laisserons à l'utilisateur le choix d'affichage, à moins de créer des portails spécifiques : nous avons vu que pour le portail sur le médicament, deux terminologies sont affichées par défaut : le thésaurus MeSH et la classification ATC. Chaque ressource indexée dans CISMeF possède de plus une notice détaillée : l'utilisateur avancé y trouvera l'indexation par chaque terminologie.

L'application de cet univers multiterminologique devrait être applicable hors des portails d'accès à l'information sur l'Internet, en particulier sur les DEP. Dans ce cas, l'objectif serait de mieux suivre les patients et bien cerner l'information requise lors des consultations et hospitalisations. Ainsi, un exemple de requête possible au sein du DEP peut être sous cette forme : *infarctus du myocarde [CIM 10] [manuel] ET bilirubine > 20 ET (même hospitalisation – même identifiant d'hospitalisation-)*. Cette requête se traduit en langage naturel par : rechercher tous les patients qui ont un infarctus du myocarde codé manuellement en CIM-10 et dont un examen biologique de bilirubine est supérieur à 20 pendant la même hospitalisation. Cet exemple illustre l'intérêt de notre travail pour améliorer l'inclusion de patients dans des essais cliniques. En collaboration avec la société Oracle dans le projet PSIP, nous étudions l'intérêt d'intégrer de la sémantique lors de l'interrogation du catalogue afin d'obtenir des réponses plus pertinentes selon le contexte de l'utilisateur.

Plusieurs nouvelles terminologies, classifications, thésaurus et nomenclatures médicaux seront intégrées au *backoffice* de CISMeF, pour permettre une recherche d'information encore plus efficace en termes de rappel :

- Le thésaurus *Orphanet* sur les maladies rares, en collaboration avec le service commun n°11 de l'INSERM (équipe *Orphanet*) ;
- *LOINC* pour les examens biologiques, en collaboration avec la direction du système de l'information de l'AP-HP ;
- *MedDRA* et *WHO-Adverse reaction terminology* (ART) pour les effets secondaires sur le médicament, en collaboration avec le Laboratoire d'enseignement et de recherche sur le traitement de l'information médicale (LERTIM) de Marseille.

Conflits d'intérêts : aucun.

Remerciements : Depuis 2005, ce travail a été en partie financé par les projets suivants : VODEL (ANR-05-RNTL-041-204), PIH (ANR-05-RNTS-00-601), PSIP (Grant agreement n°216130) InterSTIS (ANR-07-TECSAN-010-02) et L3IM (ANR-08-TECS-007-02), ALADIN (ANR-08-TECS-001-02).

Références

- [1] Keselman A, Browne AC, Kaufman DR. Consumer health information seeking as a hypothesis testing. *J Am Med Inform Assoc* 2008;15:484-95.
- [2] Jansen BJ, Spink A. How are we searching the World Wide Web? A comparison of nine search engine transaction logs. *Information Processing and Management* 2006;42:248-63.
- [3] 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.
- [4] Koch T. Quality-Controlled subject gateways: definitions, typologies, empirical overview, Subject gateways. *Online Information Review* 2000; 24:24-34.
- [5] 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 avril 2009].
- [6] Gehanno JF, Thirion B, Darmoni SJ. Evaluation of meta-concepts for information retrieval in a quality-controlled Health Gateway. *AMIA Annu Symp Proc* 2007; 269-73.
- [7] Névéol A, Soualmia LF, Douyère M, Rogozan A, Thirion B, Darmoni SJ. Using CISMeF MeSH "Encapsulated" terminology and a categorization algorithm for health resources. *Int J Med Inform* 2004;73(1):57-64.
- [8] Humphrey SM, Névéol A, Gobeil J, Ruch P, Darmoni SJ, Browne A. Comparing a Rule Based vs. Statistical System for Automatic Categorization of MEDLINE Documents According to Biomedical Specialty. *J Am Soc Inform Sci Technol* 2009 (accepted).
- [9] Soualmia L, Dahamna B, Thirion B, Darmoni SJ. Strategies for health information retrieval. *Stud Health Technol Inform* 2006;124:595-600.
- [10] Unified Medical Language System: URL : <http://www.nlm.nih.gov/research/umls/>. [Date de consultation Avril 2009].
- [11] Cornet R, de Keizer N. Forty years of SNOMED: a literature review. *BMC Med Inform Decis Mak* 2008;8(Suppl 1):S2.
- [12] Zais A, Hanser S. The french common classification of procedures CCAM. An option for Germany. *Bundesgesundheitsblatt Gesundheitsforschung* 2007;50:944-52.
- [13] Soler JK, Okkes I, Wood M, Lamberts H. The coming of age of ICPC: celebrating the 21st birthday of the International Classification of Primary Care. *Fam Pract* 2008;25:312-7.
- [14] Morel M. Pourquoi un dictionnaire des résultats de consultation en médecine générale ? *La Revue du praticien. Médecine générale* 1996;118(311) (83-36).
- [15] Baron S, Linden M. The role of the 'International Classification of Functioning, Disability and Health. ICF' in the description and classification of mental disorders. *Eur Arch Psychiatry Clin Neurosci* 2008;258(Suppl 5):81-5.
- [16] Skrbo A, Begovic 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.
- [17] Letord C, Sakji S, Pereira S, Dahamna B, Kergoulay I, Darmoni SJ. Recherche d'information multi-terminologique: application à un portail d'information sur le médicament en Europe. *IRBM* 2008;29: 350-6.
- [18] PSIP (Patient safety through intelligent procedures in medication). URL : www.psip-project.eu. [date de consultation avril 2009].
- [19] 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>. [Date de consultation Avril 2009].
- [20] Health Administration systems. URL : <http://www.has-sante.fr/portail/jcms/c71889/de-la-certification-des-lap>. [Date de consultation avril 2009].
- [21] L'assurance maladie en ligne URL: <http://www.ameli.fr/index.php>.
- [22] Dittmar PG, Stobaugh RE, Watson CE. The chemical abstracts service chemical registry system I. *General Design. J Chem Inf Comput Sci* 1976;16:111.