CISMeF Catalogue et Index des Sites Médicaux de langue Française

Terminologies et données massives

Stefan J. Darmoni, MD, PhD

Service d'Informatique Biomédicale (SIBM), CHU de Rouen &
TIBS, LITIS EA4108, Université de Normandie, France
LIMICS, INSERM U1142

Email: Stefan.Darmoni@chu-rouen.fr

Données massives en santé

- Les données massives en santé existent :

  - au sein du dossier du patient informatisé
    - ex : 8 millions de CR depuis 1984 au CHU de Rouen, plusieurs centaines de millions d'examens biologiques

  - données agrégées
    - SNIIRAM +++ objet de nombreuses convoitises
    - projet TOLBIAC : matrice de cooccurrences

  - Internet des objets
    - 10^{21} en 2020 (à vérifier !!!) ; un zettaoctet de données de santé (bien-être)

  - **terminologies, ontologies, classification**
    - **millions de concepts, 100 millions de relations**

Terminologies de santé

- Dictionnaire : pas de hiérarchie ; renvoi entre terme
- Classification monoaxiale : CIM10
- Thésaurus multiaxial : MeSH
- Ontologie formelle, avec inférence possible (raisonnement) : FMA (anatomie)

- La santé est la discipline où coexistent de très nombreuses terminologies et ontologie (T/O), avec le droit
- Très peu de chose en informatique => absence de base de données bibliographiques comme MEDLINE/PubMed, fondé sur un thésaurus

Trois grands serveurs de terminologies en santé

- UMLS
  - NIH, Bethesda (USA)
  - Plus de 150 T/O
  - Essentiellement en anglais
  - La référence internationale pour la diffusion, mais pas pour la consultation
- BioPortal*
  - NCBO, Stanford (USA)
  - Plus de 400 T/0 (beaucoup en biologie, avec peu de concepts)
  - Essentiellement en anglais (silo pour les autres langues)
  - La référence pour poster une ontologie
- HeTOP*
  - SIBM, Rouen (France)
  - 68 T/0 en 23 langues
  - La référence inter-lingue (navigation entre les langues) et dans le monde francophone

*_Grosjean J et coll._ An Approach to Compare Bio-Ontologies Portals. **Stud Health Technol Inform**, 2014;205:1008-1012.

HeTOP : principaux chiffres

Mai 2010

Terminologies & ontologies	Concepts	Synonymes	Définitions	Relations & hiérarchies
25	> 580 000	> 840 000	> 220 000	> 1 200 000

Mai 2011

Terminologies	Concepts	Synonymes	Définitions	Relations
32	> 980 000	> 2 300 000	> 220 000	> 4 000 000

Avril 2013

Terminologies	Concepts	Synonymes	Définitions	Relations
45	≈ 1 620 000	≈ 3 700 000	≈ 220 000	≈ 5 500 000

Octobre 2015

Terminologies	Concepts in English	Concepts in French	Synonyms	Definitions	Relations
68 (17 UMLS)	1,743,772	1,031,230	8,611,170	278,687	9,862,198

Chiffres importants

Utilisateurs inscrits (obligatoire pour CISP2) +++ À discuter +++	> 2 200
trafic	15 000 hits/jours (600 utilisateurs par jour ouvré)

Terminologies en français dans l'UMLS (2015AA)

Source vocabulary (Translator) [N lang.]	Number of Strings (PT + syn. + acro.)	Number of CUIs (% Translation in French)
MeSH Fr* [14]	105,758	41,229
MedDRA Fr [9]	73,860	73,608
WHO-ART Fr [5]	3,631	3,091
MTHMST Fr	1,833	1,636
ICPC2 [19]	**702**	**722 (100%)**
*MeSH Fr HeTOP		
Descriptors (INSERM) [16]	105,274	27,324 (100%)
Supplementary Concepts (CISMeF) [2]	58,514	43,177 (18.84%)
Concepts (INSERM & CISMeF) [2]	99,184	93,622 (26.52%)

Terminologies en français mais pas dans l'UMLS (2015AA)

Source vocabulary (Translator) [N lang.]	Number of Strings	Number of CUIs (% of translation Fr)
ICD10 Fr (WHO) [12]	26,337	12,143 (100%)
ICD10 PCS (CISMeF) [2]	7,297	7,297 (5%)
ICD9 Fr (WHO) [1]	10,716	7,356 (100%)
ICDO Fr (WHO+CISMeF) [3]	1,462	1,362 (100%)
ICF (WHO) [2]	1,496	1,495
FMA Fr (U. of Washington)	4,564	4,452
FMA Fr in HeTOP (CISMeF) [7]	15,923	15,865 (19.58%)
ICNP [2]	2,811	1,158 (92%)
SNOMED Int. [2]	139,792	96,756 (94.51%)
ATC (WHO) [3]	5,834	5,757 (100%)

Terminologies en français mais pas dans l'UMLS (2015AA)

Source vocabulary (Translator) [N lang.]	Number of Strings Fr/En/% of tranlsations	Number of CUIs via CISMeF mappings
WHO-ICPS [2]	424	105
RADLEX (CISMeF) [2]	7,633/42,313/18.04	240
LOINC (APHP & SFIL) [2]	58,950	58,500 (60.71%)
MEDLINEplus Fr (CISMeF & LIMSI) [2]	849	846 (100%)
NCIT Fr (CISMeF) [2]	56,032	50,938 (54,23%)

Terminologies en français mais pas dans l'UMLS (2015AA)

Source vocabulary (Translator) [N lang.]	Number of Strings	Number of CUIs
OMIM Fr (CISMeF) [2]	7,770	6,668 (88.66%)
HRDO (Orphanet) [2]	13,535/id/100	4,943
HPO (CISMeF) [2]	11,100/11,908/93.21	1,541
CCAM (procedure) [1]	10,121	0
BNPC (toxicology) [2]	91,751	11,539
Q-Codes	**184**	
… including interface terminologies	in biology and imaging	

Overall, number of distinct CUI with at least one French translation in HeTOP = 334,935 vs. 85,685 in UMLS

108 millions of RDF triplets (health big data) in 2014

Autres outils intégrés à HeTOP

- ECMT Extracteur de Concepts Multi Terminologiques
  - Capable d'extraire les concepts médicaux d'un CR en ½ seconde (NoSQL)
  - Valorisation Alicante
  - Utilisation en pratique courante au GHICL (Catho de Lille) ; DIM : Arnaud Hansske
  - Près d'un million de CR indexés avec ECMT
- InfoRoute, un InfoButton à la française
  - URL: inforoute.chu-rouen.fr
  - Accès contextualisé à la connaissance (expansion sémantique, fondé sur les alignements entre T/O)
- MT@HeTOP, outil d'alignement et de traduction automatique
- Moteur de recherche sémantique générique
  - Doc'CISMeF (URL: doccismef.chu-rouen.fr) sur la littérature grise en santé sur l'Internet (10^5 ressources)
  - LISSA (URL : www.lissa.fr), un PubMed à la française (0,7 x 10^6 citations d'articles), dont Exercer
  - RIDOPI, moteur de recherche dans le DPI (8 x 10^6 CR ; 10^8 données numériques à Rouen)

Accueil Se déconnecter **Projets SIFADO, TerSan et RAVEL**

Recherche multi-patients | Dossiers patients consultables | Aide

Recherche simple

OK

Recherche avancée

Exemples
Exemples multi-patients

Entité | Libellé

Utilisez Ctrl-Espace pour voir les propositions de mots réservés

```
1  analyse(codeEXEResultatBiologique(label="Sodium") AND valeurNumer
   iqueAnalyse<borneInfAnalyse AND patient(id="DM_PAT_125"))
```

analyse(codeEXEResultatBiologique(label="Sodium") AND valeurNumeriqueAnalyse<borneInfAnalyse AND patient(id="DM_PAT_125"))

Items per page: 20 << < Page: 1 / 2 > >> Filtre

Type d'examen biologique	Date d'examen	Résultat	Patient
Sodium	2009/10/13	131 mmol/l [135 - 145]	125
Sodium	2007/09/08	120 mmol/l [135 - 145]	125
Sodium	2009/08/08	133 mmol/l [135 - 145]	125
Sodium	2009/10/30	134 mmol/l [135 - 145]	125
Sodium	2009/03/04	134 mmol/l [135 - 145]	125
Sodium	2009/08/07	132 mmol/l [135 - 145]	125
Sodium	2009/10/19	130 mmol/l [135 - 145]	125
Sodium	2009/02/26	134 mmol/l [135 - 145]	125
Sodium	2007/09/11	127 mmol/l [135 - 145]	125
Sodium	2009/10/20	132 mmol/l [135 - 145]	125
Sodium	2009/08/11	131 mmol/l [135 - 145]	125
Sodium	2007/09/10	125 mmol/l [135 - 145]	125
Sodium	2007/09/10	125 mmol/l [135 - 145]	125
Sodium	2007/09/14	130 mmol/l [135 - 145]	125
Sodium	2009/08/18	134 mmol/l [135 - 145]	125
Sodium	2009/09/07	131 mmol/l [135 - 145]	125

Accueil Se déconnecter **Projets SIFADO, TerSan et RAVEL**

Recherche multi-patients | Dossiers patients consultables | Aide

Recherche simple

infarctus du myocarde OK

Infarctus du myocarde

Recherche avancée

Exemples

Exemples multi-patients

Entité Libellé

Utilisez Ctrl-Espace pour voir les propositions de mots réservés

```
patient(id="DM_PAT_1022")) OR compteRendu(T_DESC_MESH_DESCRIPTEUR
(id="MSH_D_000789") AND patient(id="DM_PAT_1022")) OR compteRendu
(T_DESC_MESH_DESCRIPTEUR(id="MSH_D_006342") AND patient(id
="DM_PAT_1022")) OR compteRendu(T_DESC_MESH_DESCRIPTEUR(id
="MSH_D_009203") AND patient(id="DM_PAT_1022")) OR compteRendu
(T_DESC_MESH_DESCRIPTEUR(id="MSH_D_056989") AND patient(id
="DM_PAT_1022")) OR compteRendu(T_DESC_NCIT_CODE(id="NCI_CO_C2799
6") AND patient(id="DM_PAT_1022")) OR compteRendu(T_DESC_NCIT_COD
E(id="NCI_CO_C71066") AND patient(id="DM_PAT_1022")) OR compteRen
du(T_DESC_NCIT_CODE(id="NCI_CO_C35519") AND patient(id="DM_PAT_10
22"))
```

OK

Aide du portail terminologique de santé
URL : www.hetop.eu

4 entrées trouvées (1,67 s)

Items per page: 20 << < Page: 1 / 1 > >> Filtre

Unité médicale	Date	CR	Patient
Compte-rendu de séjour (ab)	2008-07-29 00:00:00	CR	1022
Compte-rendu de séjour (ab)	2008-07-18 00:00:00	CR	1022
Compte-rendu de séjour (ab)	2008-07-04 00:00:00	CR	1022
Compte-rendu de séjour (ab)	2008-06-26 00:00:00	CR	1022