

Santé Publique – Santé Numérique

Systemes d'Organisation des Connaissances (SOC) en Santé



Julien Grosjean, PhD

Kévin Billey, MSc

Stéfan Darmoni, MD, PhD

Lina Soualmia, PhD

Florent Desgrippes (APHP)

Faculté de Médecine, Université de Rouen et APHP

Courriel : Stefan.Darmoni@chu-rouen.fr



Normandie Université



Les Systèmes d'Organisation des Connaissances (SOC)

- Les terminologies
- Notion de "concept terminologique"
→ connaissance~intelligence (inférence, subsumption, lexiques, etc.)
- Les classifications ⇔ vocabulaires contrôlés (ou lexiques) structurés
- Depuis l'antiquité, évolution des classifications en même temps que la connaissance
- Fort intérêt depuis la fin du XVIIIème siècle avec la classification scientifique des espèces puis grand développement jusqu'à nos jours

Ontologie

Terminologie

Thesaurus

Nomenclature

Taxinomie

Dictionnaire

Lexique

Liste

↑ complexité

Terminologie

- “Terminologie” = tout ensemble de termes consacrés

- Les Terminologies ont un rapport étroit:

- à la “connaissance” / aux croyances
- aux “standards” et aux “normes”
- à l’éducation
- à la souveraineté / l’hégémonie
 - Scientifique, technique, économique
 - Logistique
 - Culturelle
- à la propriété intellectuelle
- à la confiance
- au pouvoir en général

⇒ sujet quasi-philosophique

⇒ sujet sensible (“knowledge is power” = “soft-power”)

⇒ lien fort avec l’OpenGov (données ouvertes portées par le gouvernement français)

La santé est le plus riche domaine en terme d'existence et de développement des SOC

- A l'inverse, les sciences de l'ingénieur ou l'informatique sont très faiblement représentées dans les SOC... de même pour les bases de données bibliographiques.
- En santé, 200 SOC dans UMLS (méta-base états-unienne), 500 dans BioPortal (portail « ouvert » états-unien), 90 dans HeTOP (portail « fermé » français, en 52 langues = interlingue)
- Un SOC pour chaque domaine : CIM-10 pour les maladies, MeSH pour la documentation scientifique, FMA pour l'anatomie, LPP & Cladimed pour les dispositifs médicaux, NCIT pour le cancer, LOINC pour la biologie... et SNOMED qui aspire à être la SOC pivot
- Mais les SOC se recoupent... beaucoup de concepts en commun (attention aux subtilités!)

Langage documentaire

- Langage artificiel, constitué de notions et de relations entre les notions
- Objectif : dans un système documentaire, formaliser les données dans les documents et dans les demandes des utilisateurs (AFNOR)
- Les langages documentaires se répartissent en deux grandes familles :
 - les langages à structure hiérarchique (classifications) utilisant des indices symboliques,
 - et les langages à structure combinatoire (thésaurus, listes d'autorité), utilisant les mots du langage naturel

Classification vs. thésaurus

- Deux principales méthodes pour réaliser une indexation
 - Méthode synthétique : du général au spécifique
 - Méthode analytique : décomposition des concepts et combinaison entre eux
- Deux familles de langages documentaires
 - Méthode synthétique ➡ indexation systématique ➡ langage avec structure hiérarchique = **classification**
 - Méthode analytique ➡ indexation analytique ou indexation alphabétique ➡ langage avec structure analytique ou structure combinatoire = **thesaurus**

Precoordination & Postcoordination

- Contradictory organization and use of documentary languages
- In precoordinated languages (classifications)
 - Terms of indexing are mainly words or composed indices (e.g. 121.2.1) covering the entire notion
 - Coordination between the concepts are performed when indexing
- In postcoordinated languages (thesaurus)
 - Notions are defined by the most simple constituents
 - Several descriptors are necessary to describe the entire notion
 - Coordination between the concepts are performed when information retrieval using query operators (e.g. Boolean)

Classification

- La connaissance est séparée en petites unités... de connaissances
 - Séparée en sous-domaine, puis sous-sous-domaine... pour aboutir à la granularité la plus fine (la « feuille » de l'arborescence)
- Langage avec structure hiérarchique
 - CIM10/ICD10 = structure monoaxiale
 - Langage avec des codes (ex : J45 pour l'asthme)
- Possibilité de grouper les codes par thèmes

Caractéristiques des classifications

- Le lien sémantique qui préside à la catégorisation peut être un lien d'affectation hiérarchique (type *est-un*, par exemple, une rubéole est une maladie infectieuse) ou de partition (type *fait-partie-de*, par exemple, le genou fait partie du membre inférieur)
- Une classification monoaxiale répartit en plusieurs classes disjointes l'ensemble des objets et revient à construire une hiérarchie de classes à partir d'une racine unique et commune. Les classes d'un niveau doivent couvrir l'ensemble du domaine de ce niveau (exhaustivité) sans se recouvrir (exclusivité) afin qu'un objet trouve une place et une seule.

Caractéristiques des classifications

- Très difficile en pratique de répartir les objets du domaine selon un seul critère, ce qui a entraîné le développement de classifications multiaxiales.
- Les classifications multiaxiales ou à facettes sont modulaires et combinent des termes appartenant à des systèmes différents, eux-mêmes organisés de façon hiérarchique
- Ainsi, une sémantique de juxtaposition s'ajoute à la sémantique de catégorisation.

Caractéristiques des classifications

- Cependant, même la juxtaposition de termes ne suffit pas à exprimer la diversité des liens sémantiques existant entre les termes médicaux (relation de causalité, d'association, d'évolution, etc...).
 - Aussi ces liens sont-ils explicités par des connecteurs (équivalents des groupes verbaux : est un, a, traite, est traité par, cause, est causé par, etc...) dans certaines classifications ou dans les langages artificiels.
- En outre, à côté des termes (les mots) et des connecteurs (les verbes), certains systèmes ajoutent des modificateurs, faisant fonction d'adjectif, venant préciser le degré de certitude, l'évolutivité ou l'intensité.

Definitions of ontology

- Philosophy

- Part of metaphysics, which applies to the nature of being, becoming, existence, or reality, as well as the basic categories of being and their relations. Traditionally listed as a part of the major branch of philosophy known as metaphysics, ontology often deals with questions concerning what entities exist or may be said to exist, and how such entities may be grouped, related within a hierarchy, and subdivided according to similarities and differences.

- Computer Science & Knowledge Engineering

- “An ontology is a shared specification of a conceptualization” (by Tom Gruber, 1990)
- Formal naming and definition of the types, properties, and interrelationships of the entities that really or fundamentally exist for a particular domain of discourse. It is thus a practical application of philosophical ontology, with a taxonomy.

Ontologie

- Ontologie – une description formelle des concepts, rôles et relations qui existent pour un agent ou une communauté d'agents
- Les ontologies fournissent une compréhension commune d'un domaine qui peut être communiquée, elles jouent un rôle majeur dans les échanges d'information
- $\exists$ plusieurs ontologies formelles en santé
 - FMA en anatomie
 - HPO pour les maladies rares
- $\exists$ aussi des quasi-ontologies en santé
 - SNOMED CT
 - NCIT pour la cancérologie

Quelques exemples de SOC en santé

Classification Internationale des Maladies

CIM (ICD)

- La Classification Internationale des Maladies (CIM ou ICD en anglais) de l'Organisation Mondiale de la Santé (OMS ou WHO) a été originellement conçue pour coder les causes de décès dans une optique épidémiologique.
- Elle est maintenant également utilisée pour l'évaluation des soins médicaux et l'indexation des dossiers...
- Mais surtout la CIM sert à coder les diagnostics lors des hospitalisations (et consultations) => PMSI => budget de l'hôpital
- Dans sa version 10 (depuis 1994), classification monoaxiale avec
 - 21 chapitres principaux, dont
 - 17 concernent des maladies et
 - quatre concernent les signes et résultats anormaux, les causes de traumatismes, d'empoisonnement ou de morbidité, l'état de santé et les facteurs de recours aux soins.
- Traduite en 43 langues
- Version 11 in 2022 ???
- CIM 9, 10 & 11 sont intégrées au sein d'HeTOP –voir plus loin-

CIM

- Les catégories de maladies sont définies en fonction d'un caractère commun qui peut être l'étiologie (1 = Maladies infectieuses, lettres A et B), la topographie (9 = maladies de l'appareil circulatoire, lettre I), la physiologie (15 = Grossesse et accouchement, lettre O) ou la pathologie (II = Tumeurs).
- La classification aboutit par subdivisions successives à un code à 3 caractères (une lettre correspondant au chapitre puis 2 chiffres) pour les maladies définies à un niveau général, décliné par l'ajout d'un quatrième chiffre (derrière un point) pour désigner les diagnostics précis et les formes cliniques ; le sous-code 9 désignant l'absence de précision (SAI = sans autre indication) et le sous-code 8 les autres formes non précédemment définies.
- Dans certains cas, un cinquième chiffre a été rajouté afin d'améliorer la finesse de description. La CIM-10 a introduit la notion de troubles iatrogènes. Elle compte au total 80.000 termes.

CIM

- La classification de l'OMS sert en France :
 - au codage des causes de décès (codage des diagnostics inscrits sur le certificat de décès)
 - au regroupement des séjours hospitaliers en groupes homogènes de malades –GHM- (programme de médicalisation du système d'information ou PMSI) dont le but est l'analyse médico-économique de l'activité hospitalière.

CIM

- ICD-10 top tree
 - Diseases of the respiratory system
 - Chronic lower respiratory diseases
 - Asthma
 - Bronchiectasis
 - Bronchitis, not specified as acute or chronic
 - Emphysema
 - Centrilobular emphysema
 - Emphysema, unspecified
 - MacLeods syndrome
 - Other emphysema
 - Panlobular emphysema
 - Other chronic obstructive pulmonary disease
 - Chronic obstructive pulmonary disease with acute exacerbation, unspecified
 - Chronic obstructive pulmonary disease with acute lower respiratory infection
 - Chronic obstructive pulmonary disease, unspecified
 - Other specified chronic obstructive pulmonary disease
 - Simple and mucopurulent chronic bronchitis
 - Mixed simple and mucopurulent chronic bronchitis
 - Mucopurulent chronic bronchitis
 - Simple chronic bronchitis
 - Status asthmaticus
 - Unspecified chronic bronchitis

CIM version 11

- Nouvelle classification
- Multi axiale, comme le MeSH, SNOMED ou NCIT
- Plus de 59 000 concepts
- Intégré dans HeTOP, traduction automatique fournie par l'ANS (URL : <https://www.hetop.eu/hetop/rep/fr/CIM-11/>)
- Devrait être opérationnel en janvier 2022... déjà utilisé dans certains pays du tiers monde ;
 - la traduction en espagnol et en mandarin quasi terminée ;
 - Celle en français est pour l'instant automatique

CCAM

- Les classifications des actes et des procédures sont plus variables car elles répondent à différents objectifs
- Le catalogue des actes médicaux par nature (CDAM puis désormais CCAM) a été préparé dans le cadre du PMSI afin de permettre le codage des actes effectués au cours d'un séjour hospitalier
- En 2021, les classifications des actes et procédures sont nationales, donc différentes d'un pays à l'autre
 - CCAM en France, CPT aux EtatsUnis...
- Tout comme la CIM pour les maladies, l'OMS travaille sur une terminologie internationale sur les interventions de santé
 - URL : <https://www.who.int/standards/classifications/international-classification-of-health-interventions>
 - ICHI a été intégré à HeTOP en 2021 (URL : <https://www.hetop.eu/hetop/rep/fr/ICHI/>)

Structure de SOC : exemple du MeSH

- Each descriptor is linked to other via several relations
 - Hierarchy
 - IS A
 - PART OF
 - Merge of these two relations => BTNT & NTBT ; wrong for ontologies with reasoning capabilities
- Most used thesaurus in medicine = MeSH used to index article citations in MEDLINE/PubMed bibliographic database
 - N (MeSH Descriptors) $\simeq$ 29,000 (2021)
 - N (MeSH Supplementary Concepts) $\simeq$ 318,000 (chemistry, rare diseases)
 - N (MeSH Concepts) $\simeq$ 454,000 > N(SNOMED CT)

MeSH

- MeSH (Medical Subject Headings) sert à indexer, cataloguer et retrouver des **références d'articles scientifiques dans le domaine de la Santé.**
- Il a été conçu à la **National Library of Medicine (NLM), du National Institute of Health (NIH)*** aux Etats-Unis comme support de l'Index Medicus, répertoire des principales publications scientifiques, et est utilisé par les bases de données bibliographique Medline/PubMed (URL : www.pubmed.gov)... Mais aussi en France, CISMef (URL : www.cismef.org) ou LiSSa (URL : www.lissa.fr)
- MeSH est organisé en deux parties : une liste alphabétique de termes (lexique) et une structure multiaxiale.
 - Les 200.000 termes du lexique sont distribués selon 15 axes, allant de l'anatomie à la géographie.
 - Descripteurs MeSH, plus de 27.000 en 2021
 - Concepts Supplémentaires MeSH, plus de 248 000 en 2021, chimie & maladies rares
 - Concepts MeSH (union des deux premiers + d'autres), avec une volonté d'ontologiser le MeSH : plus de 378 000 concepts MeSH... mais pas recherchables en tant que tel dans PubMed... et recherchables dans CISMef

Griffon N, et coll. SJ. Searching for rare diseases in PubMed: a blind comparison of Orphanet expert query and query based on terminological knowledge. BMC Med Inform Decis Mak. 2016 Aug 2;16:101.

MeSH

- Les descripteurs s'organisent selon une structure hiérarchique multi axiale (un fils peut avoir plusieurs pères +++)
 - La relation hiérarchique « mélange » deux relations formelles différentes IS-A (EST UN) et PART-OF (PARTIE-DE), applicable en anatomie, avec un nombre limité : doigts PARTIE-DE mains (généralement 5... mais pas toujours)
 - Relation hiérarchique aboutit à BTNT (broader than-narrower than) et sa relation inverse (NTBT)
- Des connecteurs permettent des références explicites entre termes et expriment les relations de synonymie, de voisinage ou d'association tandis que des qualificatifs permettent de considérer les différentes facettes d'un concept (par exemple : tuberculose/traitement).
- Ce système indexe actuellement plus de 20 millions de références de la base de données bibliographique PubMed et 1,2 de LiSSa (pour le français) et est mis à jour régulièrement pour suivre l'évolution des connaissances.

☐ No wildcard search☐ Do not search into definitions☒ Terminologies selection

428 matches in 0,27 s

Top terms

- abnormalities [MeSH Qualifier]
- pupil disorders [MeSH Descriptor]
- pupil malformations [MeSH concept]
- head and brain malformations [MedlinePlus Topic]
- Limbic malformations [HPO term]
- Abnormality of the uterus [HPO term]
- Retinal malformation [HPO term]
- vascular malformations [MeSH Descriptor]
- Malformations multiple [MedDRA LLT]
- MALFORMATIONS MULTIPLE [WHO-ART Preferred Term]

+ MeSH (161)

+ Genes & Proteins (12)

+ HPO (9)

+ HRDO (27)

+ ICD-10 WHO (150)

+ MedDRA (9)

+ MedlinePlus (2)

+ NCIt (6)

+ OMIM (47)

+ SNOMED CT (4)

+ WHO-ART (1)

Pupil disorders (MeSH Descriptor)



Description

Hierarchies

Relations

Full tree

- MeSH top tree
 - Diseases Category
 - eye diseases
 - pupil disorders
 - anisocoria
 - + miosis
 - mydriasis
 - + tonic pupil
 - nervous system diseases
 - neurologic manifestations
 - pupil disorders
 - anisocoria
 - + miosis
 - tonic pupil
 - pathological conditions, signs and symptoms
 - signs and symptoms
 - neurologic manifestations
 - pupil disorders
 - anisocoria
 - + miosis
 - tonic pupil

☐ No wildcard search☒ Terminologies selection ☐ Do not search into definitions

428 matches in 0,27 s

Top terms

- abnormalities [MeSH Qualifier]
- pupil disorders [MeSH Descriptor]
- pupil malformations [MeSH concept]
- head and brain malformations [MedlinePlus Topic]
- Limbic malformations [HPO term]
- Abnormality of the uterus [HPO term]
- Retinal malformation [HPO term]
- vascular malformations [MeSH Descriptor]
- Malformations multiple [MedDRA LLT]
- MALFORMATIONS MULTIPLE [WHO-ART Preferred Term]

+ MeSH (161)

+ Genes & Proteins (12)

+ HPO (9)

+ HRDO (27)

+ ICD-10 WHO (150)

+ MedDRA (9)

+ MedlinePlus (2)

+ NCIIt (6)

+ OMIM (47)

+ SNOMED CT (4)

+ WHO-ART (1)

Pupil disorders (MeSH Descriptor)



Description

Hierarchies

Relations

: Intra-terminologic

: Inter-terminologic

Semantic type(s) (1)

Allowable qualifiers (37)

See also (suggested by CISMef) (1)

Concepts Supplémentaires MeSH en relation (8)

Record concept(s) (16)

Metaterm(s) (2)

Curated CISMef NLP mapping (3)

UMLS correspondences (same concept) (5)

Automatic exact mappings (from CISMef team) (34)

False automatic mappings (1)

☐ No wildcard search
☐ Do not search into definitions

☒ Terminologies selection

428 matches in 0,27 s

Top terms

- abnormalities [MeSH Qualifier]
- **pupil disorders [MeSH Descriptor]**
- pupil malformations [MeSH concept]
- head and brain malformations [MedlinePlus Topic]
- Limbic malformations [HPO term]
- Abnormality of the uterus [HPO term]
- Retinal malformation [HPO term]
- vascular malformations [MeSH Descriptor]
- Malformations multiple [MedDRA LLT]
- MALFORMATIONS MULTIPLE [WHO-ART Preferred Term]

MeSH (161)

Genes & Proteins (12)

HPO (9)

HRDO (27)

ICD-10 WHO (150)

MedDRA (9)

MedlinePlus (2)

NCIt (6)

OMIM (47)

SNOMED CT (4)

WHO-ART (1)

Pupil disorders (MeSH Descriptor)

Description Hierarchies Relations

☒ Intra-terminologic ☒ Inter-terminologic

Semantic type(s) (1)

Allowable qualifiers (37)

See also (suggested by CISMef) (1)

Concepts Supplémentaires MeSH en relation (8)

congenital corneal opacities, cornea guttata, and corectopia	MeSH Supplementary Concept
pupil disorders/congenital:microcoria, congenital	MeSH Descriptor/MeSH Qualifier: Me
ectopia lentis with ectopia of pupil	MeSH Supplementary Concept
ectopia pupillae	MeSH Supplementary Concept
McPherson robertson Cammarano syndrome	MeSH Supplementary Concept
microphthalmia, isolated, with corectopia	MeSH Supplementary Concept
pierson syndrome	MeSH Supplementary Concept
ptosis, strabismus, and ectopic pupils	MeSH Supplementary Concept

Record concept(s) (16)

Metaterm(s) (2)

Currated CISMef NLP mapping (3)

UMLS correspondences (same concept) (5)

SNOMED

- SNOMED (CT) (Systematized Nomenclature of Medicine Clinical Terms) combine une quasi-ontologie de plus de 360.000 termes... mais moins que le nombre de MeSH Concepts (> 400K)
- systematically organized computer processable collection of medical terms providing codes, terms, synonyms and definitions used in clinical documentation and reporting.
- Objectif : très utile pour indexer les dossiers patients informatisés en amont (questionnaire, peu utilisé en France) ou en aval (en utilisant des outils de traitement automatique de la langue (voir ECMT plus bas))
- A “vocabulary” that enables the clear exchange of health information for all.

SNOMED

- A ce jour, 20 pays sont membres de l'organisation SNOMED international
- Historiquement, la terminologie « SNOMED International » a été créé par le Pr. Roger Côté (Canada) pour étendre les cas d'usage auparavant centrés sur l'anatomie-pathologie (SNOP Systematized Nomenclature of Pathology)

Concepts parents

Acute renal
failure
syndrome

Acute digestive
system disorder

Hepatic
failure

Attributes

liver structure

kidney structure

sudden onset
AND/OR short duration

measurement
of renal function
impaired

**Hepatorenal
syndrome**

Hepatorenal
syndrome due
to a procedure

Familial
arthrogryposis
cholestatic
hepatorenal
syndrome

Hepatorenal
syndrome as a
complication of
care

Hepatorenal
syndrome
following
delivery

Concepts enfants

ICD-11 (Foundation)

Foundation Id : <http://id.who.int/icd/entity/1015890899>

Hepatorenal syndrome

Parent(s)

- Certain specified diseases of liver
- 5-74 Other diseases of the digestive system

Synonyms

- hepatorenal failure
- HRF - [hepatorenal failure]
- hepatorenal degeneration
- hepatic renal syndrome
- hepatic renal failure
- renal hepatic failure
- urohepatic syndrome

Exclusions

- Hepatorenal syndrome following labour or delivery ⇒

Body Site

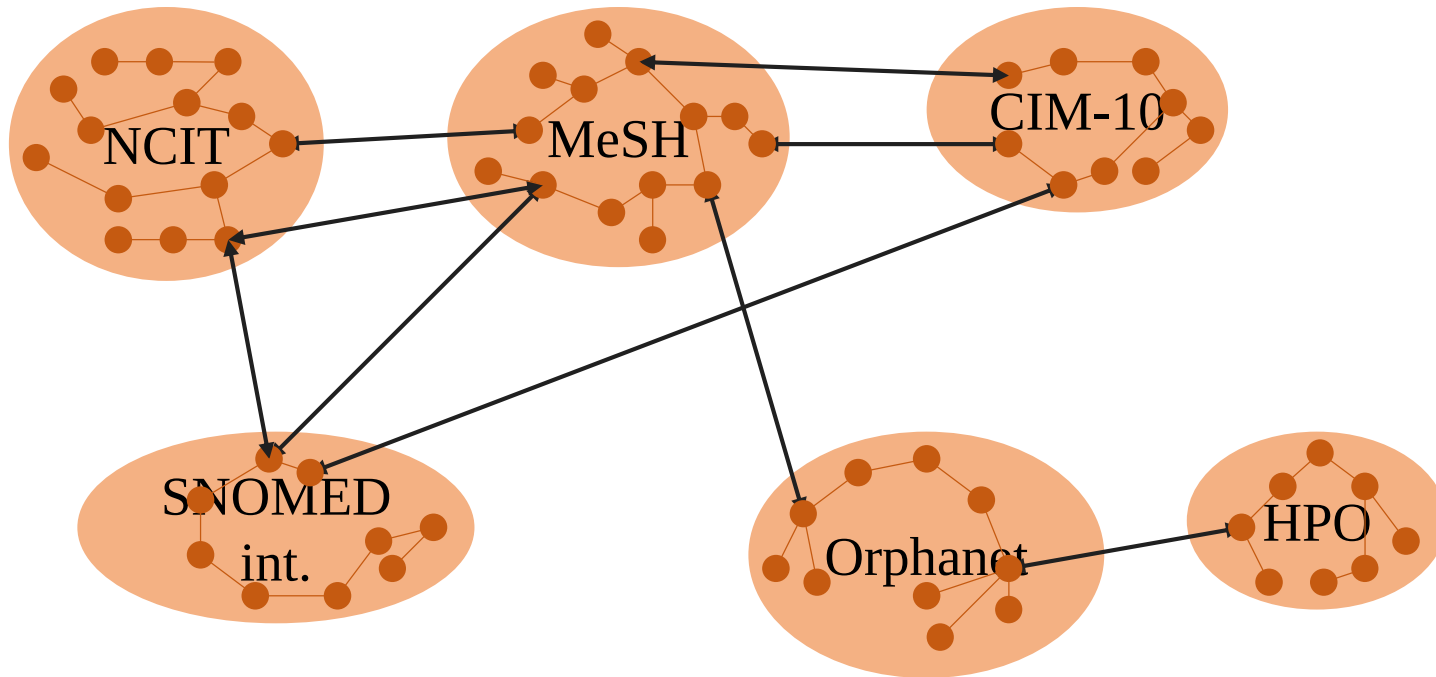
- Kidney structure (body structure)
- Entire kidney (body structure)
- Entire liver (body structure)
- Liver structure (body structure)

Construire et manipuler des SOC

Transcodage

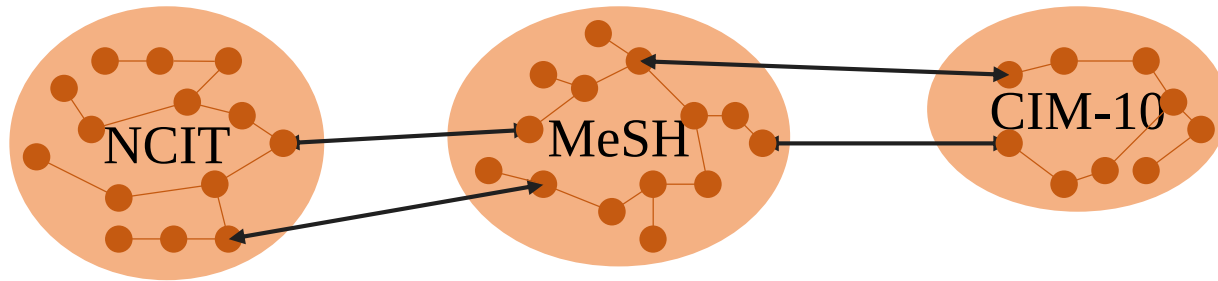
- La multiplicité des classifications répond à la diversité des informations à traduire.
- Mais parfois, pour représenter la même réalité, les acteurs de la santé utilisent différents systèmes de classification.
- Ainsi, pour indexer précisément les dossiers de ces patients en fonction des actes réalisés, un chirurgien utilisera une classification ou un système de mots clés spécifiques de sa spécialité, alors que le codage des actes dans le cadre du PMSI utilise la CCAM
- Cette multiplicité perturbe la vision de l'activité quand elle ne dégrade pas l'information, notamment lors de la communication entre systèmes différents.
- Un transcodage est une « passerelle » entre deux classifications. La qualité de ces passerelles dépend du type et de la congruence des classifications ainsi reliées. Par exemple le transcodage de CIM9 en SNOMED ne permet pas toujours une transcription explicite des concepts en raison de la différence de structuration des deux systèmes.
- Les transcodages exposent ainsi souvent à une perte d'information mais, dans certaines situations, ils sont impossibles.

Interopérabilité sémantique



https://www.hetop.eu/hetop/documentation/sem_net.htm

Alignements inter-terminologiques

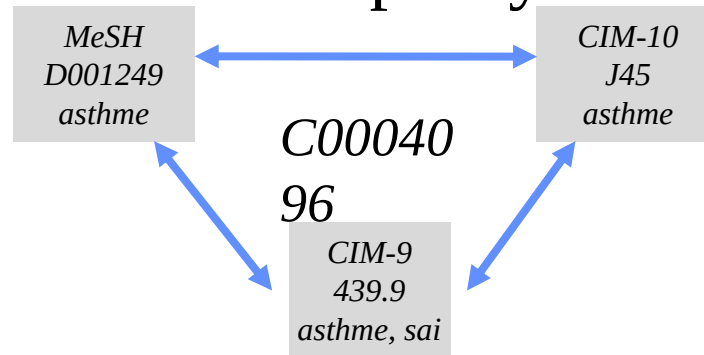


Générer automatiquement des alignements :

- par correspondance des CUI d'UMLS ;
- par correspondance de termes ;
- par fermeture transitive.

Génération automatique des alignements

- A partir des données d'UMLS :
Identifier les concepts ayant le même CUI UMLS.



Génération de 270 000 correspondances UMLS,
dont 130 000 nouveaux alignements dans HeTOP.

Génération automatique des alignements

- Par correspondance de termes :
 - Identifier des concepts ayant des termes en commun.
 - Plus de 200 000 alignements générés automatiquement.

Problème : imprécision des résultats.

- normalisation et racinisation des mots
 - ex : "avoir une réaction" et "avion à réaction"
- acronyme
 - ex : RAISIN (Réseau d'Alerte, d'Investigation et de Surveillance des Infections Nosocomiales)

Validation des alignements automatiques

Développement d'une interface dédiée aux validateurs

- Alignements exacts $A=B$
- Alignements faux $A \neq B$
- Alignements BTNT $A > B$
- Alignements NTBT $A < B$
- Relations 'Voir aussi' $A \sim B$
- Relations inconnues $A ? B$

¼ des alignements par correspondance de termes ont été validés à ce jour.

J45 asthme (Catégorie CIM-10)

Description

Hiérarchies

Relations

PubMed / Doc'CISMeF

Curation

[Nouvel alignement](#)

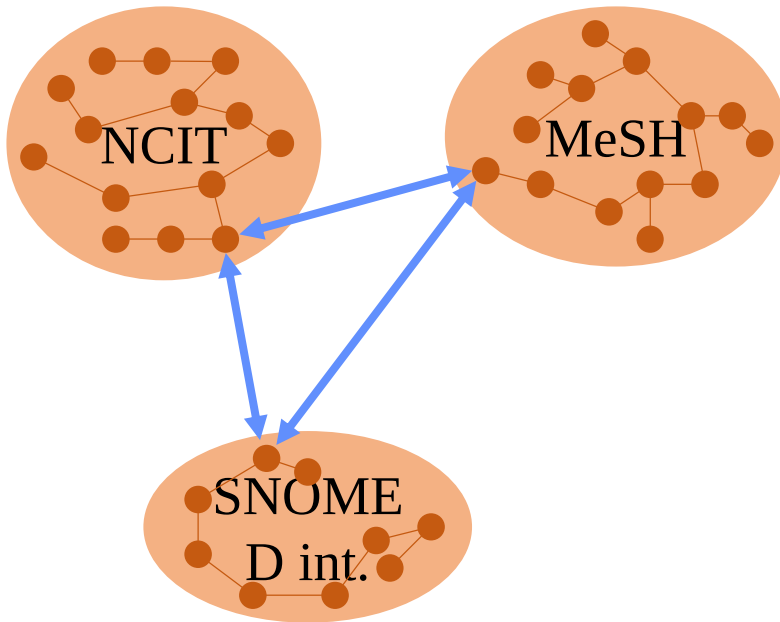
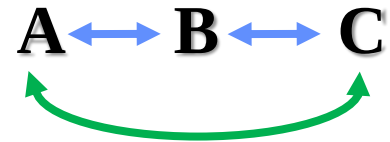
voie de l'asthme	Concept NCIt	CISMeF; 5	☒	☒	BTNT	NTBT	RT	?
Asthme bronchitique	Terme de bas niveau MedDRA	UMLS;	☒	☒	BTNT	NTBT	RT	?
Bronchite asthmatique	Terme de bas niveau MedDRA	UMLS;	☒	☒	BTNT	NTBT	RT	?
Asthmatique	Terme de bas niveau MedDRA	UMLS;	☒	☒	BTNT	NTBT	RT	?

Déjà supervisés ou ajoutés (32)

asthme*	Descripteur MeSH	☒	☒	BTNT	NTBT	RT	?
Asthme SAI	Terme de bas niveau MedDRA	☒	☒	BTNT	NTBT	RT	?
asthme	Concept TUV	☒	☒	BTNT	NTBT	RT	?
J45 - asthme	Catégorie CIM-10 (oms)	☒	☒	BTNT	NTBT	RT	?
asthme	Concept NCIt	☒	☒	BTNT	NTBT	RT	?

Utilisation de la fermeture transitive

Si A aligné avec B et B avec C
alors A aussi aligné avec C.



- Créer/valider automatiquement des alignements
- Détecter des anomalies (alignements différents ou faux)

55 000 alignements ont ainsi été créés et 68 000 validés.

Utilisation de la fermeture transitive

J45 asthme (Catégorie CIM-10)

Description

Hiérarchies

Relations

PubMed / Doc'CISMeF

Curation

Validation



2 nouveaux alignements générés

J45 asthme	Catégorie CIM-10	asthma*	Disease (Nov.)	☒	☒	BTNT	NTBT	RT	
J45 asthme	Catégorie CIM-10	J45.9 - asthme, sans précision	Sous Catégorie CIM-10 (oms)	☒	☒	BTNT	NTBT	RT	

OK

asthme*	Descripteur MeSH	☒	☒	BTNT	NTBT	RT	
Asthme SAI	Terme de bas niveau MedDRA	☒	☒	BTNT	NTBT	RT	
asthme	Concept TUV	☒	☒	BTNT	NTBT	RT	
J45 - asthme	Catégorie CIM-10 (oms)	☒	☒	BTNT	NTBT	RT	
asthme	Concept NCIt	☒	☒	BTNT	NTBT	RT	

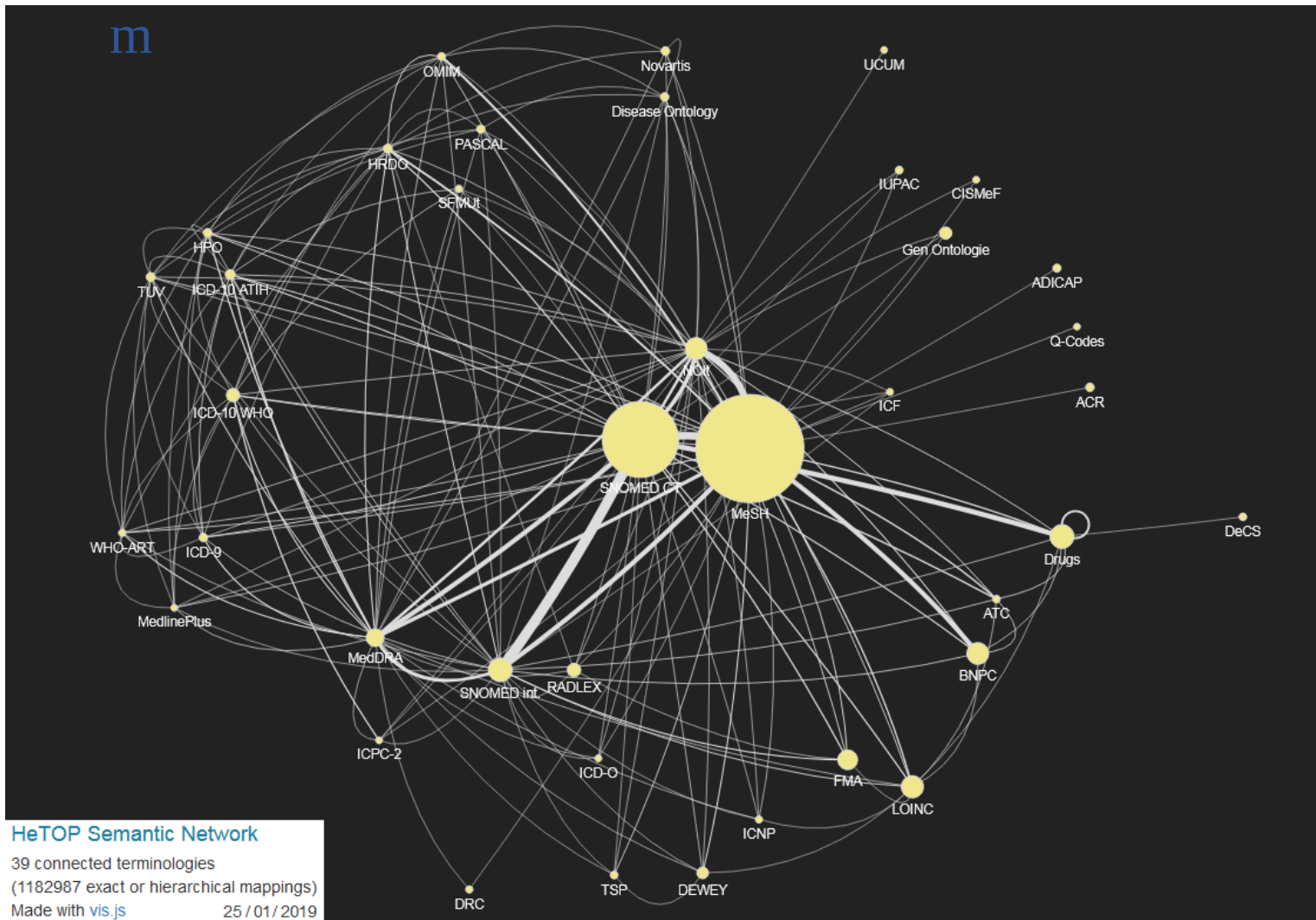
✚ Effectuer la fermeture transitive

Alignements terminologiques au sein d'HeTOP

	Source Term (Terminology)	Target Term (Terminology)	Number of relations in HeTOP
UMLS _{alignment}	<i>Myocardial Infarction</i> (MeSH)	<i>Myocardial infarction,</i> NOS (SNOMED Int)	1.15 M
CISMeF _{manual}	<i>Riedel thyroiditis</i> (HRDO)	<i>Riedel's thyroiditis</i> (MedDRA)	107 K
CISMeF _{exact}	<i>appetite stimulants</i> (ATC)	<i>Appetite stimulated</i> (WHOART)	2.2 M
CISMeF _{Supervised} 21/08/2015	<i>Gonadotropin releasing hormone</i> (MeSH)	<i>Luteotropin-releasing factor</i> (FMA)	868 K

Le réseau sémantique d'HeTOP

https://www.hetop.eu/hetop/documentation/sem_net.ht



Utilisations et exploitations de SOC

Usages et objectifs des SOC

- Apporter du sens aux données (sémantique)
- Exploiter ce sens via les lexiques et les structures informatiques
- Représenter l'information
 - Stocker
 - Décrire
 - Structurer
- Rechercher l'information
 - Indexation / annoter / coder des ressources (décrire)
- Inférer des faits
 - Règles => systèmes d'aide à la décision
 - Raisonnement
 - Subsumption (hiérarchie)
- Enseigner / apprendre (connaissances)



Cas d'usage pratique des SOC en santé

- Statistiques & épidémiologie
 - Depuis le début du XIXème siècle, statistiques de mortalité utilisant le CIM, permettant l'émergence de l'épidémiologie
- Indexation contrôlée (sciences de l'information)
 - Base de données bibliographiques +++ MEDLINE/PubMed aux EtatsUnis ; www.pubmed.gov, LiSSa en France ; www.lissa.fr
 - Littérature grise (recommandations, rapport, cours, brochures pour les patients) : CISMeF en Francophonie ; www.cismef.org
- Utilisation et réutilisation des données cliniques et omiques (clinomiques) = données de vie réelle (voir cours Entrepôts de données de santé)
 - Soins (visualisation agrégées de ces données de santé)
 - Epidémiologie
 - Recherche clinique (essais cliniques, cohorte, registre)
 - Prévention, santé publique
 - Indicateurs...

Rôles des serveurs de terminologies

- Donner accès aux principales terminologies de santé
 - Aux humains
 - Aux machines (services Web)
- Diffuser les principales terminologies de santé
 - CIM-10, CCAM
 - LOINC, ATC
 - MeSH
 - SNOMED CT

Autres systèmes académique multi-SOC en santé

- UMLS (US National Library of Medicine)

Opérationnel depuis 1986



Bodenreider, O. (2004). The Unified Medical Language System (UMLS): Integrating biomedical terminology. *Nucleic Acids Res*, 32:267–270.

- BioPortal (Stanford)



Noy, N.F et al. (2009). BioPortal: ontologies and integrated data resources at the click of a mouse. *Nucleic Acids Research; Web Server Issue 10*.

- Ontology Lookup Service (UK)



- LexGrid / NCI Term browser (USA)



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

- **Mondeca (ITM, KB), société française**

- Bioloinc
 - <https://bioloinc.fr/bioloinc/KB/>
- TerMef
 - <https://terminologie.finances.gouv.fr/index>

- **TopQuadrant (TopBraid)**

- OWL, ShaCL
- <https://www.w3.org/TR/shacl/>

- **Apelon (santé)**

- <https://www.apelon.com>

- **PoolParty**

- SKOS
- <https://www.poolparty.biz/>

- **OntoText (GraphDB)**

- <https://graphdb.ontotext.com/>

- **MultiTes, Synaptica...**

Standards

- Quel normes ? ISO 25964, ...
- Quels formats ? SKOS, OWL, RDF/XML, JSON-LD ... ?
- Quels modèles ?
- Quel formalisme ?
- Quels outils de gestion ? CTS2, ... ?
- OMOP (santé, backend)
- FHIR (santé, frontend)
- FAIR
- SQL, SparQL, GraphQL, NoSQL

SERVEURS MULTI-TERMINOLOGIQUES: RÉSUMÉ DES FONCTIONNALITÉS

Axe	Fonctionnalités
Domaine	Généraliste / thématique (santé...)
Architecture	Client/Serveur, Client léger-lourd/Serveur...
Générales	Gestion de profils utilisateurs
	CMS (édition de pages...)
	Import/Export
	Backup/Archivage
	Intégration à l'infrastructure (LDAP, CAS...)
	Règlementaire (RGPD, RGAA...)
Publication	Metadatas
	Recherche dans des termes
	Navigation dans les termes
	Visualisation arborescente
	Téléchargement
Edition/Gestion	Edition des métadonnées
	Versioning
	Edition des termes, des arborescences (CRUD)
	Traduction
	Alignements automatiques
Fonctions avancées	Workflow, Gestion de licences (ack, reporting...)
	Lien inter-terminologie
	Validation de formes (ShaCL, SHEX)
	Annotation...
API	REST, FHIR, SparQL...
Interop/Standards	OWL, CSV...

Les terminologies en santé

L'Agence du Numérique en Santé (**ANS**), (URL : <https://esante.gouv.fr/>), par l'intermédiaire de son Centre de Gestion des Terminologies de Santé (**CGTS**), est en charge (URL : <https://esante.gouv.fr/interoperabilite/espace-des-terminologies-de-sante/cgts>):

- de la définition,
- de la production,
- du maintien et
- de la publication de Terminologies sur lesquelles s'appuient les systèmes d'information de santé.

L'ANS développe un serveur multiterminologique (SMT) en charge des tâches ci-dessus

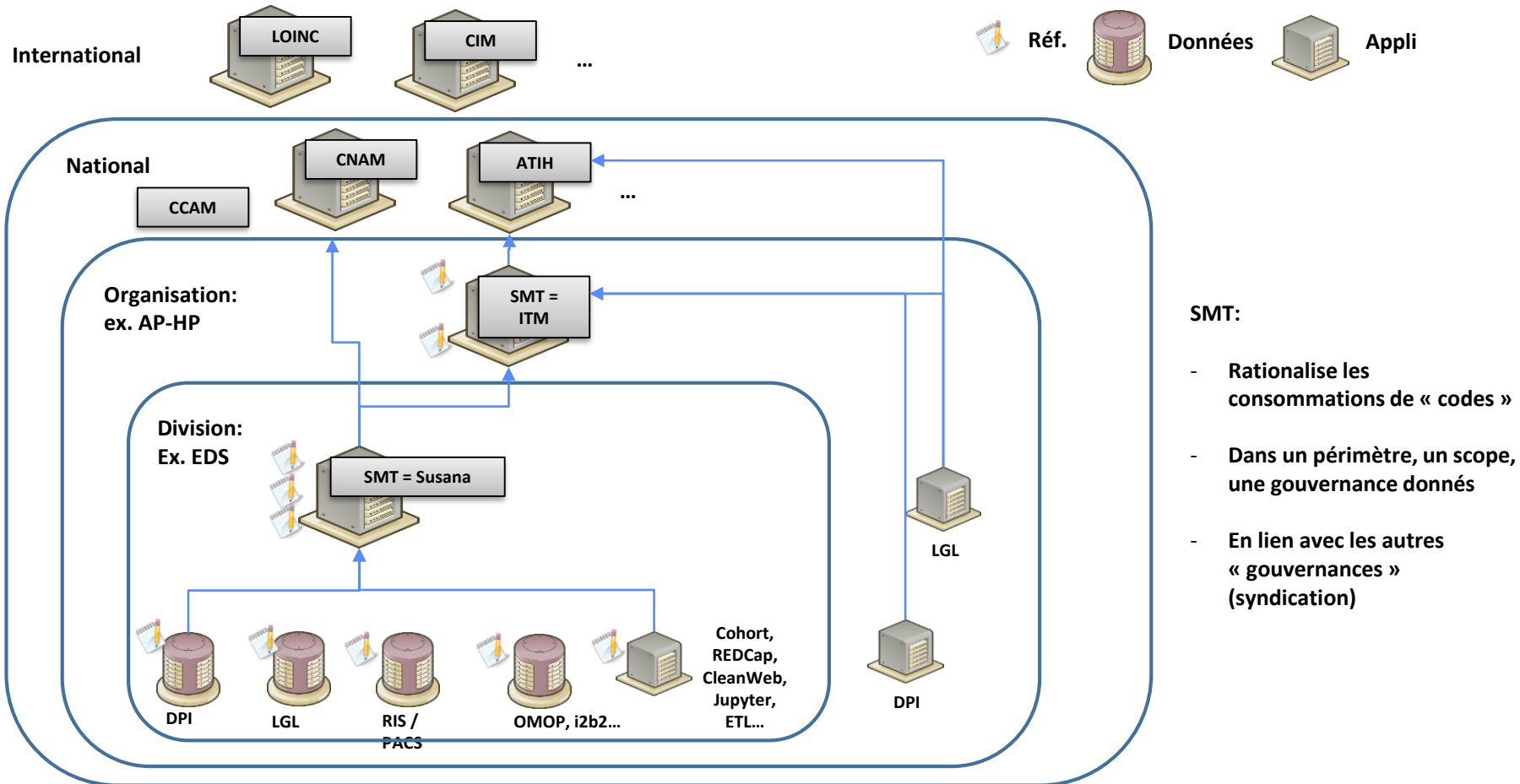
Dans un monde idéal, le D2IM du CHU de Rouen pourrait servir d'anté-serveur avec HeTOP (URL : www.hetop.eu), c'est-à-dire qu'une structure hospitalo-universitaire pourrait servir de tests sur les terminologies de santé en français (« bac à sable »)

TERMINOLOGIES EN USAGE: LA SANTÉ

Terminologies listées par le CGTS, en usage via le CI-SIS



PAS DE D'OUTILS (ICI SERVEURS MULTI-TERMINOLOGIES - SMT) SANS GOUVERNANCE



■ Enjeux :

- Rationalisation de la conception / validation / publication / documentation de Terminologies
- Rationalisation de la consommation des Terminologies

■ Objectifs :

- Mettre à disposition un outil technique industriel collaboratif provisionné avec des Terminologies qualifiées consommé par:
 - Silos techniques: ETL, Base de données, outils de création de cohortes...
 - Silos fonctionnels: PS, MOA, recherche clinique...

<https://icd.who.int/dev11/l-m/en#/http://id.who.int/icd/entity/1480627948>

Login, ...

The screenshot shows the ICD-11 web interface. At the top, there is a navigation bar with the text "ICD-11 (ICD-11 Mortality and Morbidity Statistics)" and a search bar containing "Pneumococcal pneumonia". To the right of the search bar are links for "Advanced Search", "Foundation", "Linearizations", "Proposals", and "Info". Below the search bar, a left-hand navigation menu lists categories: "Lung infections", "CA40 Pneumonia", and "CA40.0 Bacterial pneumonia". Under "CA40.0 Bacterial pneumonia", several sub-items are listed, with "CA40.07 Pneumonia due to Streptococcus pneumoniae" highlighted in orange. To the right of the menu, the main content area displays the details for "CA40.07 Pneumonia due to Streptococcus pneumoniae". This section includes a "Parent" link to "CA40.0 Bacterial pneumonia", a "Description" paragraph, and an "Inclusions" list with one item: "Bronchopneumonia due to S. pneumoniae".

Recherche

Codes

Navigation

Attributs / définitions



HeTOP

- Health Terminology/Ontology Portal
- Crosslingual: matrix navigation, among terminologies and languages, with two pivotal languages (English & French) and two pivotal terminologies (MeSH for indexing & SNOMED CT for annotating)
- D2IM, CHU de Rouen – LIMICS
- URL : www.hetop.eu
- **95 KOS, 52 languages**
- Partially free (44 KOS); some others after authentication (ICPC2)
- over 1,000 uniques machines per working day
- More than 4,700 subscribers
- In 2020, creation of the HUI (HeTOP Unique Identifier) for French concepts
- Technologies:
 - Meta-model
 - NoSQL database since 2016

Principales langues dans HeTOP

Langue	Nombre de termes	%
Anglais	5,382,259	36,84
Français	3,089,151	21,14
Espagnol	1,251,243	8,56
Portugais	406652	2,78
Italien	310262	2,12
Allemand	232798	1,59
Mandarin	227028	1,55
Néerlandais	214766	1,47
Russe	191574	1,31
Japonais	112062	0,77

Politique éditoriale de HeTOP

- Approche mixte contenu/contenant
- Multi-terminologique avec terminologie pivot (MeSH)
- Inter-lingue avec langues pivots (français/anglais) : *navigation matricielle sur les langues*
 - Intégration systématique des SOC en français (BNPC, CCAM, LPP, Cladimed, DRC, NABM, etc.) + versions françaises traduites par des tiers (OMS – CIM-10, CISP-2, CIF, ATC, ... - LOINC, ICNP, etc.)
- Choix des SOC centré sur les cas d'usages académiques (PMSI, DPI, indexation documentaire, recherche d'information, flux de terminologies, soins primaires, maladies rares, etc.)
- Dernières versions uniquement
- Enrichissements sémantiques : traductions (français essentiellement) et alignements

HeTOP

long feeding tube

Coronavirus Upd

קורונה - לוח בקרה

Coronavirus (COV

The Pandemic's H

Travel Med Int - S

Publications - CISMef

Artificial Intelligen

GPT-3 - Wikipéd

DBGUI

Home - L

https://www.hetop.eu/hetop/en?rr=LNC_HP_LP30689-1&oti=all&q=allergenes

HeTOP

English

allergenes

No wildcard search

Do not search into definitions

Terminologies selection

filter translated concepts

Your queries

0 matches in 0,04 s

Occupational allergens

(LOINC hierarchical part)

Description

Hierarchies

Relations

PubMed / Doc'CISMeF

Curation*

Translation

Full tree

LOINC top tree

Laboratory

Allergy

Occupational allergens*

Mixes°

Single allergens°

Abachi Wood Dust (Triplochiton scleroxylon)°

Acrylic fiber°

Alkalase*

American Beech (Fagus grandifolia) wood dust°

Amylase*

Benzene ring°

Bromelin*

Carageenan°

Cedar dust°

Chloramin T*

Coffee bean*

Cotton fibers*

Cotton gin dust°

Cotton linters°

Cottonseed°

Cultivated Cotton°

Ethylene oxide*

European Ash (Fraxinus excelsior) wood dust°

European silver fir wood dust°

Ficus sp°

Formaldehyde*

Glutaraldehyde*

Grain*

Grain dust*

Grain elevator dust°

Grain mill dust°

Green Coffee Bean°

- ☐ No wildcard search
- ☐ Do not search into definitions

Terminologies selection filter translated concepts

Your queries

0 matches in 0,04 s

Abachi Wood Dust (Triplochiton scleroxylon) (LOINC component) ⓘ

Description Hierarchies Relations PubMed / Doc'CISMeF Curation° Translation



Add a metadata

Origin ID

LP14168-6

Preferred Label

Abachi Wood Dust (Triplochiton scleroxylon)

poussière de bois Abachi (Triplochiton scleroxylon)

LOINC status

ACTIVE

LOINC short name

Triplochiton scleroxylon

poussière de bois Abachi Triplochiton scleroxylon

Synonyme automatique par l'ANS

Abachi Poussière de bois (ABACHI)

UMLS CUI

C1441303

For practice

- ☐ No wildcard search
☐ Do not search into definitions
☒ Terminologies selection ☐ filter translated concepts

Your queries

59 matches in 0,04 s

Top terms

- L95 vascularite [angéite] limitée à la peau, non classée ailleurs [ICD-10 category]
- Vasculitis limited to skin, not elsewhere classified [ICD-10 category (who)]
- **L95 osteoporosis [ICPC-2 Rubric]**
- L950 vascularite livédoïde [ICD-10 Sub-category]
- L951 erythema elevatum diutinum [ICD-10 Sub-category]
- L958 autres vascularites limitées à la peau [ICD-10 Sub-category]
- L959 vascularite limitée à la peau, sans précision [ICD-10 Sub-category]
- Lower Arm and Wrist Tendon, Right [PCS menu]
- Livedoid vasculitis [ICD-10 Sub-category (who)]
- Erythema elevatum diutinum [ICD-10 Sub-category (who)]

MeSH (17)

Genes & Proteins (5)

ICD-10 ATIH (5)

ICD-10 PCS (19)

ICD-10 WHO (5)

ICPC-2 (1)

NCIt (7)

L95 osteoporosis (ICPC-2 Rubric)



Description Hierarchies Relations PubMed / Doc'CISMeF



Show all languages

On

 Add a metadata

ICPC2 in 20 languages

Origin ID

L95

Preferred Label

- | | |
|---|---|
|  osteoporosis | |
|  ostéoporose | |
|  Osteoporose |  Osteoporose |
|  Osteoporosis |  Osteoporosi |
|  Osteoporoz |  osteoporosis |
|  Osteoporose |  骨粗鬆症 |
|  osteoporoz |  osteoporose |
|  osteoporos |  Osteoporose |
|  Loğng xurong |  Osteoporoz |
| |  Osteoporoz |

ICPC2 definition

-  characteristic imaging appearance
-  画像診断上特有の所見を有すること
-  sinais imagiológicos característicos
-  aspect imagistic caracteristic

Inclusion

-  pathological fracture due to osteoporosis
-  fracture pathologique dues à l'ostéoporose
-  patologiska frakturer beroende på osteoporos
-  骨粗鬆症による病理学的骨折
-  fraturas patológicas devido a osteoporose
-  fracturi patologice datorate osteoporozel
-  pathological fractures due to osteoporosis

Knowledge

HeTOP

 en 

marfan syndrome



☐ No wildcard search

☐ Do not search into definitions

☒ Terminologies selection ☐ filter translated concepts

Your queries

56 matches in 0,02 s

Top terms

- [marfan syndrome \[MeSH Descriptor\]](#)
- [marfan syndrome \[MeSH concept\]](#)
- [3.1711 - Marfan syndrome \[ACR pathology\]](#)
- [4.1711 - Marfan syndrome \[ACR pathology\]](#)
- [5.1971 - Marfan syndrome \[ACR pathology\]](#)
- [9.1971 - Marfan syndrome \[ACR pathology\]](#)
- [Marfan syndrome \[HRDODisease\]](#)
- [marfan's syndrome \[MedDRA Preferred Term\]](#)
- [marfan syndrome \[MedlinePlus Topic\]](#)
- [Marfan Syndrome \[NCIt concept\]](#)

☒ **MeSH (14)**

☒ **ACR (4)**

☒ **Gene Ontology (1)**

☒ **Genes & Proteins (1)**

☒ **HRDO (10)**

☒ **ICD-10 WHO (1)**

☒ **ICPC-2 (1)**

☒ **MedDRA (1)**

☒ **MedlinePlus (1)**

☒ **NCIt (2)**

☒ **OMIM (6)**

☒ **PASCAL (1)**

☒ **SNOMED CT (10)**

☒ **SNOMED int. (3)**

Marfan syndrome (MeSH Descriptor)



Description

Hierarchies

Relations

PubMed / Doc'CISMeF

Full tree

MeSH top tree

Diseases Category

cardiovascular diseases

cardiovascular abnormalities

heart defects, congenital

marfan syndrome

heart diseases

heart defects, congenital

marfan syndrome

congenital, hereditary, and neonatal diseases and abnormalities

congenital abnormalities

abnormalities, multiple

marfan syndrome

cardiovascular abnormalities

heart defects, congenital

marfan syndrome

genetic diseases, inborn

marfan syndrome

musculoskeletal diseases

bone diseases

bone diseases, developmental

marfan syndrome

skin and connective tissue diseases

connective tissue diseases

marfan syndrome

Knowledge

HeTOP

 en

marfan syndrome



T/O



☐ No wildcard search

☐ Do not search into definitions

☒ Terminologies selection ☐ filter translated concepts

Your queries

56 matches in 0,02 s

Top terms

- [marfan syndrome \[MeSH Descriptor\]](#)
- [marfan syndrome \[MeSH concept\]](#)
- [3.1711 - Marfan syndrome \[ACR pathology\]](#)
- [4.1711 - Marfan syndrome \[ACR pathology\]](#)
- [5.1971 - Marfan syndrome \[ACR pathology\]](#)
- [9.1971 - Marfan syndrome \[ACR pathology\]](#)
- [Marfan syndrome \[HRDODisease\]](#)
- [marfan's syndrome \[MedDRA Preferred Term\]](#)
- [marfan syndrome \[MedlinePlus Topic\]](#)
- [Marfan Syndrome \[NCIt concept\]](#)

MeSH (14)

ACR (4)

Gene Ontology (1)

Genes & Proteins (1)

HRDO (10)

ICD-10 WHO (1)

ICPC-2 (1)

MedDRA (1)

MedlinePlus (1)

NCIt (2)

OMIM (6)

PASCAL (1)

SNOMED CT (10)

SNOMED int. (3)

Marfan syndrome (HRDODisease)



Description

Hierarchies

Relations

PubMed / Doc'CISMeF

Intra-terminologic

Inter-terminologic

Add a metadata

Mappings

Semantic type(s) (1)

Related sign(s) (48)

Abnormal dentition/dental position/implantation/unerupted/dental ankylosis	HRDO Sign
Aortic dissection	HRDO Sign
Aortic root dilatation/dilation/aneurysm	HRDO Sign
Aortic valve anomaly/incompetence/insufficiency/regurgitation/bicuspid	HRDO Sign
Arterial aneurism (excluding aorta)	HRDO Sign
Arterial rupture	HRDO Sign
Articular/joint pain/arthralgia	HRDO Sign
Asthenia/fatigue/weakness	HRDO Sign
Autosomal dominant inheritance	HRDO Sign
Cleft palate without cleft lip/submucosal cleft palate/bifid uvula	HRDO Sign
Coxa profunda/acetabular protrusion	HRDO Sign

Related MeSH (1)

UMLS correspondences (same concept) (9)

Currated CISMeF NLP mapping (15)

SNOMED CT concept(s) (1)

MedDRA code(s) (1)

OMIM phenotype(s) (2)

Towards practice with the web of data...

HeTOP es

☐ No wildcard search
☐ Do not search into definitions
☒ Terminologies selection ☐ filter translated concepts

[Su consulta](#)

2321 Recurso(s) encontrado en 0,07 s

Top terms

- Evaluación [CISMeF resources Type]
- tiene evaluación [SNOMED CT concept]
- evaluación del coeficiente intelectual [SNOMED CT concept]
- autoevaluación [CISMeF resources Type]
- evaluación de la piel [SNOMED CT concept]
- Evaluación de Daños [DeCS]
- Evaluación en Salud [DeCS]
- evaluación del dolor [SNOMED CT concept]
- evaluación de salud [SNOMED CT concept]
- QD33 evaluación del estado de salud [Q-code]

MeSH (27)

CISMeF (10)

DeCS (28)

MedDRA (12)

SNOMED CT (2234)

Q-Codes (10)

- Q-code (10)**
 - **QD33 evaluación del estado de salud**
 - QP23 competencia cultural
 - QP41 satisfacción del paciente
 - QR21 farmacoepidemiología
 - QS46 farmacéutico
 - QT3 garantía de la calidad
 - QT13 evaluación de enseñanza
 - QT33 lectura crítica de la literatura
 - QT35 acreditación
 - QT52 evento adverso

QD33 evaluación del estado de salud (Q-code) http://www.hetop.eu/hetop/?la=es&r=CGP_QC_QD33

Descripción

Jerarquías

Relaciones

PubMed / Doc'CISMeF

Show all languages

Origin ID

QD33

Preferred Label

es evaluación del estado de salud
en health status assessment fr évaluation de l'état de santé

Category label

es Evaluación del estado de salud
en Health status assesment fr Evaluation de l'état de santé

Q-Code entry term

en Activities of Daily Living (ADL) en functional status fr Activités de la Vie Journalière (AVJ)

Q-Code acronym

en ADL fr AVJ

Q-Code definition

es
La evaluación objetiva y subjetiva de la capacidad de una persona para realizarse y adaptarse a su entorno durante un periodo de tiempo determinado. (Woncadic)
en
assessment of the ability of a person to perform and adapt to his environment, measured both objectively and subjectively over a stated period of time.(Woncadic)
fr
évaluation de la capacité d'une personne à fonctionner et s'adapter à son environnement, mesuré à la fois objectivement et subjectivement sur une période déterminée. (Woncadic)

Contextual queries

HeTOP  en    T/O  julien.grosjean@chu-roen.fr 

☐ No wildcard search
☐ Do not search into definitions
☒ Terminologies selection ☐ filter translated concepts

Your queries

56 matches in 0,06 s

Top terms

- [marfan syndrome \[MeSH Descriptor\]](#)
- [marfan syndrome \[MeSH concept\]](#)
- [3.1711 - Marfan syndrome \[ACR pathology\]](#)
- [4.1711 - Marfan syndrome \[ACR pathology\]](#)
- [5.1971 - Marfan syndrome \[ACR pathology\]](#)
- [9.1971 - Marfan syndrome \[ACR pathology\]](#)
- [Marfan syndrome \[HRDODisease\]](#)
- [marfan's syndrome \[MedDRA Preferred Term\]](#)
- [marfan syndrome \[MedlinePlus Topic\]](#)
- [Marfan Syndrome \[NCIt concept\]](#)

MeSH (14)

- MeSH Descriptor (3)**
 - [Fibrillin-1](#)
 - [Loeys-Dietz syndrome](#)
 - [marfan syndrome](#)
- MeSH Supplementary Concept (6)**
- MeSH concept (5)**

ACR (4)

Gene Ontology (1)

Genes & Proteins (1)

HRDO (10)

ICD-10 WHO (1)

ICPC-2 (1)

MedDRA (1)

MedlinePlus (4)

Marfan syndrome (MeSH Descriptor)   

[Description](#) [Hierarchies](#) [Relations](#) [PubMed / Doc'CISMMeF](#)

1. Allowed qualifier(s) for this descriptor:

[Display the qualifiers hierarchical list](#)

☐ blood	☐ cerebrospinal fluid	☐ chemically induced
☐ classification	☐ complications	☐ diagnosis
☐ diagnostic imaging	☐ diet therapy	☐ drug therapy
☐ economics	☐ embryology	☐ enzymology
☐ epidemiology	☐ ethnology	☐ etiology
☐ genetics	☐ history	☐ immunology
☐ metabolism	☐ microbiology	☐ mortality
☐ nursing	☐ parasitology	☐ pathology
☐ physiopathology	☐ prevention and control	☐ psychology
☐ radiotherapy	☐ rehabilitation	☐ surgery
☐ therapy	☐ urine	☐ veterinary
☐ virology		

2. Options

☐ only the main ones

☐ without explosion

all types 

3. Queries:

  
Littérature Scientifique en Santé

You may also use the Query Builder:

CRBM
Constructeur de Requêtes Bibliographiques Médicales

HeTOP: principaux chiffres

May 2010

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

May 2011

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

April 2013

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

March 2021

Terminologies	Concepts	Terms	Attributes	Relations
92 (21 UMLS)	3.3 M	14.6 M	23.5 M	15.6 M

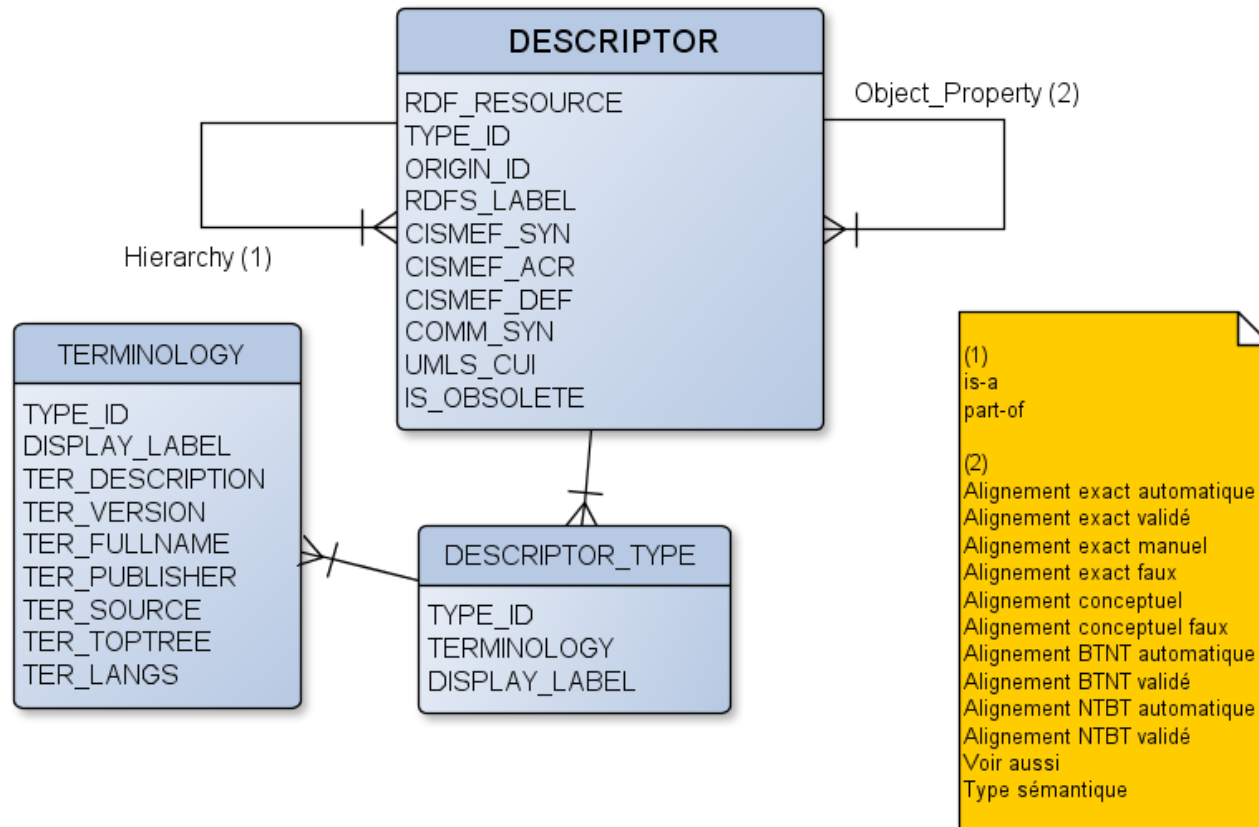
Au total dans HeTOP, nombre de CUI ayant au moins une traduction française disponible ≈ **520 000** vs. **144 000** concepts en français dans UMLS (2020AB) vs. dans HeTOP (≈ x3,6)

Base de données à gros volume (108 millions de triplets RDF en 2014) => Big data en connaissance sur la Santé

HeTOP : Health Terminology/Ontology Portal

- Évolution constante des données (mises à jours et nouvelles ressources)
- > 4,800 resgistered users (+850 en 2019) :
 - 36% students;
 - 30% physicians ;
 - 6% librarians;
 - 28% others (translators, researchers, paramedics, ...).
- Around 3,000 queries by working day
- Around 20,000 uniques users per month

Modèle générique HeTOP

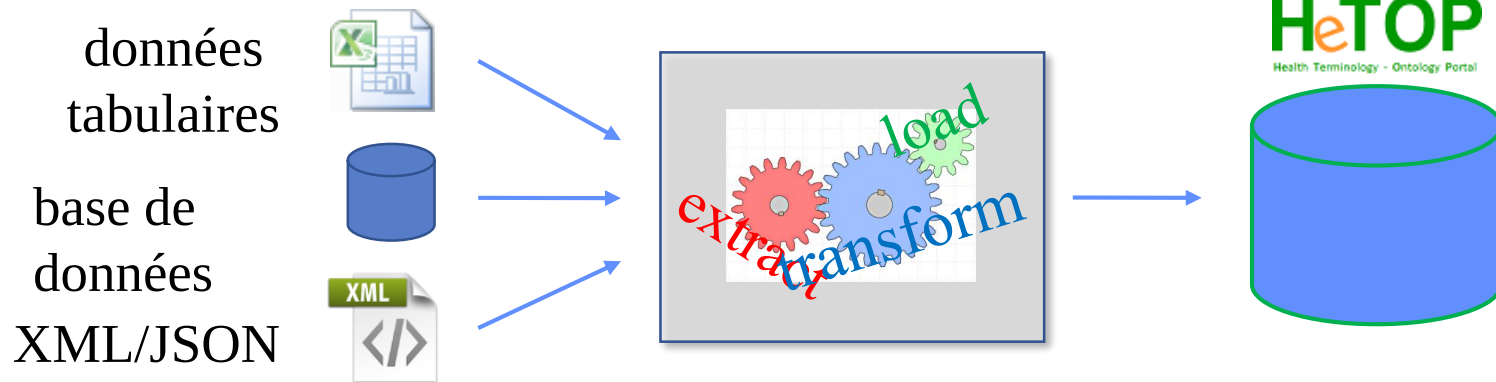


- Compatible avec les recommandations de Tao et al. (2013) et norme ISO 25964 (thésaurus)

Intégration des SOC dans HeTOP

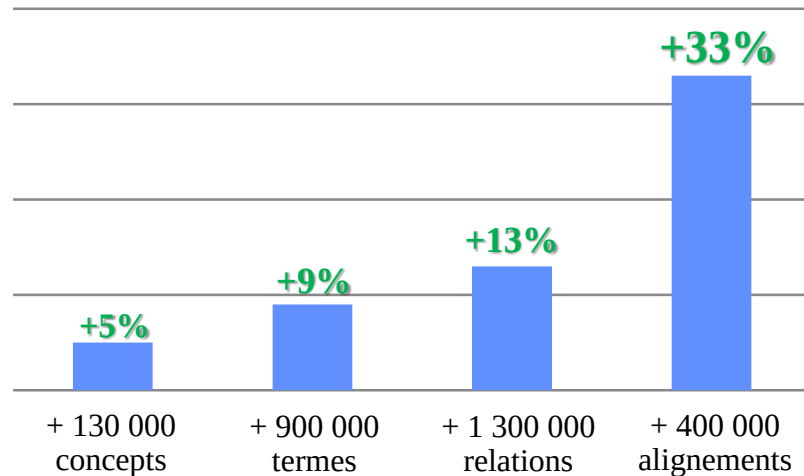
Création de programmes ETL (Extract-Transform-Load)
développés en JAVA.

Sources SOC :



Évolution d'HeTOP sur 2018

Enrichissement du contenu
d'HeTOP :



Traductions : 20 000 concepts traduits en français.

- Traductions automatiques simples (30%)
- Validation de traductions proposées via les alignements (30%)
- Traductions manuelles

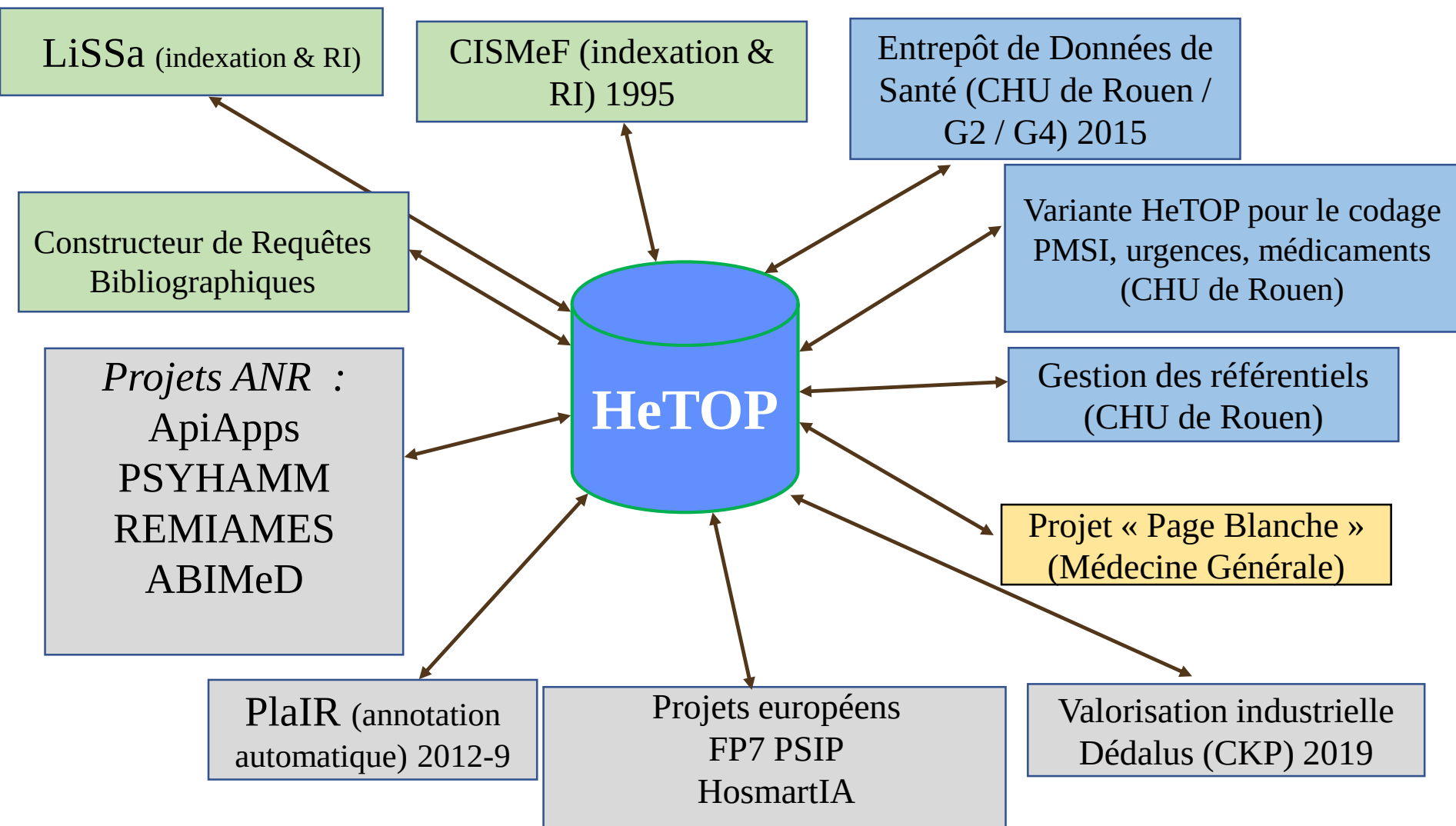
Cas d'usage HeTOP

- En interne, au D2IM, département de l'informatique et de l'information médicales du CHU de Rouen
- Serveur terminologique pour :
 - Indexation multi-terminologique (CISMeF, LiSSa)
 - Support pédagogique (étudiants en médecine, sciences de l'information)
 - Médicaments Virtuels Nationaux : short list (annoncé par HAS)
 - SNOMED CT (coopération avec le Pr. Lemoine, Hop. Erasme, Bruxelles)
 - CISP-2 ? (discussion avec le CISP Club)
 - Support pour la maintenance de la traduction française MedDRA (ICH)
 - Création de SOC
 - Q-Codes (Marc Jamouille, MD, doctorant thèse de sciences)
 - Ontologie bio-informatique (startup OMICX)
 - Thésaurus Informatique Médicale (EFMI), financé par le projet européen HospIA

Cas d'usage HeTOP

- Serveur de référentiel au CHU de Rouen depuis fin 2016
 - Terminologies d'interface alignées à des terminologies de référence (biologie, imagerie) <= projet TerSan ANR
- Instanciation HeTOP
 - Version DIM (simplifiée) pour les TIM (CIM-10, CCAM)
 - Version Urgences pour coder les RPU (CIM-10, SFMU)
 - Version Médicament : www.hetop.eu/hetop/drugs
- BDD HeTOP : support terminologique et lexical de l'outil d'indexation automatique ECMT (commercialisé par Alicante)
- Projets de recherche (ApiApps, PSYHAMM, REMIAMES, ABIMeD, TerSan, SYNODOS, RAVEL, SIFADO, PlaIR2.018, C3-Cloud, ...) : service web + « vitrine » (L3IM pour VCM, ...)

Plateforme terminologique sémantique



Plateforme terminologique

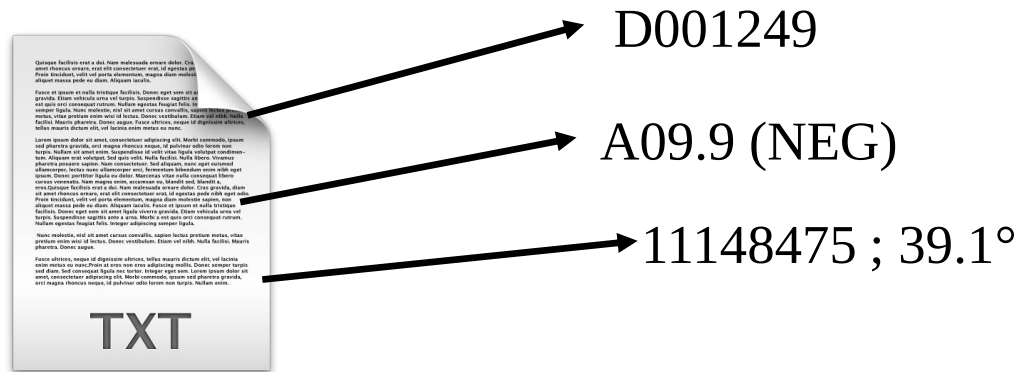
- Utilisation par des producteurs de contenu (MedDRA, Q-Codes, CISP-2)
- Indexation (bibliothèques, centres documentaires, ...)
- Enseignement (facultés de Rouen, Paris, Marseille, Québec, ...)
- Traduction
- Accès à la connaissance (PubMed)

HeTOP : utilisations

- Plus de 3 900 utilisateurs inscrits (800 en 2018) :
 - 40% étudiants ;
 - 32% médecins ;
 - 8% documentalistes ;
 - 20% autres (traducteurs, paramédicaux, chercheurs, ...)
- Environ 2 500 recherches par jour en moyenne
- Environ 1 000 utilisateurs uniques par jour en moyenne

ECMT

- Extracteur de Concepts Multi-Terminologique



- Extraire et structurer les informations (événements/connaissances)
- 13 M de documents : 24h de traitement => 2 milliards de concepts après filtrage (6ms par document)
- nombreuses études d'évaluation (ISP & IMG +++): quantitative et qualitative

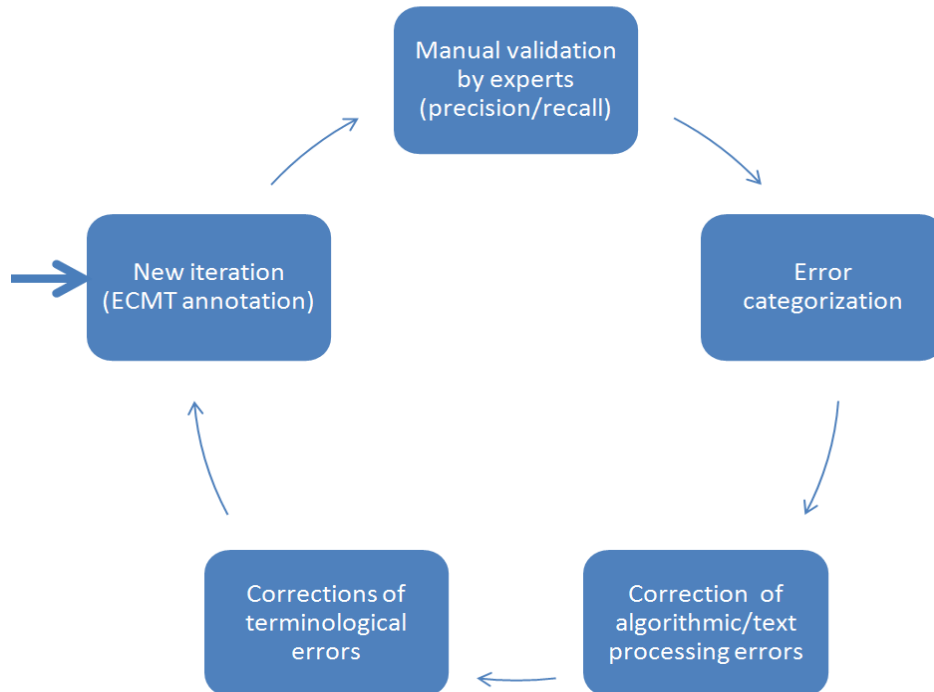
ECMT

- Evaluations :

	P/R 1 ^{ère} itération	P/R 2 ^{ème} itération
EDS Rouen	0,36/0,63	0,62/0,68
LiSSa	0,72/0,85	0,91/0,87
ApiApps	0,80/0,84	

Annotation de documents médicaux - EDSaN

- Projet 2017-2018 : annotation complète de 12 millions de documents médicaux (CHU de Rouen à partir de 2000).
- Méthodologie d'amélioration itérative :



Grosjean J, Lahrach M, Ndangang M, Lelong R, Dahamna B & Darmoni SJ. Health documents automatic annotation in French: a qualitative evaluation. Submitted to MEDINFO 2019

ECMTE

- Interface web pour tester, visualiser voire valider les annotations automatiques

MMSE 18 10 17
BREF 4 4 6
Mattis DRS 72 63 92
SAPS (signes positifs) 59 33 36
SANS (signes négatifs) 78 81 35
ADL 4,5 3 3,5
FBI (signes frontaux) 17 29 39

On notera donc une fluctuation cognitive importante puisque le MMSE après avoir perdu 8 points reprend 7 points, la BREF semble stable et la MATTIS, après s'être discrètement aggravée s'améliore. Néanmoins, on notera cliniquement une aggravation sur le plan frontal puisque l'on note plus de persévérations et un comportement de préhension. Il est tout à fait notable également que l'attention lors des tests était bien meilleure cette fois ci, et que le délire était moindre, on notera notamment dans les scores de signes négatifs de schizophrénie, une amélioration par rapport à l'évaluation précédente témoignant d'une moindre apathie notamment.

L'interrogatoire par téléphone de l'équipe soignante retrouve que l'autonomie au quotidien, après s'être aggravée dans un premier temps, reste stable et que les signes frontaux semblent plus importants qu'auparavant.

La ponction lombaire a été pratiquée, les résultats standards sont : leucocytes < 2/mm3, hématies < 2/mm3, protéinorachie = 0,6g/L. L'index gamma ainsi que les résultats des biomarqueurs Tau et Abeta vous seront communiqués dès réception.

Au total, le contrôle des tests neuropsychologiques ne permettent pas de conclure avec certitude. En effet, les variations des scores aux tests neuropsychologiques peuvent être en partie expliquées les variations sur le plan thymique et sur le plan des signes psychotiques ainsi que les modifications thérapeutiques. Néanmoins, l'aggravation des signes frontaux et l'aggravation de l'autonomie par rapport à 2010 sont toujours compatibles avec l'hypothèse d'une dégénérescence lobaire fronto-temporale qui était étayée par la présence d'une atrophie modérée antérieure et surtout d'une hypoperfusion antérieure sur la scintigraphie.

Dans tous les cas, le suivi neuropsychologique reste nécessaire, de même qu'un contrôle de l'IRM et de la scintigraphie courant 2012.

En vous remerciant de votre confiance.

Afficher	10	▼ éléments	Filtrer :		
#	Concept	Évaluation			
51	bref	✓	✗	?	!
52	stable	✓	✗	?	!
53	discret	✓	✗	?	!
54	pire	✓	✗	?	!
55	Amélioré	✓	✗	?	!
56	pire	✓	✗	?	!
57	Plans frontaux	✓	✗	?	!
58	coronal	✓	✗	?	!
59	persévération	✓	✗	?	!
60	Comportement	✓	✗	?	!

Affichage de l'élément 51 à 60 sur 125
éléments

Précédent

1

...

5

Overall ECMT process

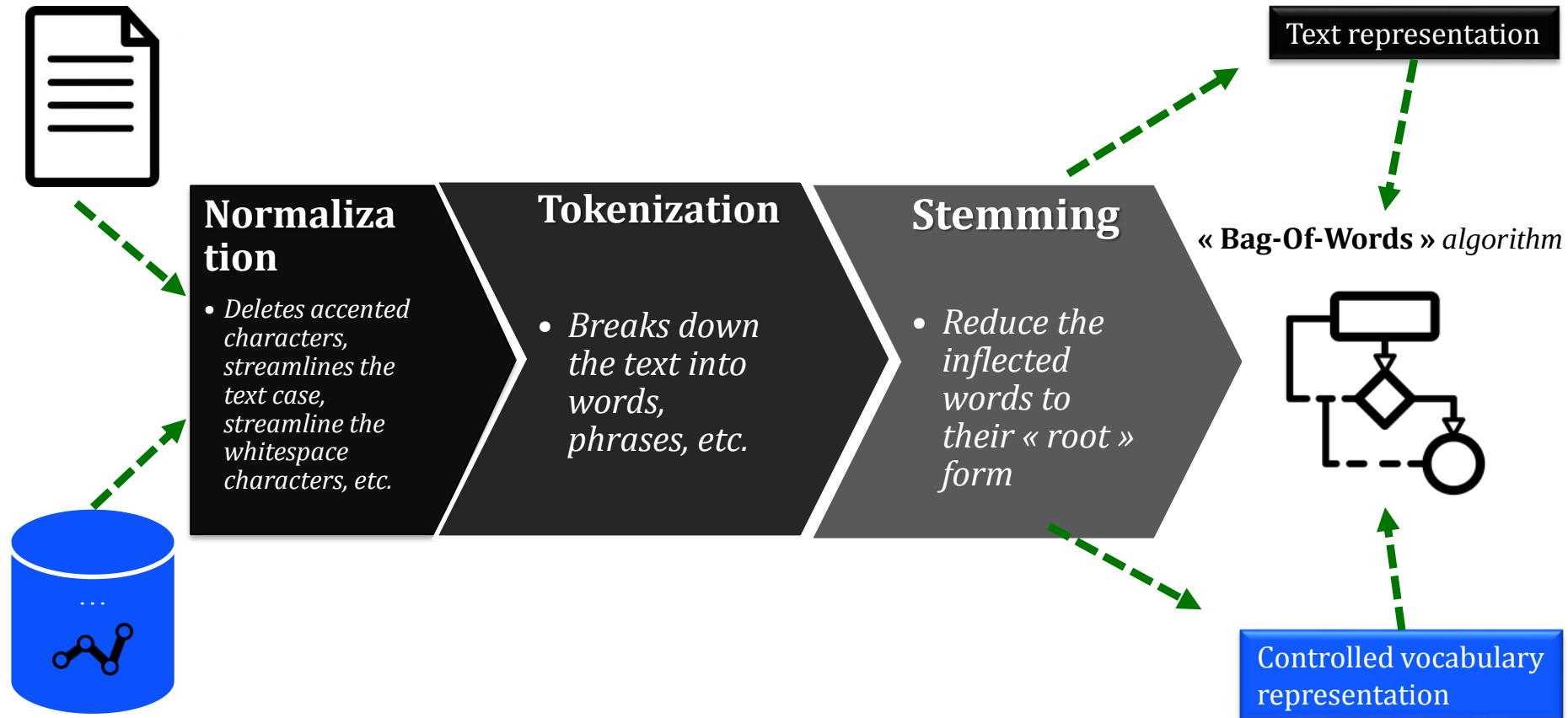


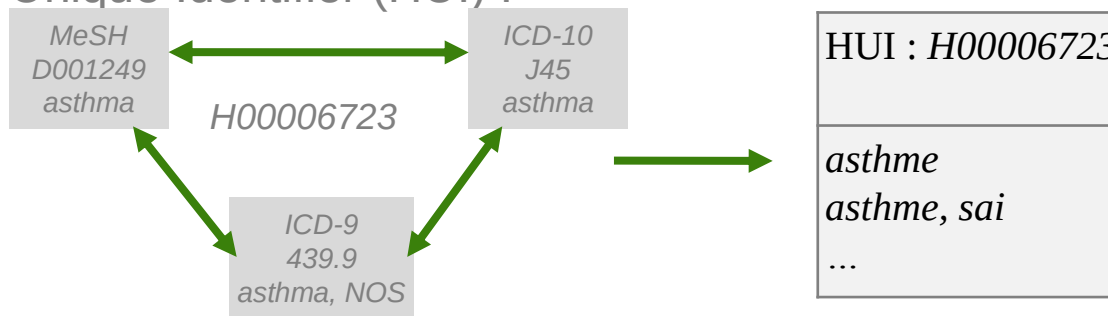
Table 7. System performance for ICD10 coding on the **French raw** test corpus in terms of Precision (P), recall (R) and F-measure (F). A horizontal dash line places the frequency baseline performance. The top part of the table displays official runs, while the bottom part displays non-official and baseline runs.

ALL					EXTERNAL				
Team		P	R	F	Team		P	R	F
Official runs	SIBM-run1	.857	.689	.764	SIBM-run1	.567	.431	.490	
	LITL-run2	.666	.414	.510	LIRMM-run1	.443	.367	.401	
	LIRMM-run1	.541	.480	.509	LIRMM-run2	.443	.367	.401	
	LIRMM-run2	.540	.480	.508	LITL-run2	.560	.283	.376	
	LITL-run1	.651	.404	.499	LITL-run1	.538	.277	.365	
	TUC-MI-run2 - - - - -	.044	.026	.033	TUC-MI-run2 - - - - -	.010	.004	.005	
	TUC-MI-run1	.025	.015	.019	TUC-MI-run1	.006	.005	.005	
average		.475	.358	.406	average	.367	.247	.292	
median		.541	.414	.508	median	.443	.283	.376	
Non-official	LIMSI-run2	.872	.784	.825	LIMSI-run2	.700	.594	.643	
	LIMSI-run1	.883	.760	.817	LIMSI-run1	.709	.559	.625	
	TUC-MI-run1-corrected	.883	.539	.669	TUC-MI-run1-corrected	.780	.290	.423	
	TUC-MI-run2-corrected	.882	.536	.667	TUC-MI-run2-corrected	.767	.283	.414	
	UNIPD-run1	.629	.468	.537	UNIPD-run2	.350	.381	.365	
	UNIPD-run2	.518	.384	.441	UNIPD-run1	.362	.251	.296	
	Mondeca-run1 - - - - -	.375	.131	.194	Mondeca-run1	.335	.228	.271	
Frequency baseline		.339	.237	.279	Frequency baseline	.381	.110	.170	

Level of French translations in HeTOP

HUI HeTOP Unified Concepts

For each HeTOP concept, attribution of a HeTOP Unique Identifier (HUI) :



>600 000 HeTOP Unified Concepts (CoMUF), similar to UMLS CUI

UMLS (Unified Medical Language System)



US Meta-thesaurus (NLM, NIH) with more than 200 health KOS
> 4 millions of unique concepts (CUI Concept Unique Identifier)



julien.grosjean@chu-rouen.fr
kevin.billey@univ-rouen.fr
stefan.darmoni@chu-rouen.fr
florent.desgrippes@aphp.fr

- L'ensemble de ce document relève des législations française et internationale sur le droit d'auteur et la propriété intellectuelle. Tous les droits de reproduction de tout ou partie sont réservés pour les textes ainsi que pour l'ensemble des documents iconographiques, photographiques, vidéos et sonores.

Ce document est interdit à la vente ou à la location. Sa diffusion, duplication, mise à disposition du public (sous quelque forme ou support que ce soit), mise en réseau, partielles ou totales, sont strictement réservées à l'université de Rouen.

L'utilisation de ce document est strictement réservée à l'usage privé des étudiants inscrits à l'UFR de médecine de l'université Rouen, ainsi que ceux inscrits au C2I Santé, et non destinée à une utilisation collective, gratuite ou payante.

Ce document a été réalisé par la Cellule TICE Médecine de la Faculté de Médecine de Rouen (Courriel : Florence.Charles@univ-rouen.fr).