Alignements de terminologies

04 Mai 2021
Stefan Darmoni
Florent Desgrippes
Tayeb Merabti

Module Interopérabilité et Standards de Données Médicales **(ISDM) -** Majeure Santé EPITA

- Définir l'activité d'alignement terminologique
- Partager ses enjeux et ses objectifs
- Décrire les principales méthodes d'alignement
- Les illustrer au travers d'un SMT institutionnel: HeTOP
- Faire le lien entre alignement automatique et validation humaine
- Perspectives

- **DPI Stefan DARMONI**
  - PU-PH, CHU de Rouen / D2IM / LIMICS
  - Liens d'intérêt : actionnaire de la société de conseil MedaïS Scientific en santé numérique
  - Pas de rémunération pour ce cours
- **DPI Florent DESGRIPPES**
  - Ingénieur, AP-HP
  - Pas d'intérêt
  - Pas de rémunération pour ce cours
- **DPI Tayeb MERABTI**
  - Docteur en informatique, ANS
  - Pas d'intérêt
  - Pas de rémunération pour ce cours
- **Licence du support**
  - Licence Ouverte v2
    - LOv2
    - https://www.etalab.gouv.fr/licence-ouverte-open-licence
    - auteurs: ANS / AP-HP / D2iM

Plan du cours

1. SOC (rappels)
2. Alignements de terminologies

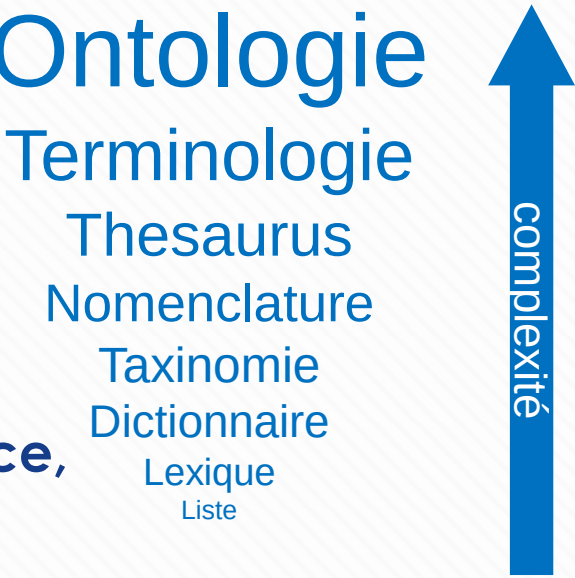
   Définitions, Approches
3. Méthodes d'évaluation des alignements
4. Aligner pour traduire

   Comment les alignements peuvent aider à traduire: les approches UMLS, HeTOP
5. Validation humaine d'alignements avec HeTOP
6. Outils sémantiques: autres cas d'usage et perspectives

1.

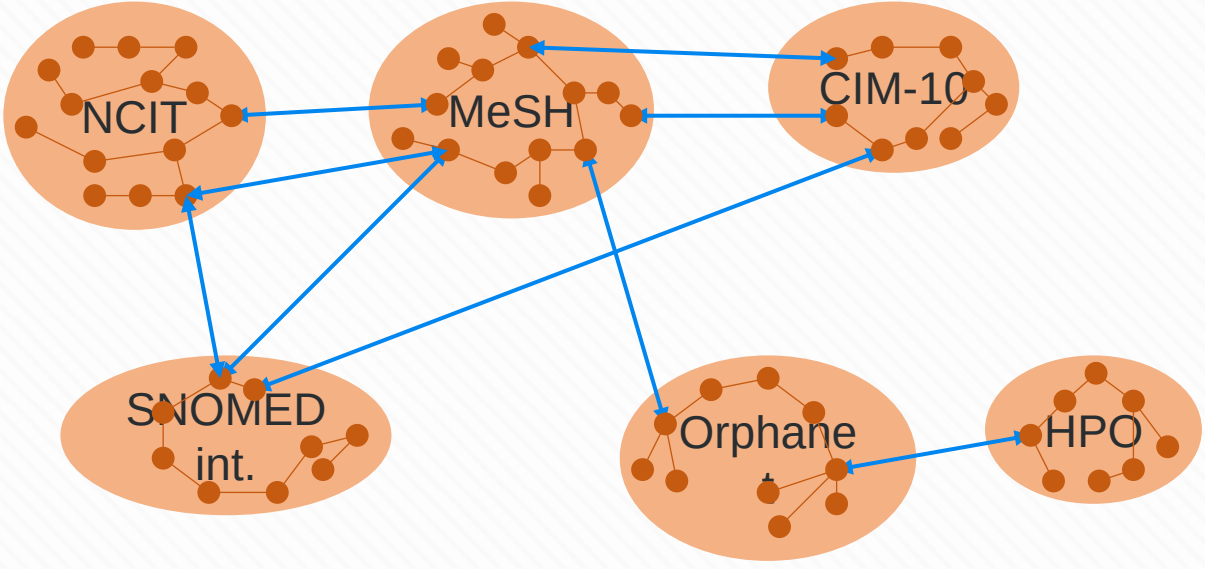
SOC (Rappels)

- **Les Systèmes d'Organisation des Connaissances (SOC)**
- **Les terminologies**
- **Notion de concept terminologique**
- **connaissance~intelligence (inférence, subsomption, lexiques, etc.)**
- **Le domaine de la Santé**



- Apporter du sens aux données (sémantique)
- Exploiter ce sens via les lexiques et les structures informatiques.

- Indexer / annoter / coder des ressources (décrire)
- Ranger
- Rechercher
- Classer
- Inférer
- Enseigner / apprendre (connaissances)



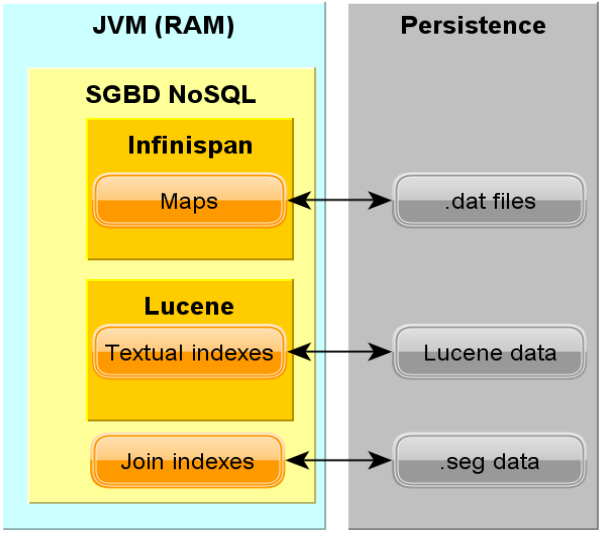
https://www.hetop.eu/hetop/documentation/sem_net.htm

- **Health Terminology/Ontology Portal**
- **D2IM, CHU de Rouen – LIMICS**
- **URL : www.hetop.eu**
- **85 KOS, 45 languages**
- **Partially free (27 KOS); some others after authentification (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**

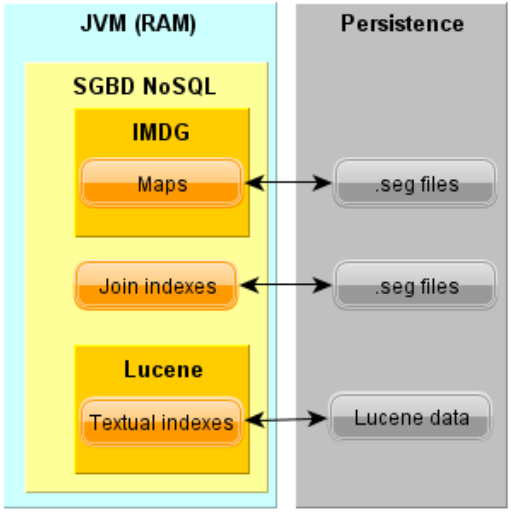
- **Technologies:**
  - Meta-model
  - BDD NoSQL since 2016

- **UMLS (US National Library of Medicine)**
- **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)**

- Implémentation complexe (changement de paradigme = changement de méthodes, repenser certaines applications)
- Construction d'un SGBD « sur-mesure » (intégration/évolution)



- **Adaptation à de très grands ensembles de données (Téraoctets) : persistance / éviction / index / types**



HeTOP: MAIN FIGURES

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 2020

Terminologies	Concepts	Terms	Attributes	Relations
85 (17 UMLS)	**3.3 M**	**13.1 M**	**21.3 M**	**13.3 M**

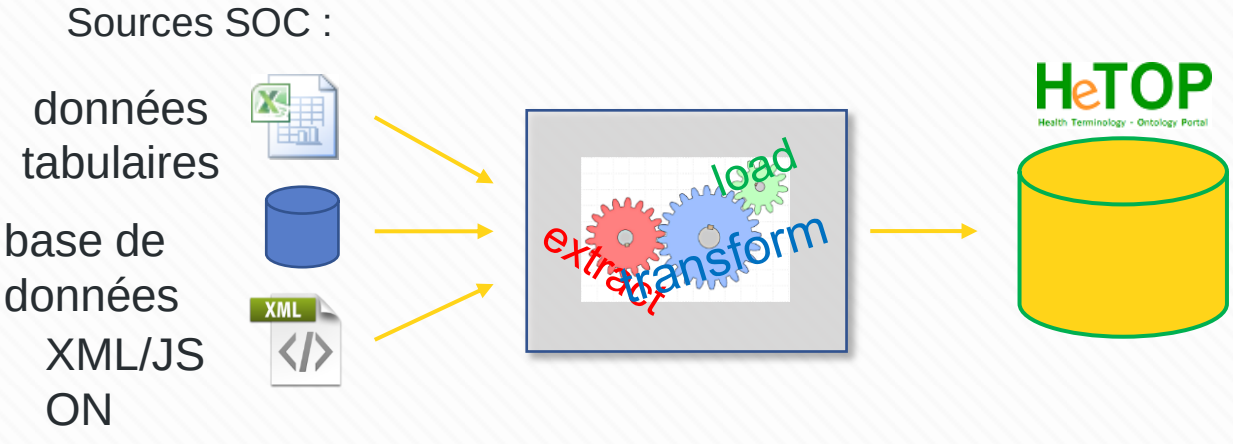
- **Migration terminée**
- **Évolution constante des données (mises à jours et nouvelles ressources)**
- **Nouveaux modules**
- **Volumétries :**
  - 75 SOC
  - 2 677 925 concepts : **1 035 298** CUI UMLS dont **438 195** (42%) avec traduction française vs **143 871** dans UMLS
  - 11 055 147 termes
  - 18 413 704 attributs
  - 11 485 117 relations
    - 1 522 715 alignements

      103 602 alignements manuels

      1 419 113 alignements automatiques

      - *509 322 alignements validés*

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



- CIM-10 (Classification Internationale des Maladies)
- CIM-9
- CISP-2 (Classification Internationale des Soins Primaires)
- Disease Ontology
- NCIT (National Cancer Institute Thesaurus)
- Orphanet (Orphanet Rare Disease Ontology)

Pathologies

- CCAM (Classification Communes des Actes Médicaux)
- NABM (Nomenclature des Actes de Biologie Médicale)

Actes médicaux

- FMA (Foundational Model of Anatomy)
- RadLex (Radiology Lexicon)

Anatomie

- ATC (Anatomical Therapeutic Chemical)
- BDPM (Base de Données Publique des Médicaments)
- Médicaments virtuels (MédicaBase)

Médicaments

- HPO (Human Phenotype Ontology)
- OMIM (Online Mendelian Inheritance in Man)

Génétique / Phénotypes

- Cladimed (Classification des Dispositifs Médicaux)
- LPP (Liste des Produits et des Prestations)

Dispositifs médicaux

- MeSH / DeSC (Medical Subject Headings)

Indexation

- **UMLS (Unified Medical Language System)**
- **Méta-thésaurus : mise en correspondance de terminologies**
- **Regroupe près de 200 SOC dont 17 en commun avec HeTOP (MeSH, CIM-10, SNOMED-CT, MEDDRA, MEDLINE+, etc.)**
- **Plus de 10 millions de concepts uniques UMLS (CUI)**
- **Fournit des bases de données pour son exploitation**

- **Concept Unique Identifiers (CUI) attribués à tous les concepts des SOC contenus dans UMLS.**

  - *MeSH D001249 asthme*
  - *CIM-10 J45 asthme*
  - *CIM-9 439.9 asthme, sai*

C0004096

- **Attribution des CUI UMLS aux concepts contenus dans HeTOP :**
  - 85 000 CUI importés (92% d'insertions et 8% de mises à jour) ;
  - 15 000 CUI propagés vers d'autres terminologies.

Nombre de concepts UMLS avec des termes en français :

- UMLS : 143 871 concepts (143 762 pour UMLS 2017)
- HeTOP : plus de 520 000 concepts (425 269 début 2018)

Soit environ un facteur 3 entre UMLS et HeTOP

Explique en grande partie l'intérêt d'HeTOP pour le français médical

Plus de 2200 comptes sur HeTOP (dont 120 traducteurs/traductrices)

Valorisation industrielle en 2019 (société Dédalus pour créer son CKP Clinical Knowledge Portal)

2.

Alignements de terminologies (définitions et approches)

- **Plusieurs termes sont utilisés par différents auteurs pour définir le mécanisme de mise en correspondance entre les termes de différentes terminologies : mapping, alignement…**
- **Le terme « alignement » est utilisé pour désigner le processus permettant la mise en relation entre terminologies**
  - Par extension, « alignement » désigne parfois le résultat du processus
- **Le « mapping » est une version directe de l'alignement qui exige principalement que l'objet mappé soit égale à son image, i.e., que la relation entre termes est une relation d'équivalence.**
  - mapping (en français) ≠ alignement.
- **Alignement : une fonction qui prend en entrée deux ontologies (terminologies) et qui retourne un ensemble de correspondances**

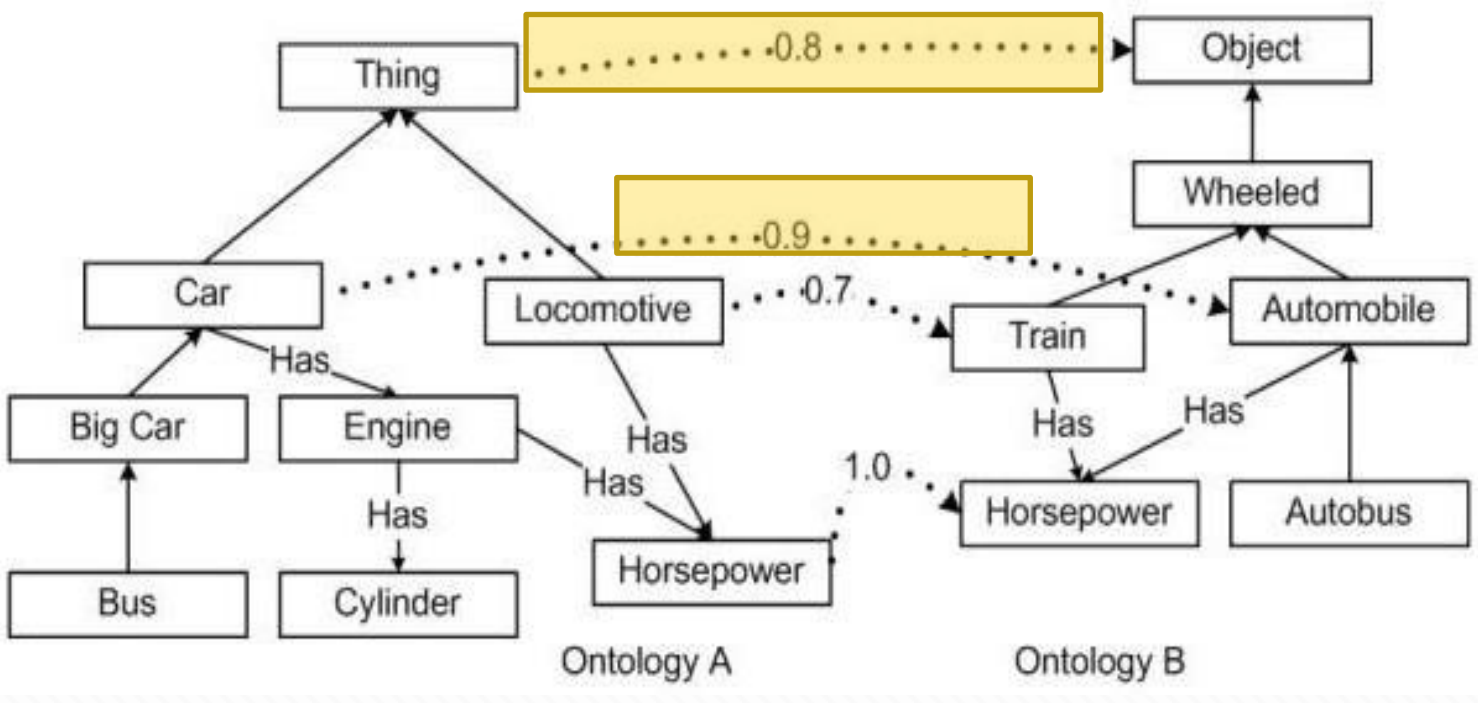
  (Euzenat, J. et Shvaiko, P. (2007). Ontology Matching).
- **Correspondance :** est une relation existante ou supposée existante suivant un algorithme particulier entre entités de différentes terminologies.
  - Exacte, Hiérarchique, proche,…

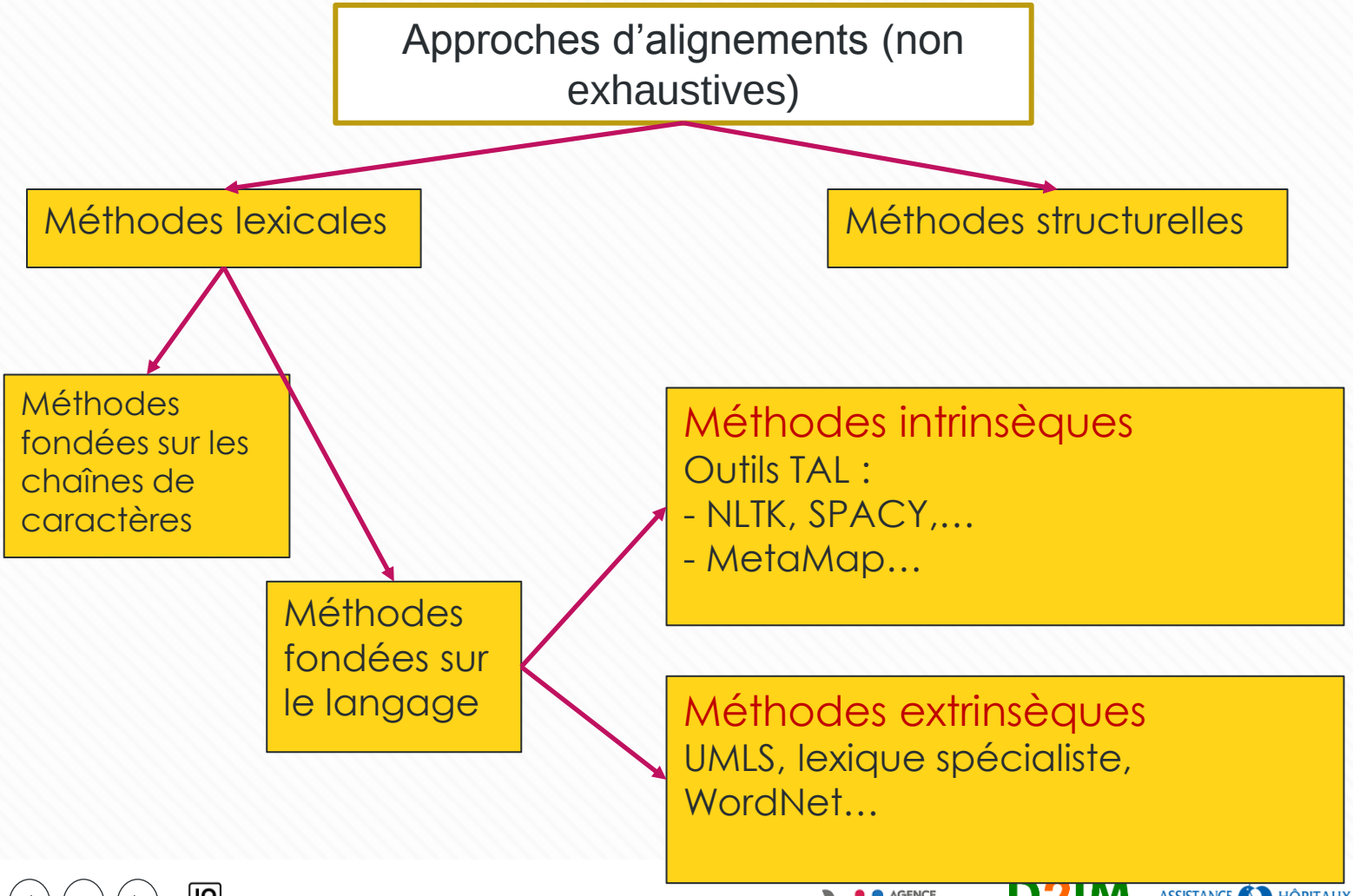
- **Parfois, des « alignements » sont directement fournis avec la Terminologie d'intérêt**
  - Le processus y ayant conduit n'est pas toujours connu, documenté ou reproductible
- **Ex. de Terminologies fournies avec des alignements:**

Terminologie « source »	Alignement vers:
SNOMED-CT	CIM-10, CISP-2
ICHI (OMS, actes)	ICF (OMS, handicap)
…	

- **La combinatoire {sources; cibles} est telle qu'il est rare d'avoir l'alignement correspondant à son besoin**

ALIGNEMENTS ENTRE ONTOLOGIES





Méthodes lexicales :

- les méthodes qui s'appuient sur des propriétés lexicales des termes de chaque terminologie.
- L'approche la plus triviale d'identifier les correspondances entre termes.
- Très utilisées dans le domaine de la médecine où la plupart des terminologies partagent un grand nombre de termes similaires.

Méthodes lexicales

1. Méthodes fondées sur les chaînes de caractères

- Les libellés (termes) sont considérés comme des séquences de caractères.
- Distance de Hamming, Distance de Jaccard, Distance de Dice, Distance d'édition, Distance de Levenshtein, Distance de Stoilos…
- Exemple : L'outil OnaGUI (OntoAlignementGUI)

2. Les méthodes fondées sur le langage

- Elles considèrent les termes comme étant des mots dans un langage naturel.

Méthodes intrinsèques :

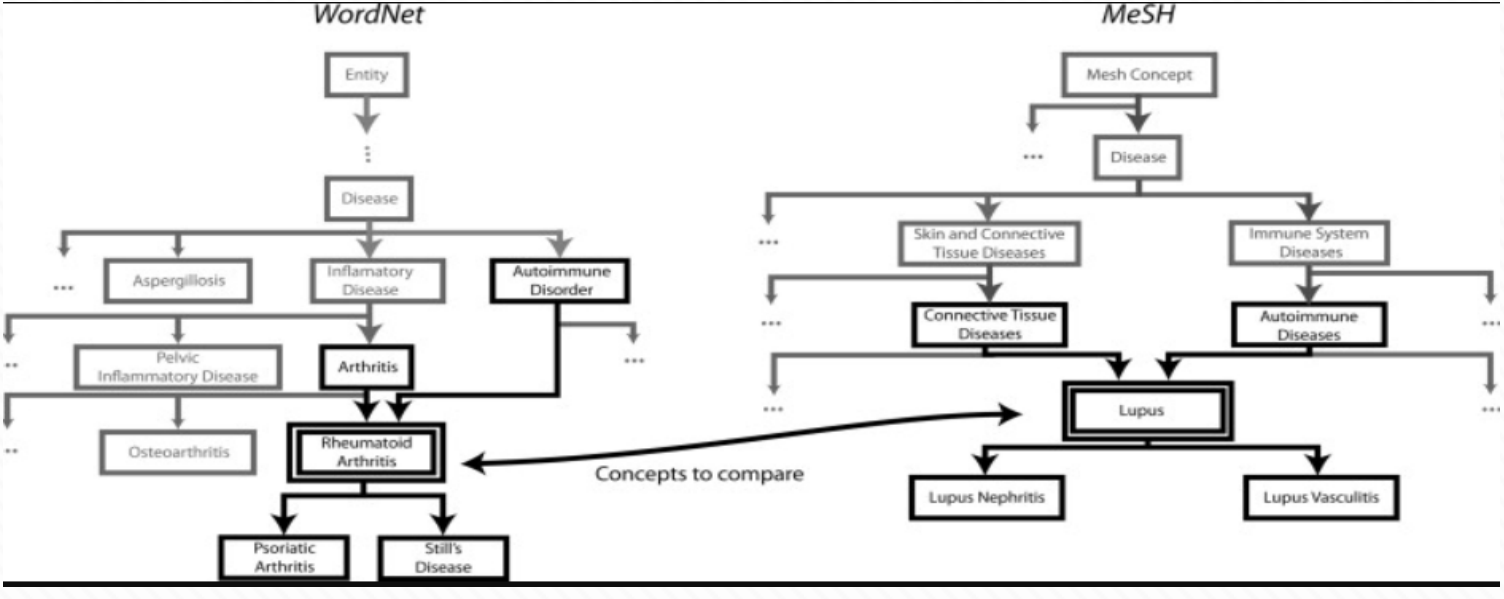
- Outils TAL pour normaliser les termes : Tokenization, Déssufixation, Elimination des mots vides…

Méthodes extrinsèques :

- Ressources externes : WordNet, dictionnaires…

Méthodes Structurelles (sémantiques):

- Propriétés structurelles des terminologies pour établir des correspondances possibles.
- Définissent des mesures de similarité fondées sur les structures hiérarchiques : distance Wu-Palmer, distance de Resnik, distance de Lin…

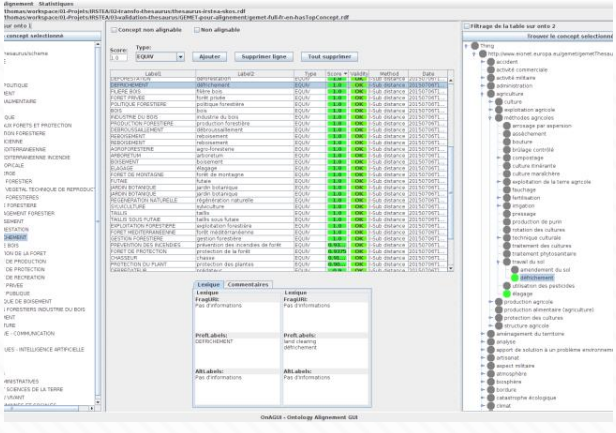


Exercice :

- Ci-joint deux listes de termes :
  - L1 = [« Alexandre, maladie », « Asthme de l'enfance »]
  - L2 = [« maladie d'Alexandre », « Asthme chez l'enfant »]

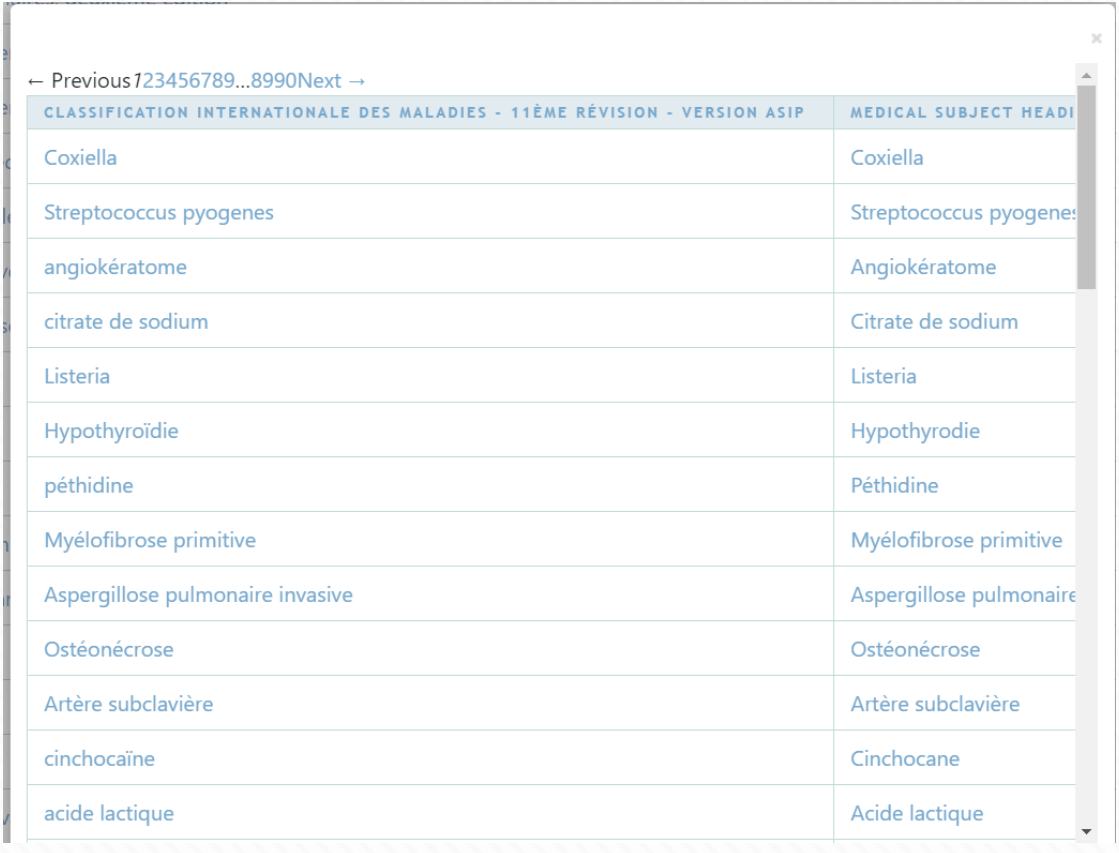
1. Quelle approche peut-on utiliser pour aligner L1 et L2.
2. Ecrire un script python qui permet d'aligner les termes de L1 vers L2 (Utiliser des bibliothèques de type NLTK par exemple).

- **« ontology alignment/matching tools »**
- **OnAGUI**
- **https://oaei.ontologymatching.org/**
  - **Ontology Alignment Evaluation Initiative**
  - 13 « challenges »
  - Outils listés dans les résultats
    - http://oaei.ontologymatching.org/2020/results/
  - LogMap: voir intégration de OWL2Vec
  - AML…



Credits: Sparna

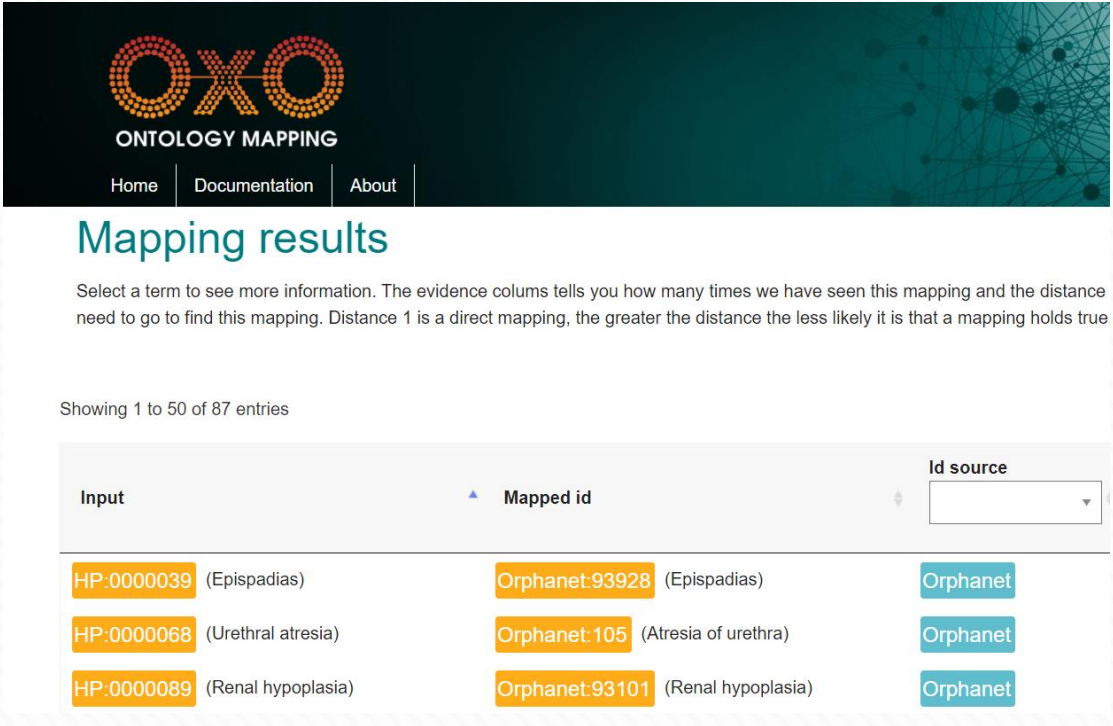
- **http://bioportal.lirmm.fr/ontologies/CIM-11/?p=mappings**
  - (CIM-11 vs MeSH)



← Previous 1 2 3 4 5 6 7 8 9 ... 89 90 Next →

CLASSIFICATION INTERNATIONALE DES MALADIES - 11ÈME RÉVISION - VERSION ASIP	MEDICAL SUBJECT HEADI
Coxiella	Coxiella
Streptococcus pyogenes	Streptococcus pyogenes
angiokératome	Angiokératome
citrate de sodium	Citrate de sodium
Listeria	Listeria
Hypothyroïdie	Hypothyrodie
péthidine	Péthidine
Myélofibrose primitive	Myélofibrose primitive
Aspergillose pulmonaire invasive	Aspergillose pulmonaire
Ostéonécrose	Ostéonécrose
Artère subclavière	Artère subclavière
cinchocaïne	Cinchocane
acide lactique	Acide lactique

- **https://www.ebi.ac.uk/spot/oxo/datasources/HP**
  - EMBL-EBI OxO: https://github.com/EBISPOT/OXO (Neo4j, Solr)
  - Ici: HPO vs Orphanet



Exercice :

- Compiler OnAGUI et LogMap
- Exécuter le « Mouse Challenge » (anatomie)
- Discuter les résultats

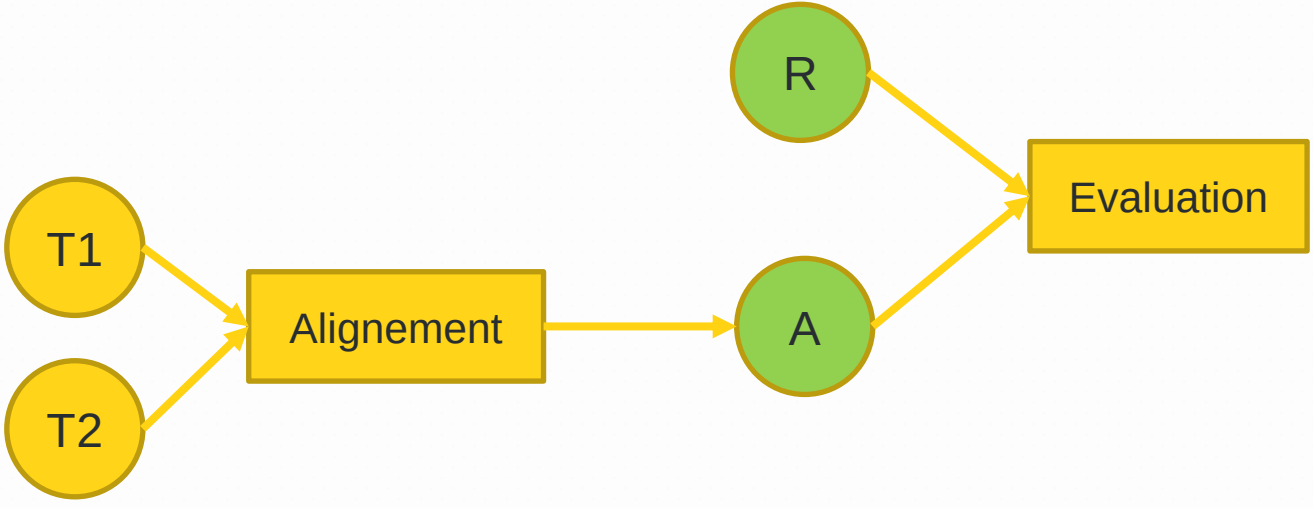
3.

Méthodes d'évaluation des alignements

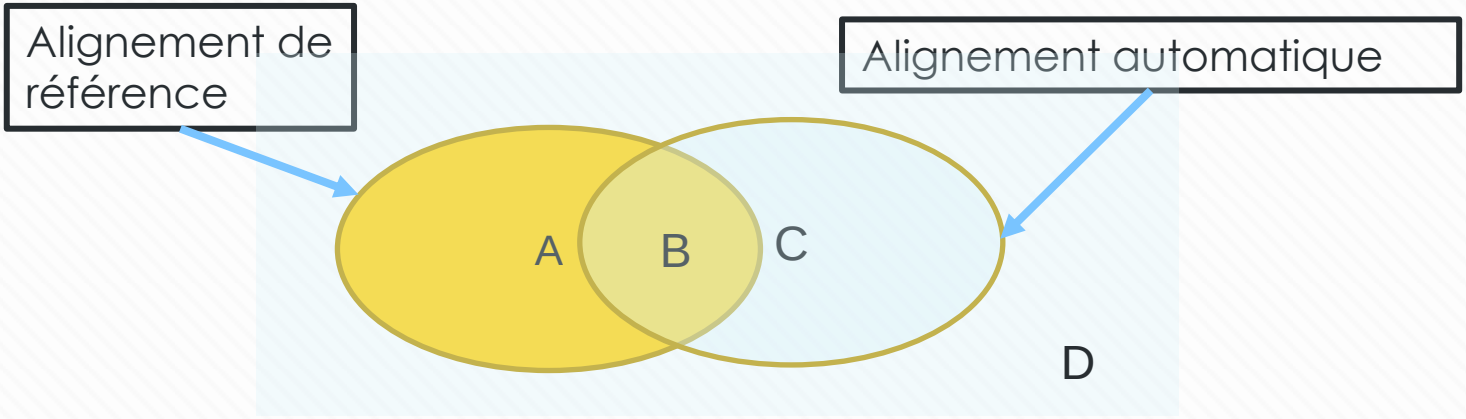
- L'évaluation des alignements consiste à mesurer la qualité des correspondances obtenues.
- Elles peuvent être manuelles ou automatiques.
- L'évaluation manuelle peut être réalisée en demandant à un expert indépendant d'évaluer la qualité de l'alignement.
- L'évaluation automatique consiste à extraire des échantillons et appliquer des mesures comme la précision et le rappel.
- Ces mesures permettent de fournit une approximation de l'exactitude et l'exhaustivité des alignements.

Evaluation automatique entre deux terminologies T1 et T2 :

1. **Extraire un « Gold standard » ou un alignement de référence (R) :** l'ensemble des correspondances entre T1 et T2 effectuées manuellement ou validées au préalable par un expert humain comme des correspondances correctes.
2. **Calculer la « matrice de confusion »** entre l'ensemble de correspondances obtenues automatiquement (A) et l'alignement de référence (R).



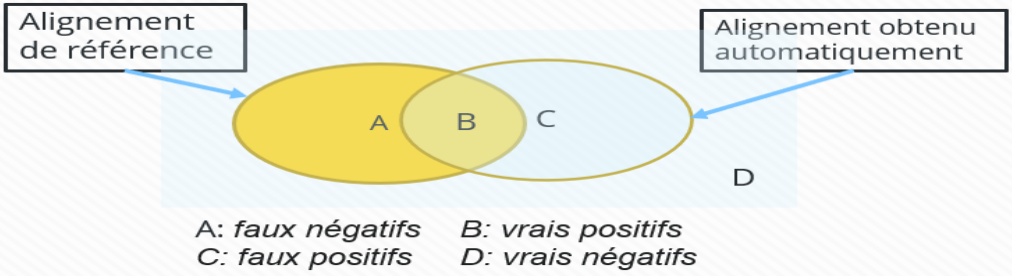
Matrice de confusion



A: *faux négatifs* B: *vrais positifs*
C: *faux positifs* D: *vrais négatifs*

		Alignement automatique	
Alignement de Référence		**Correspondances trouvées**	**Correspondances non trouvées**
	Correspo. Vrais	B	A
	Correspo. Faux	C	D

Précision, Rappel et F-mesure



- **Précision : le nombre de correspondances correctes trouvées par rapport au nombre total de correspondances trouvées.**

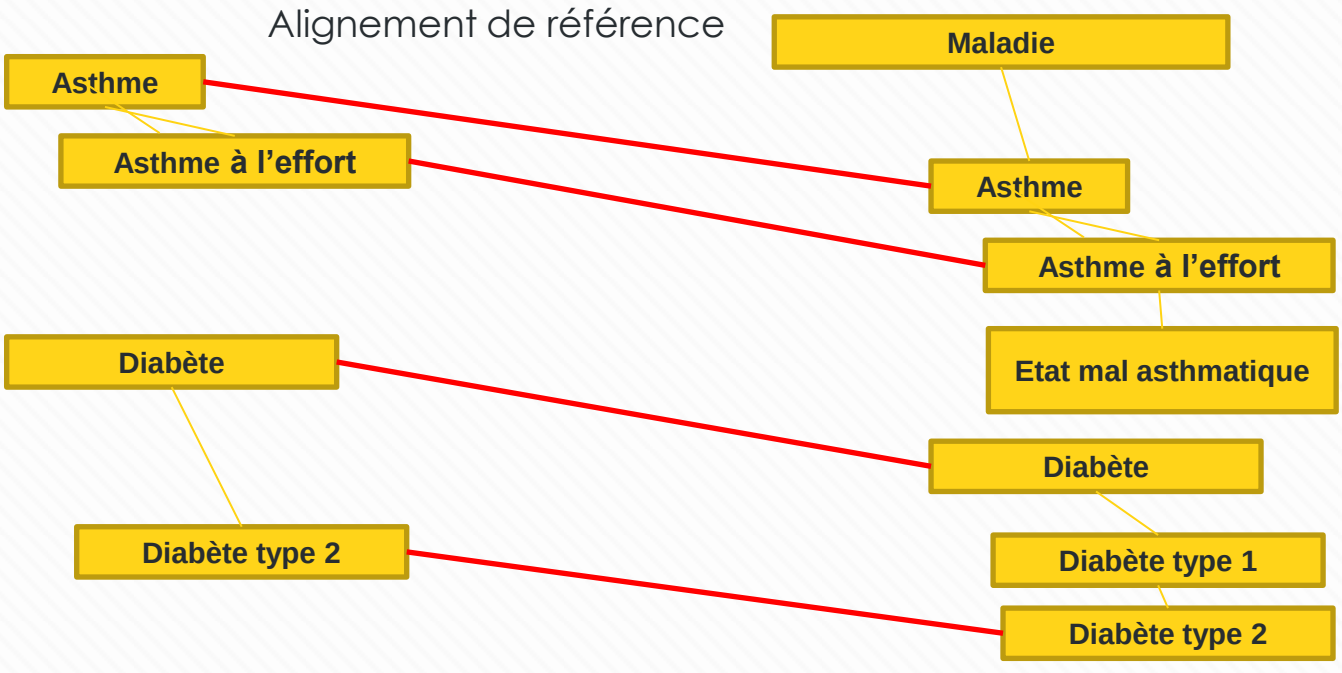
$$\text{précision} = \frac{|B|}{|B|+|C|}$$

- **Rappel : le nombre de correspondances correctes trouvées par rapport au nombre réel de correspondances correctes.**

$$\text{rappel} = \frac{|B|}{|A|+|B|}$$

- **F-mesure** $= 2 * \frac{précision * rappel}{précision + rappel}$

Exercice :



Alignement automatique :

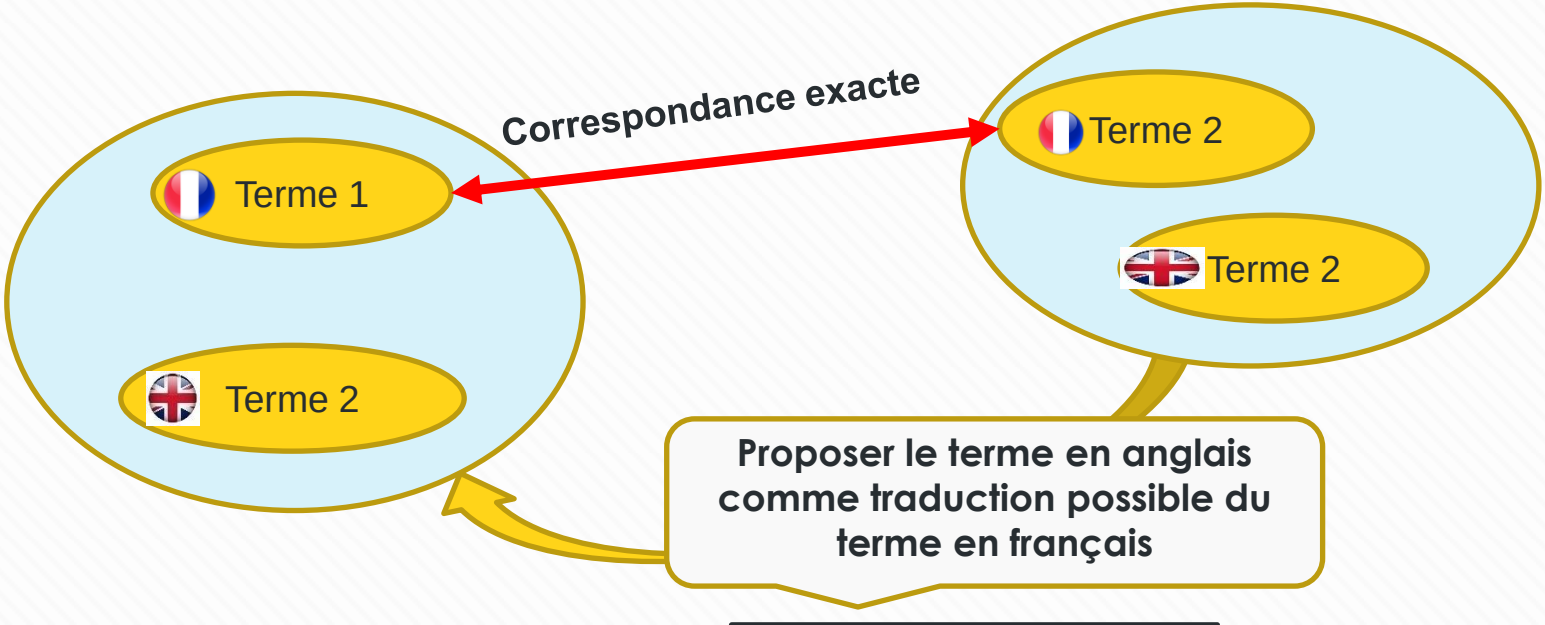
1. asthme -> asthme
2. Diabète type 2 -> Diabète type 2
3. Diabète type 1 -> Diabète type 2

Précision et rappel ?

4.

Aligner pour traduire

- **L'idée :**
  - Utiliser les alignements entre terminologies multilingues pour proposer des traductions manquantes d'une terminologies par rapport à une autre.



Approche UMLS pour traduire vers le français

MÉTATHÉSAURUS DE L'UMLS

Base de données multilingue de l'UMLS

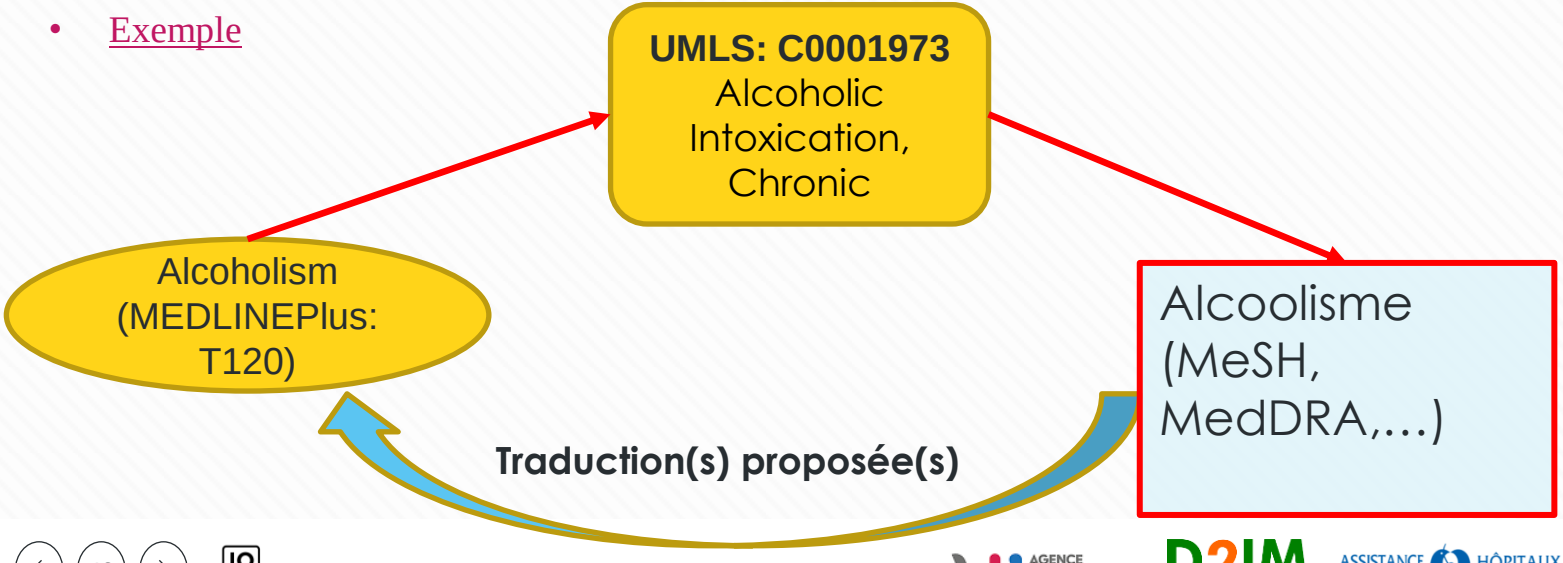
- Information à propos de tous les concepts médicaux et biomédicaux (terminologies, variations lexicales,…)
- Regroupe tous les termes (différentes terminologies, différents noms, différentes langues) partageant un même sens sous un même concept identifié par CUI (Concept Unique Identifier).

Exemple

Les termes : Fibrillation auriculaire, FA, Afib, FIBRILLATION AURICULAIRE sont dans le même CUI (Concept)

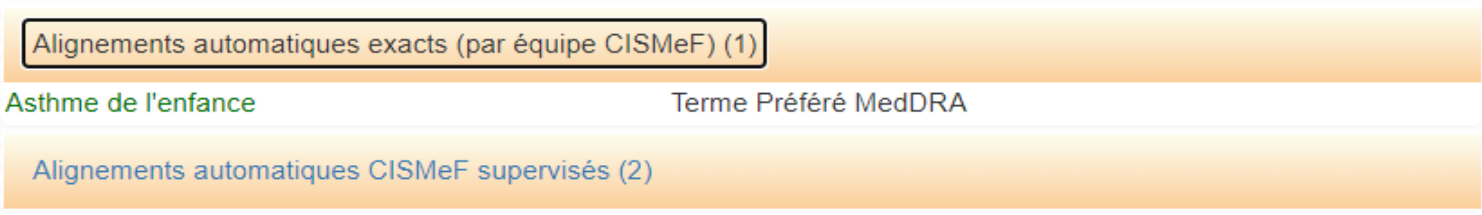
Approche UMLS pour traduire vers le français

- Cette méthode implique que chaque terme à traduire doit être inclus dans le métathesaurus.
- Utiliser les 4 terminologies en français de l'UMLS : MeSH, SNOMED Int., MedDRA, WHO-ART.
- Exemple

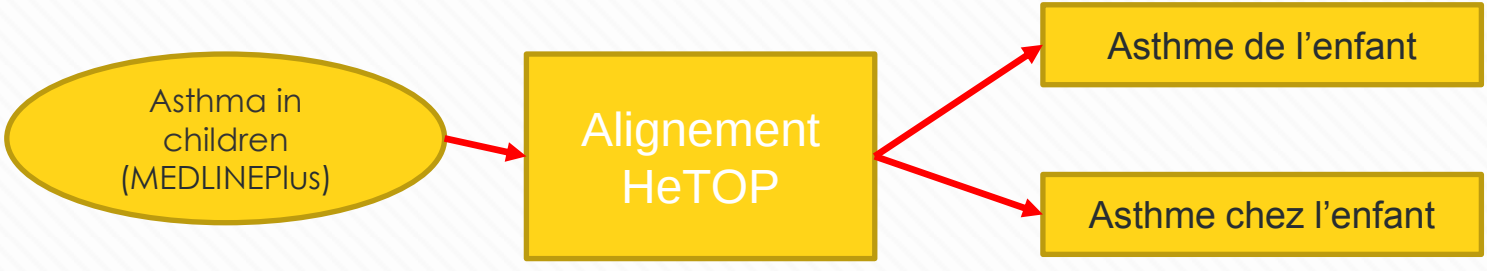


Approche HeTOP pour traduire vers le français et l'anglais

- Utiliser les approches lexicales et manuelles d'alignement développées dans HeTOP pour traduire automatiquement.



- Traduction vers le français et l'anglais.



Evaluation des traductions (BLEU score)

BLEU (Bilingual Evaluation Understudy) : est une mesure qui permet d'évaluer une traduction automatique par rapport à une référence (traduction humaine, ou une autre approche).

$$\mathrm{BLEU} = \mathrm{BP} \cdot \exp\left(\sum_{n=1}^{N} w_n \log p_n\right)$$

$$\mathrm{BP} = \begin{cases} 1 & \text{if } c > r \\ \exp\left(1 - \frac{r}{c}\right) & \text{if } c \leq r \end{cases}$$

Exemple :

Trad1 = On the mat is a cat
Trad2= There is cat sitting cat

Référence = The cat is sitting on the mat

BLEU (Trad1) = 0.454

BLEU (Trad2) = 0.59

```
from nltk.translate.bleu_score import sentence_bleu
reference = [['the', 'cat',"is","sitting","on","the","mat"]]
candidate = ["on",'the',"mat","is","a","cat"]
score = sentence_bleu( reference, candidate)
print(score)


from nltk.translate.bleu_score import sentence_bleu
reference = [['the', 'cat',"is","sitting","on","the","mat"]]
candidate = ["there",'is',"cat","sitting","cat"]
score = sentence_bleu( reference, candidate)
print(score)
```

Exercice :

En utilisant l'UMLS browser

https://uts.nlm.nih.gov/uts/umls/home

Proposer les traductions en français pour les termes :

- « frontotemporal lobar degeneration »
- « central nervous system bacterial infections »

- **Méthodes lexicales**
  - « traduire pour aligner » est souvent une nécessité
  - La question se pose souvent de l'EN vers le FR
    - Récemment: de l'IT vers le FR pour CND / EMDN en Dispositif Médical (EUDAMED)
  - Mais aussi FR vers l'EN lorsqu'on souhaite promouvoir un référentiel à l'international
    - Partie non-tumorale de ADICAP?
    - CCAM dans ICHI?
  - Certaines Terminologies sont nativement « multilingue »
    - ORDO (INSERM/Maladies Rares, 9 langues)
    - WHO-FIC: volonté initiale de l'OMS du multilingue vs. traductions
  - La traduction peut être fastidieuse et empêcher l'évaluation exhaustive de la ressource
  - La traduction automatique est possible
  - Elle porte alors sur des « textes » particuliers:
    - courts et très spécifiques
    - avec parfois des lignes éditoriales ne correspondant pas à la syntaxe du langage courant (LOINC, Cladimed par ex.)

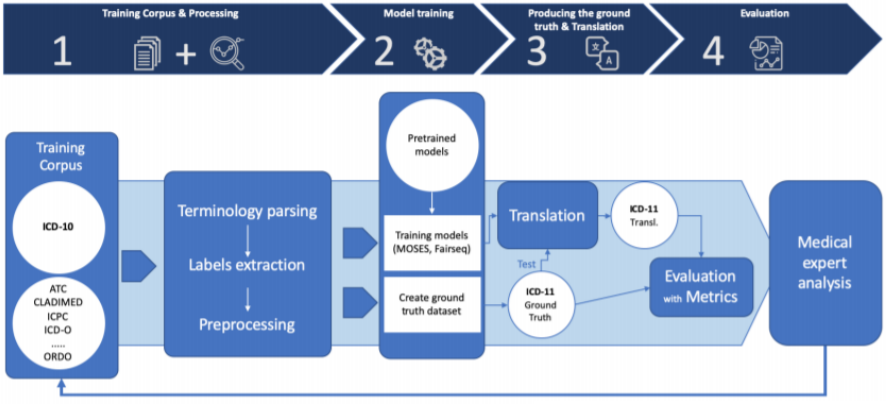


Figure: The proposed machine translation

Evaluation of Machine Translation Methods for Medical Terminologies

Skianis et al.

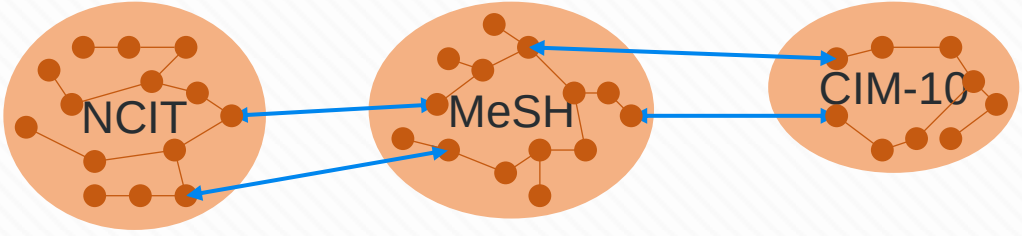
BLEU score range	English label	fairseq proposal (sys6) **Human translations**	Comments
0-0,2	Familial hypophosphataemic rickets	rickets hypophosphatémiques familiaux **Rachitisme familial hypophosphatémique**	Unknown word
	Adult-onset Still disease, buttock	apparition d'un adulte maladie mortelle, fessier **Maladie de Still survenant chez l'adulte, fesse**	Proper name misunderstood/not recognised
	common bile duct blunt injury	lésion de contour du canal biliaire commun **blessure contondante du canal cholédoque**	ambiguity of label (common)
0,21-0,5	Context of assault, gang rivalry	contexte de l'agression, rivalité entre gangs **Contexte d'agression, rivalité entre gangs**	Inappropriate insertion of article
	Barrett adenocarcinoma	adénocarcinome barrett **Adénocarcinome de Barrett**	proper name misunderstood missing coordination term
	Fracture of thumb bone	fracture du pouce **Fracture de l'os du pouce**	missing word
0,51-0,9	ureter cyst	cyste de l'uretère **kyste de l'uretère**	unknown word translated with editorialy very close term
	talipes equinovalgus	pied bot equinovalgus **talipes equinovalgus**	use of correct synonym
	Unintentional exposure to or harmful effects of oxazolidinediones	exposition non intentionnelle ou effets nocifs des oxazolidinediones **Effets nocifs ou exposition accidentelle à des oxazolidinediones**	word order and use of correct synonym

Table: Comparison of translation outputs with human translations.

Obligation de qualifier les traductions avec des experts pour le soin!

5.

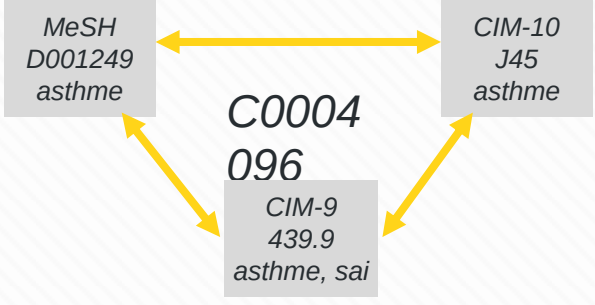
Validation humaine des alignements dans HeTop



Générer automatiquement des alignements :

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

- **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.

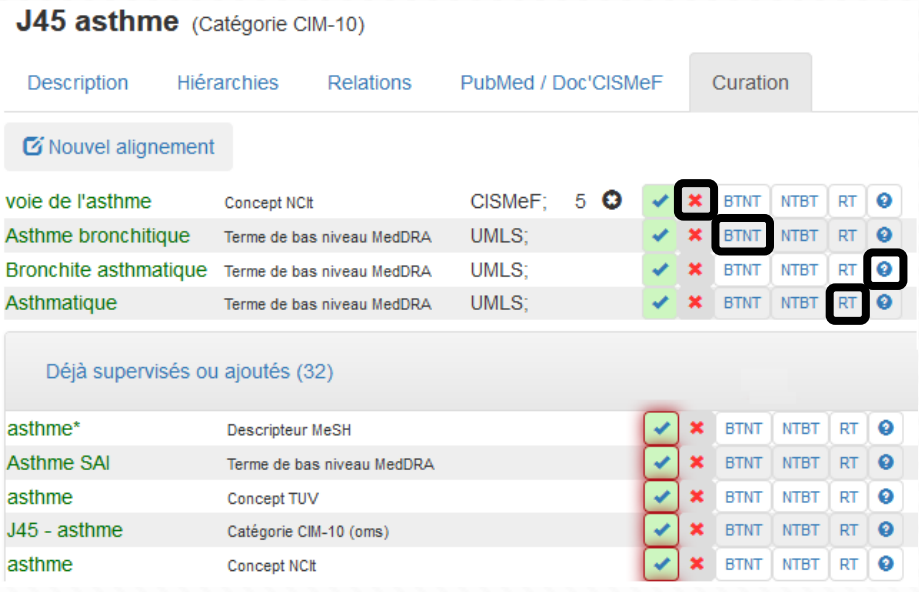
- Par correspondance de termes :
  - identifier des concepts ayant des termes en commun.
  - plus de 200 000 alignements générés automatiquement.

- 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)

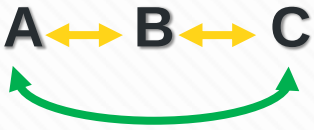
Développement d'une interface dédiée aux validateurs

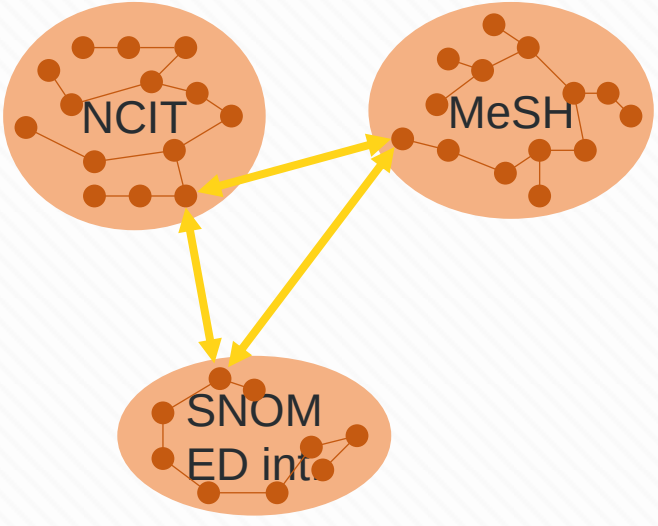
- Alignements exacts A=B
- Alignements faux A≠B
- Alignements BTNT A>B
- Alignements NTBT A<B
- Relations 'Voir aussi' A~B
- Relations inconnues A?B

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



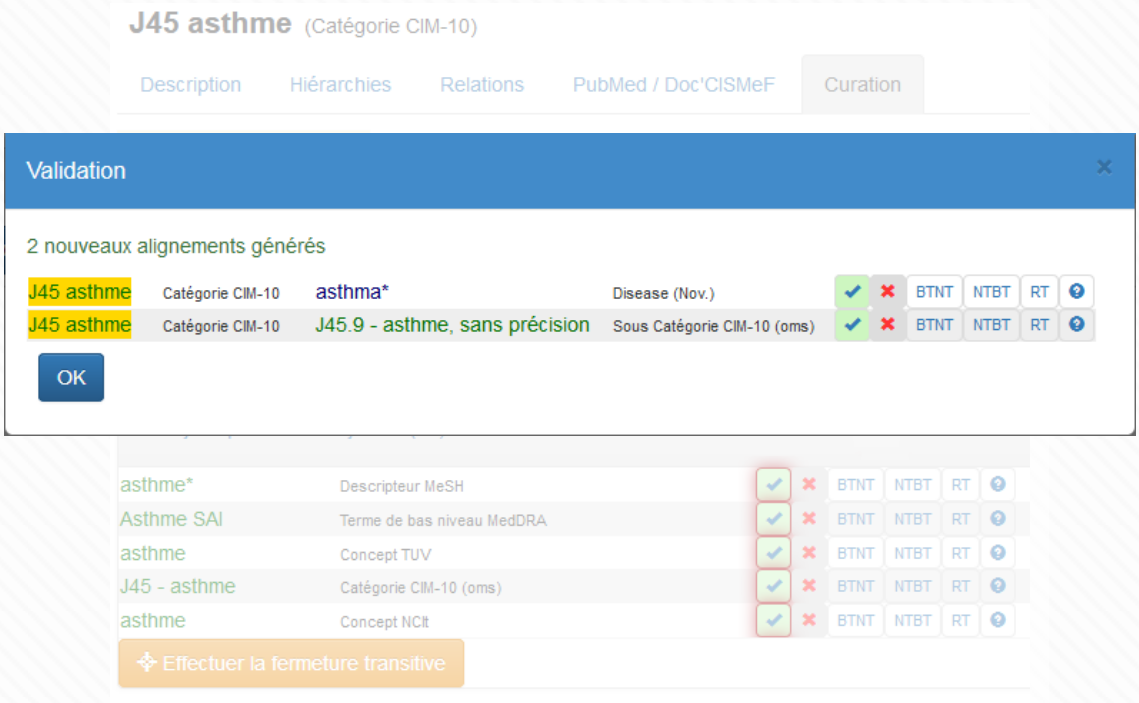
Si A aligné avec B et B avec C
alors A aussi aligné avec C.



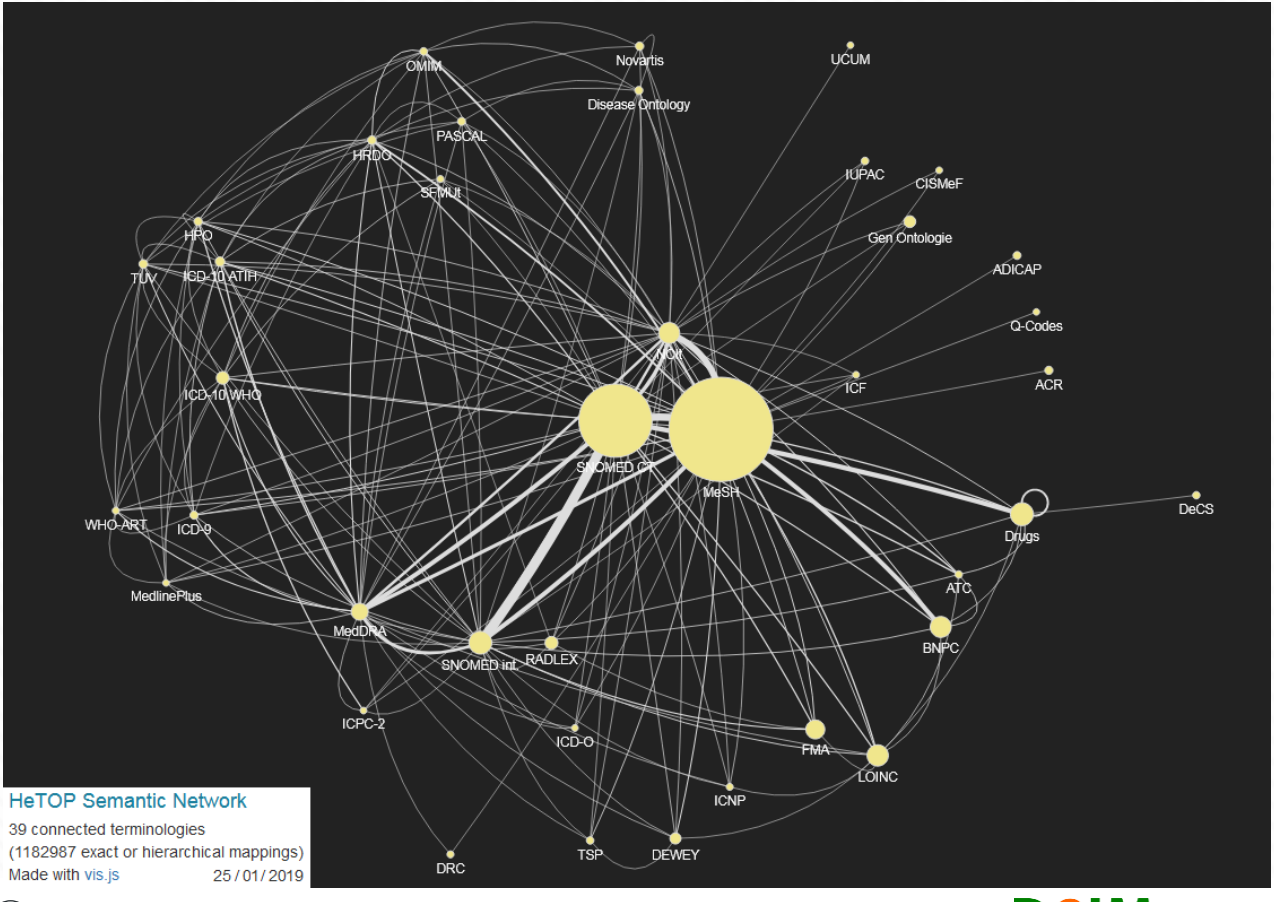


- Créer/valider automatiquement des alignements
- Détecter des anomalies (alignements différents ou faux)

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



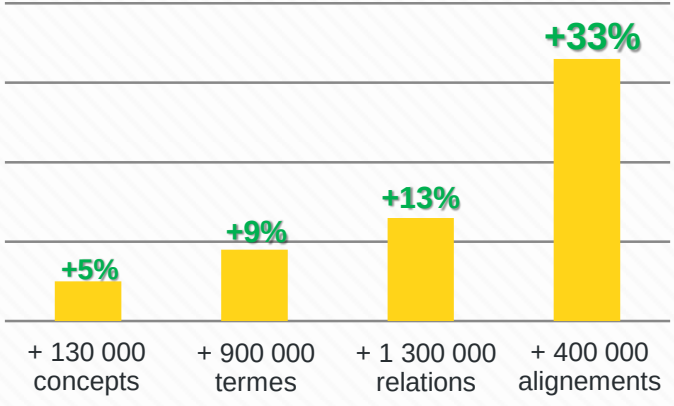
https://www.hetop.eu/hetop/documentation/sem_net.htm



6.

Outils sémantiques: autres cas d'usage et perspectives

Enrichissement du contenu d'HeTOP :



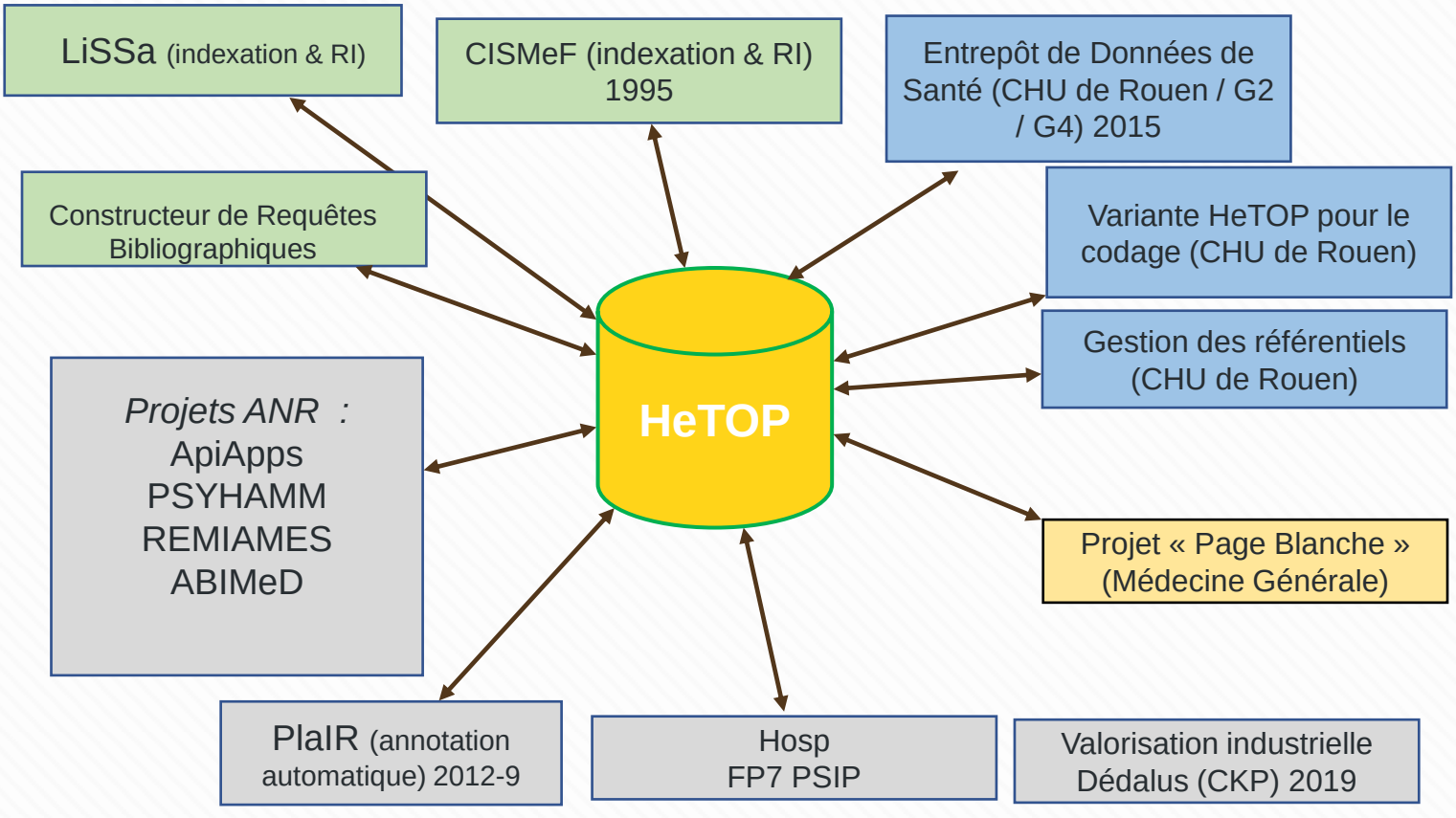
Traductions : 20 000 concepts traduits en français.

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

- **Poursuite des mises à jour des SOC de santé:**
  - CIM 11, ICHI en remplacement de CIM 10, CCAM
  - LOINC (Logical Observation Identifiers Names and Codes)
    - Participation au consortium en charge de la traduction (2021-4)
  - base de gènes NCBI
- **HeTOP 'Drugs' : HeTOP pour les médicaments !**
  - intégration de la base de données DrugBank
  - nouvelle modélisation inter-terminologique
  - création d'une interface spécifique (fiches synthétiques, etc.)
  - URL : www.hetop.eu/hetop/drugs

LOINC v2.69	Fr	En	
Code	55495	94895	
Component	28877	56666	
Property	161	222	
Aspect temporel	35	119	
System	789	3167	
Method	583	1982	
Challenge	540	1871	
Class	118	430	
Part	788	1045	
Hierarchical part	31704	40159	
Total	119090	200556	59.38%

PLATEFORME TERMINOLOGIQUE SÉMANTIQUE



- 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)

- **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**

- **Extracteur de Concepts Multi-Terminologique**

- **Extraire et structurer les informations (évènements/connaissances)**

- **Migration en NoSQL pour des gains de performances d'un facteur entre 10 et 100.**
- **Nouvelles fonctionnalités : négations, affectations de valeurs, filtres, etc.**

- **Deux axes de travaux :**
  - Nettoyage/enrichissement des SOC (bruit/silence) ;
  - Affinement des aspects sémantiques (score), association de concepts, contextes, etc.

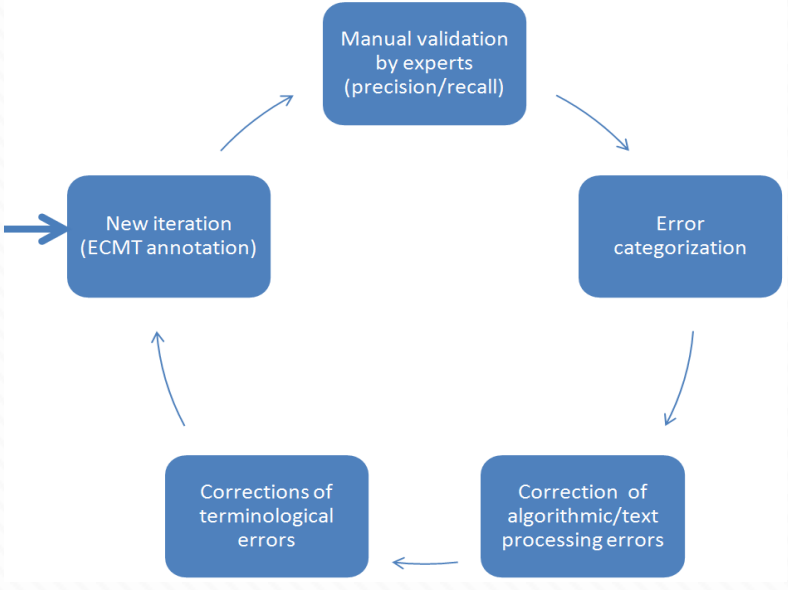
- **Travaux en cours :**

Corpus	P/R 1ère itération	P/R 2ème 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	

- **Études en cours :**
  - Nouvelles itérations sur chaque corpus
  - Corrections terminologiques
  - Segmentation documents
  - Sémantique (prise en compte du contexte)

- **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+

1 **Outils de traduction**

- Production de traductions des termes des SOC sous gestion dans tous les langues requises pour l'analyse

=> Adapter l'analyse au context culturel

2 **Outils d'alignements automatiques**

- Production "massives" d'alignements entre tous les concepts de tous les SOC sous gestion

=> détecter les "Termes **singuliers**" signant l' "innovation terminologique"

3 **Outils d'annotation sur des données massives**

=> Verification des concepts "utiles" embarqués dans les SOC sous gestion

=> Permet d'évaluer les 'nouvelles' Terminologies et/ou la pertinence de Jeux de Valeurs:

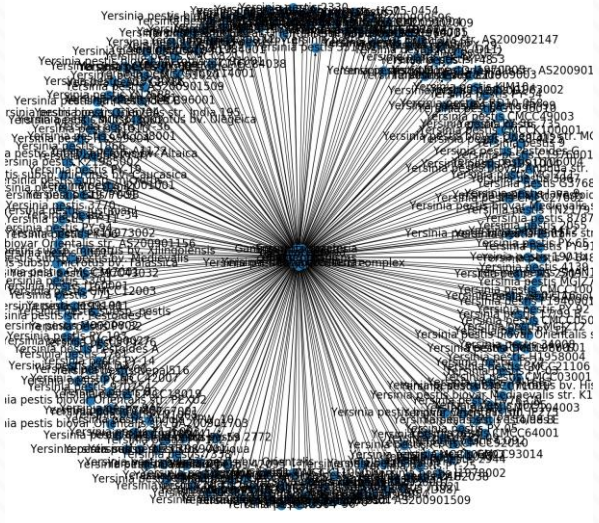
Terminologies	Total	# documents annotés	# concepts utiles	# concepts non détectés	# concepts uniques (« singuliers »)
SNOMED CT	203 282	16 048 939	49 154	154 096	18 923
CIM-10	20 053	1 833 784	4 757	15 296	2 151
CIM-11	55 255	12 925 652	11 904	43 363	6 048
MedDRA	68 137	10 505 572	28 765	39 372	16 370

=> Permet d'orienter les corrections des erreurs Terminologiques (slide précédent).

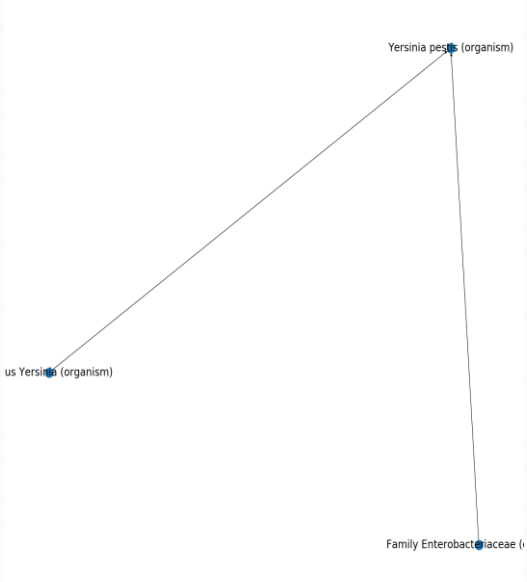
Application de l'analyse de Graphes aux Terminologies:

- **Détection automatique de différence de voisinage par analyse des egonetworks**
- **Selon la méthode: nécessite des alignements de qualité**

Ego-network extracted from NCBI



Ego-network extracted from SNOMED CT



"Yersinia pestis" dans NCBI et dans SNOMED CT (Credits: LIX)

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 serons 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'élement 51 à 60 sur 125 éléments

Précédent 1 ... 5

- Continuer l'automatisation des processus (ETL, générations de traductions/alignements à valider …).
- Améliorer les outils d'aide à la validation

- Continuer les évaluations qualitatives sur différents corpus.
- Développer les systèmes de règles et les usages liés aux contextes.

7.

Conclusion

- julien.grosjean@chu-rouen.fr
- kevin.billey@univ-rouen.fr
- stefan.darmoni@chu-rouen.fr
- Tayeb.MERABTI@esante.gouv.fr
- Florent.desgrippes@aphp.fr