

Extraction de concepts multi-terminologiques

S. Pereira ^{1 2 3 4}, Elisabeth Serrot ⁴, Michel Joubert ³, S.J. Darmoni ^{1 2}

¹ LITIS, INSA Rouen, France

² CISMef, CHU Rouen, France

³ LERTIM, Faculté de Médecine, Marseille, France

⁴ Vidal, Issy les Moulineaux, France

Plan

- Objectifs
- Extracteur multi-terminologique
- État d'avancement (À venir)
- Conclusion

Contexte

■ Indexation :

Définition Wikipédia : « consiste à repérer des mots ou des expressions particulièrement significatifs (appelés **termes**) dans un contexte donné, et à créer un lien entre ces termes et le texte original. »

■ Indexation contrôlée (ou codage): vocabulaire d'indexation déjà défini *a priori*

(ex : thesaurus, nomenclature ou classification)

■ Utilisée pour :

- Convertir un ensemble de données dans un format plus adapté pour la recherche d'informations
- Identifier et décrire le contenu d'un texte, d'un document
- Permettre des traitements informatiques (statistiques etc...)

Contexte

■ CISMef

- Recherche d'information dans les ressources médicales sur le Web
- Indexation manuelle en MeSH

■ Dossier patient

- Études épidémiologiques, objectifs budgétaires, soins du patient
- Codage manuel médico-économique en CIM10 et CCAM
- A venir : SNOMED Int v3.5
- Autres enjeux : Recherche d'information, infobutton

■ Vidal

- Alertes, recherche d'information sur le médicament
- Indexation manuelle avec le TUV

Contexte

- Indexation humaine, automatique ou semi-automatique :

Critères	Humaine	Automatique
Variabilité inter-indexeurs	élevée	aucune
Capacité de synthèse	élevée	faible
Capacité de désambiguïsation	élevée	faible
Connaissances	élevées	faibles
Temps d'indexation	élevé	faible

- Semi-automatique : le compromis

Objectifs

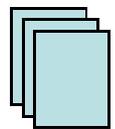
- Créer un outil « générique » destiné
 - à l'indexation automatique :
 - des dossiers patients en CIM10, CCAM et SNOMED
 - des sites médicaux en MeSH
 - des RCP en TUV
 - à intégrer une aide au codage ou à l'indexation de documents (BIBLIS chez Vidal)

FMTI

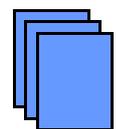
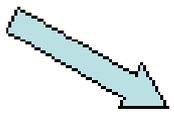
Serveur multi-terminologies

Base de Données multi-terminologique

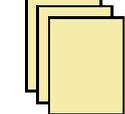
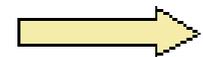
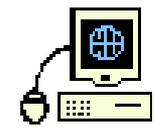
Documents électroniques



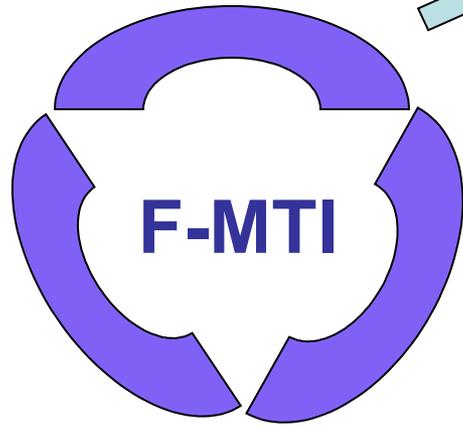
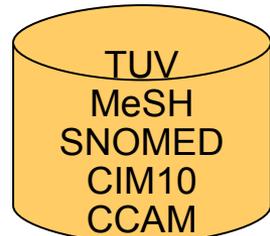
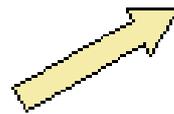
Dossier patient



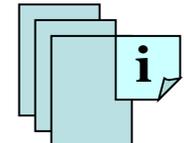
Résumé des Caractéristiques du Produit (RCP)



Sites Web médicaux

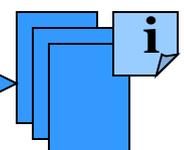


Outil d'indexation automatique multi-terminologique



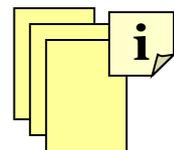
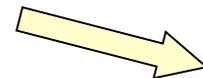
Dossier patient

SNOMED, CIM10, CCAM



TUV

RCP



MeSH

Sites Web médicaux

French Multi-Terminology Indexer : F-MTI

■ Multi-terminologies (inspirée de MTI):

● Indexation directe :

terminologies implémentées :

- CIM10 (classification des maladies)
- SNOMED Int v3.5 (nomenclature de médecine humaine et vétérinaire)
- CCAM (classification des actes)
- MeSH (thesaurus médical)
- TUV (terminologie pour les données thérapeutiques et cliniques pour le bon usage des médicaments)

French Multi-Terminology Indexer : F-MTI

■ Multi-terminologies (inspirée de MTI):

- Projection de chaque terminologie vers les autres terminologies implémentées grâce aux transcodages.

Les différents transcodages utilisés :

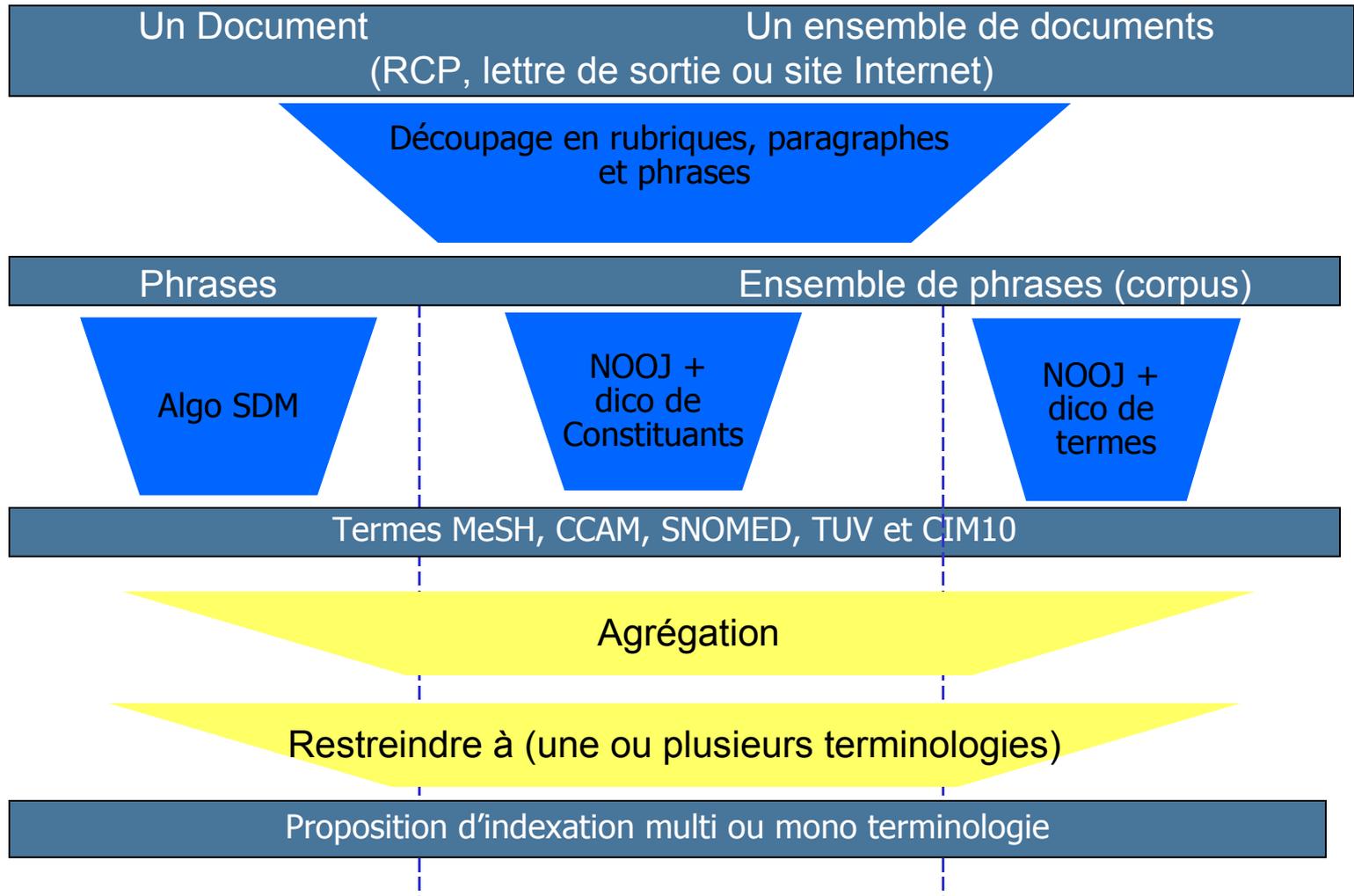
- MeSH, CIM10, SNOMED : UMLS et SFINM
- CCAM-MeSH crée en interne par P.Massari
- CIM10-CCAM de TOTHEM (enlevé)
- TUV-MeSH crée par CISMeF et validé par Vidal
- TUV-CIM10 crée par Vidal (enlevé)

F-MTI

- Multi-documents (plein texte) :
 - Lettres de sortie, sites médicaux, RCP
 - Indexation dépendante du type de document

- Multi-indexations (à paramétrer) :
 - Indexation multi-terminologie
 - Ou indexation mono terminologie

F-MTI



Évaluation de l'outil

■ Évaluation de la méthode du sac de mots :

- CIM10, SNOMED et MeSH
- CCAM pas assez de résultats

■ A venir : Évaluation du dictionnaire de termes et du dictionnaire de constituants

- TUV

■ A venir : Évaluation de la combinaison des méthodes

- TUV

Conclusion

- Implémentation d'un outil d'indexation multi-terminologique automatique et générique.
- Effectif pour la CIM10, CCAM, SNOMED, MeSH, TUV
- En cours d'implémentation et d'évaluation
- Sera intégré comme un outil d'aide au codage (BIBLIS pour Vidal)