

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

PROPOSITION D'UN MODÈLE D'AIDE À LA DÉCISION CLINIQUE POUR  
PROMOUVOIR LA MÉDECINE BASÉE SUR LES ÉVIDENCES. APPLICATION  
À LA PRESCRIPTION TRANSFUSIONNELLE.

THÈSE  
PRÉSENTÉE  
COMME EXIGENCE PARTIELLE  
DU DOCTORAT EN INFORMATIQUE COGNITIVE

PAR  
REDHA ELTAANI

NOVEMBRE 2021

UNIVERSITÉ DU QUÉBEC À MONTRÉAL  
Service des bibliothèques

Avertissement

La diffusion de cette thèse se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 – Rév.04-2020). Cette autorisation stipule que «conformément à l'article 11 du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

## REMERCIEMENTS

Je tiens à remercier tout d'abord mes directeurs de recherche Messieurs. Petko Valtchev et Pierre Poirrier pour leur encadrement et encouragement. Ces années de recherche n'auraient pas été aussi riches sans leurs précieuses discussions, les longues heures d'analyses et les innombrables corrections faites dans une étroite et respectueuse collaboration.

Je tiens à remercier particulièrement mon co-directeur, le Dr Jacques Lacroix pour son encadrement très précieux, ses encouragements et son engagement à m'ouvrir les portes de son unité de soins intensive à l'hôpital Sainte Justine.

Je voudrais remercier également Monsieur Thierry Ducret pour son aide très précieuse, que ce soit pour le suivi, la mise en œuvre ou l'analyse statistique de l'enquête auprès des médecins. Ses conseils et ses orientations ont été d'une grande valeur.

Je tiens à remercier tous les médecins et praticiens de l'hôpital Saint-Justine qui ont donné volontairement de leur temps pour répondre à l'enquête sur la transfusion pédiatrique.

Je remercie ma femme Nabila Chelabi et ma famille particulièrement Sadjia Eltaani pour m'avoir encouragé durant toutes ces longues années de recherche.

## DEDICACE

Cette thèse est dédiée à la mémoire de mon  
père SEDDIK qui voulait me voir poursuivre  
mes études.

À mon cher et regretté frère HAKIM  
ELTAANI qui serait assurément fier de moi.

À ma mère KEDDOURI ZOUBIDA, que dieu  
la garde.

À mes frères et sœurs.

## TABLE DES MATIERES

LISTE DES FIGURES.....	ix
LISTE DES ABRÉVIATIONS, DES SIGLES ET DES ACRONYMES .....	xvi
RÉSUMÉ .....	xvii
ABSTRACT .....	xviii
INTRODUCTION .....	1
CHAPITRE I UN MODÈLE D’AIDE À LA DÉCISION CLINIQUE POUR PROMOUVOIR LA MÉDECINE BASÉE SUR LES ÉVIDENCES. APPLICATION À LA PRESCRIPTION TRANSFUSIONNELLE. ....	4
1.1 Défi de l’application de la MBE.....	5
1.2 Impact de la MBE sur la pratique clinique .....	5
1.3 Les obstacles entravant le recours à la MBE.....	6
1.4 Questions de recherche.....	11
1.5 Un modèle d’aide à la décision intégrant la recherche d’information pour favoriser la MBE.....	11
CHAPITRE II MARCHE ALÉATOIRE SUR DES RESSOURCES TERMINO- ONTOLOGIQUES D’UMLS POUR L’IDENTIFICATION DES CONCEPTS PERTINENTS AU DOMAINE D’APPLICATION D’UN SADC. ....	14
2.1 Motivation.....	15
2.2 Objectif du chapitre .....	15
2.3 Question de recherche.....	16
2.4 Contributions .....	16
2.5 Structure du chapitre.....	17
2.6 Introductions et notions générales .....	18

2.6.1	Terminologie et ontologie.....	18
2.6.2	Ressources termino-ontologiques .....	20
2.7	UMLS : le métathésaurus .....	24
2.7.1	Principales sources vocabulaires et ontologiques dans UMLS.....	25
2.7.2	La structure d'UMLS .....	29
2.7.3	Les relations entre concepts dans UMLS.....	31
2.8	Revue de littérature.....	37
2.8.1	Identification des concepts potentiellement pertinents .....	37
2.8.2	Processus pour l'identification des concepts potentiellement pertinents	43
2.8.3	Notion de concepts pertinents .....	47
2.8.4	Termes de départs pour l'identification des concepts potentiellement pertinents.....	50
2.9	Méthodologie pour identification des concepts potentiellement pertinents .....	51
2.9.1	Quelles parties d'UMLS (la structure).....	54
2.9.2	MetaMapLite.....	58
2.9.3	Algorithme .....	60
2.9.4	Exemple d'identification de concepts .....	69
2.10	Expérimentations et évaluation.....	71
2.10.1	Points importants et limites à prendre en considération pour l'évaluation .....	71
2.10.2	Ressources et prétraitement.....	72
2.10.3	Variable d'évaluation .....	79
2.10.4	Impact du biais de cooccurrence .....	83
2.10.5	Impact du choix des groupes de relations .....	88
2.10.6	Impact du seuil de cooccurrence .....	89
2.10.7	Impact du nombre de marches et du seuil de visite : (2, 4, 6,8).....	92
2.11	Discussion.....	95
2.11.1	Aspects cognitifs .....	99
2.12	Conclusion et travaux futurs.....	105
CHAPITRE III RECHERCHE DE RÉSUMÉS D'ARTICLES PAR L'EXTENSION DE REQUÊTE AVEC DES CONCEPTS ISSUE DE LA MARCHÉ ALÉATOIRE SUR UMLS ET UNE ORIENTATION PICO.....		107
3.1	Contexte.....	108
3.2	Motivation.....	108
3.3	Questions de recherche .....	109

3.4	Contribution.....	110
3.5	Structure du chapitre.....	110
3.6	Introduction et notions générales en recherche d'information .....	111
3.6.1	Recherche d'information (RI).....	111
3.6.2	Évaluation des systèmes de RI.....	114
3.7	Revue de littérature.....	127
3.7.1	Extension de requêtes en RI.....	127
3.7.2	Extension de requêtes en utilisant les ontologies.....	129
3.7.3	Extension des requêtes en utilisant les ontologies dans le domaine médical .....	129
3.7.4	Utilisation des éléments PICO dans la recherche d'information .....	132
3.8	Méthodologie pour extension de requêtes à partir d'extraction ICPU avec ou sans orientation PICO.....	137
3.9	Ressources, prétraitements et préparation .....	142
3.9.1	L'outil ICPU (Identification des concepts pertinents à partir des UMLS) . .....	142
3.9.2	Les « Sujets » de la conférence TREC ad hoc pour Médecine de précision et l'aide à la décision.....	143
3.9.3	La collection de résumés scientifiques PubMed « un instantané » .....	143
3.9.4	Fichier d'appréciation de la pertinence pour les résumés scientifiques.....	144
3.9.5	Indexation de la collection documentaire PubMed à l'aide de « terrier »... ..	145
3.9.6	Extraction des concepts usuellement utilisés dans les éléments PICO des études cliniques enregistrés sur le site Clinicaltrial.gov .....	146
3.9.7	Coder les concepts PICO en utilisant UMLS (Table TrialPico).....	151
3.9.8	Expansion de requête préexistante dans Terrier.....	151
3.10	Expérimentation et évaluation .....	153
3.10.1	Objectif.....	153
3.10.2	Paramétrage pour l'identification des termes d'extension .....	156
3.10.3	Augmenter le poids des concepts d'extension ICPU .....	159
3.10.4	Augmenter le poids des concepts d'extension orientés PICO. ....	161
3.10.5	Différentes variantes pour l'extension des requêtes.....	162
3.11	Résultats et discussion .....	166
3.12	Positionnement des résultats de l'extension ICPU .....	178
3.13	Conclusion et travaux futurs.....	180

CHAPITRE IV ÉVALUATION DE L'AIDE À LA DÉCISION RENFORCÉE PAR UN ACCÈS AUX RÉSUMÉS D'ARTICLES AU MOMENT DE LA PRISE DE DÉCISION CLINIQUE : UNE ÉTUDE COMPARATIVE DANS LE CONTEXTE DE LA MÉDECINE TRANSFUSIONNELLE PÉDIATRIQUE.....	182
4.1 Résumé du protocole .....	184
4.2 Schéma.....	185
4.3 Question de recherche.....	186
4.4 Contexte.....	187
4.5 Contribution.....	189
4.6 Rôle des technologies .....	189
4.7 Aspect cognitif lié à la prise de décision médicale.....	190
4.8 La MBE appliquée à la pratique transfusionnelle.....	194
4.9 Projet de démonstration .....	200
4.9.1 Objectifs .....	200
4.9.2 Hypothèse de base.....	202
4.9.3 Type d'étude.....	202
4.9.4 Critères d'inclusion .....	203
4.9.5 Critères d'exclusion a priori.....	203
4.9.6 Critères d'exclusion a posteriori .....	204
4.9.7 Recrutement et randomisation des sujets .....	204
4.9.8 Randomisation .....	204
4.9.9 Intervention .....	205
4.9.10 Scénarios .....	205
4.9.11 Critères de jugement .....	205
4.9.12 Gestion des données ( <i>Data management</i> ).....	207
4.9.13 Déroulement de l'étude .....	208
4.9.14 Considérations statistiques.....	210
4.9.15 Considérations éthiques .....	210
4.10 Résultats.....	211
4.10.1 Taux de réponse et caractéristiques démographiques des répondants	211
4.10.2 Déterminants de la transfusion de globule rouge par scénario (groupe intervention et groupe contrôle). .....	213
4.10.3 Déterminants pour la transfusion de globule rouge suivant les caractéristiques démographiques et académiques des répondants (analyses exploratoires). .....	218



4.10.4	Déterminants de la prise de décision du groupe avec articles de référence fournis (groupe intervention).....	222
4.11	Discussion.....	226
4.11.1	Aspect cognitif.....	229
4.12	Conclusion.....	231
	CONCLUSION.....	232
	ANNEXE A PRÉSENTATION DU LOGICIEL ICPU.....	238
	ANNEXE B SCÉNARIOS UTILISÉS DANS L'ENQUÊTE (CHAPITRE VI). ..	242
	BIBLIOGRAPHIE.....	249

## LISTE DES FIGURES

Figure	Page
1.1 Modèle d'aide à la décision pour améliorer l'application de la MBE.....	12
2.1 Exemple d'une ontologie, représentation en graphe.....	20
2.2 Les différentes sources vocabulaires qui composent UMLS.....	25
2.3 Exemple de hiérarchie de concepts dans MeSH.....	27
2.4 Exemple de concept et ses descriptions dans US_SNOMED. ....	27
2.5 Les trois principales composantes d'UMLS.....	29
2.6 Intégration des espaces noms entre différentes terminologies. ....	30
2.7 Types de relations dans UMLS.....	31
2.8 Graphe qui représente les relations CHD (is_a) entourant le concept C0005841 : Blood Transfusion.....	33
2.9 Relations hiérarchiques pour le concept : C0005841 Blood Transfusion ...	34
2.10 Relation de type RO entourant le concept C0005841 : Blood Transfusion.	35
2.11 Réseau sémantique d'UMLS (Bodenreider, 2001). ....	36
2.12 Centralité d'un concept (Butt et al., 2016) .....	39

2.13 Diffusion d'information sur un graphe par marche aléatoire (Weert, 2014). .....	45
2.14 Exemple de construction d'une ontologie à partir des termes inclus dans les règles de décisions.....	51
2.15 Processus d'identification des concepts pertinents ICPU.....	52
2.16 Principales tables d'UMLS utilisées par ICPU. ....	54
2.17 Deux ensembles de résultats issus d'un processus d'exploration avec co- occurrence (gauche) et un autre sans cooccurrence (droite).....	65
2.18 Codification selon UMLS du Thésaurus du dictionnaire terminologique de la transfusion sanguine. ....	73
2.19 Arbre de décision TAXI pour la transfusion restrictive.....	75
2.20 Différents tableaux de l'impact de la cooccurrence avec marche aléatoire ; NM=10.....	83
2.21 Impact du biais de la cooccurrence sur l'exploration de l'ICPU ; NM =10.	84
2.22 Différents tableaux de l'impact de la cooccurrence avec marche aléatoire de NM=15.....	85
2.23 Impact du biais de la cooccurrence sur l'exploration de l'ICPU ; NM =15.	86
2.24 Impact du choix des groupes de relation, NM =10.....	88
2.25 Impact du choix des groupes de relation, NM =15.....	89
2.26 Impact du choix du seuil de cooccurrence, NM 10. ....	90
2.27 Impact du choix du seuil de cooccurrence, NM 15. ....	91
2.28 Impact sur le pourcentage des nouveaux concepts retrouvés (NCP).....	93

2.29	Impact du nombre de marches sur le « surplus » de concepts .....	94
3.1	Processus de recherche d'informations (Hiemstra, 2009). .....	112
3.2	Compromis entre efficacité et efficience (Lin et al. (2016)).....	121
3.3	Exemples d'opérations en utilisant le langage de requêtes de terrier.....	125
3.4	Méthodologie d'extension des « sujets » en utilisant ICPU.....	138
3.5	Variantes d'extension suivant les différents groupes de concepts PICO. ...	139
3.6	Processus d'identification des problèmes "autres" fréquemment utilisés des études cliniques.....	140
3.7	Exemple de « sujet » du TREC 2017.....	143
3.8	Exemple de fichier de jugement avec explication ( <a href="https://trec.nist.gov/data/premed/qrels-final-abstracts.txt">https://trec.nist.gov/data/premed/qrels-final-abstracts.txt</a> ). .....	144
3.9	Format du document TREC conventionnel pour la collection .....	146
3.10	Partie d'un fichier XML d'enregistrement d'un essai clinique ( <a href="https://clinicaltrials.gov/ct2/show/NCT03478891?displayxml=true">https://clinicaltrials.gov/ct2/show/NCT03478891?displayxml=true</a> ) .....	149
4.1	Schéma du déroulement de l'enquête.....	185
4.2	Arbre de décision selon le consensus de limitatif TAXI pour la transfusion. .....	196
4.3	Écart entre la pratique réelle et les évidence ((Du Pont-Thibodeau et al., 2014). .....	198
4.4	Informations sur la consultation des résumés d'articles. ....	222
4.5	Appréciations générales de résumés fournis.....	223
4.6	Impact des articles fournis sur la prise de décision. ....	224

## LISTE DES TABLEAUX

Tableau	Page
2.1 Ontologie, Thésaurus et Taxonomie.....	23
2.2 Relations dans le métathésaurus UMLS.....	32
2.3 Exemple de relations avec le concept « Blood transfusion ».....	32
2.4 Extraction des concepts pertinents pour l'ontologie du logiciel (Wu, G. et al. (2008)) .....	38
2.5 Description des colonnes de MRCONSO et un exemple pour le concept « Transfusion ».....	55
2.6 Exemple des différentes descriptions lié aux concepts CUI= C1879316....	55
2.7 Exemple de relation entre deux concepts dans la table MRRELA.RRF .....	56
2.8 Colonnes de la table/fichier MRCOC.....	57
2.9 Concepts issu du Mapping avec MetaMapLite. ....	69
2.10 List de concepts résultants de l'exploration dans UMLS à partir d'un ensemble de termes de départs d'une règle d'un SADC .....	70
2.11 Règles et concepts de départs pour l'ICPU et les différentes expérimentations.....	77
2.12 Table des acronymes et annotation des variables de paramétrage de l'ICPU.....	78

2.13	Exemple d'exploration avec un NM égale à 10.....	79
2.14	Exemple de résultats d'exploration avec les variables d'évaluation pour un NM = 10.....	81
2.15	Exemple du résultat ICPU avec vérification de l'existence des concepts dans le dictionnaire de l'INST. (LM=5; NM=10; SV=6; OC>=1; SAB=MTH, MSH, US_SNOMED). .....	82
2.16	Exemple de concepts supplémentaires potentiellement pertinents issus de l'exploration ICPU avec cooccurrences. ....	87
2.17	Exemples de réseaux cognitifs et leurs applications cognitives (Siew et al., 2019). .....	101
3.1	Exemple de calcul de la P@k pour une requête donnée.....	117
3.2	Précisions moyennes de deux requêtes ayant trois documents pertinents (notés P) et non pertinents (NP) présentés dans des rangs différents ((Abdou et Savoy, 2007). .....	118
3.3	Description des modèles de pondérations utilisés. ....	124
3.4	Catégorie PICO et exemple de concepts y afférents. ....	133
3.5	Règles d'extraction XPath des concepts pour chaque élément PICO .....	150
3.6	Ressources et procédés utilisés par les équipes participantes à la TREC 2017. ....	155
3.7	Résultats de la RI avec les termes initiaux des sujets (modèle DPH, Pub-Med).....	157
3.8	Évaluation de la MAP des paramètres pour le sujet 13 et sujet 1.....	158
3.9	Les différentes variantes d'extension de requêtes utilisées pour l'expérimentation. ....	163

3.10 (Suite) Les différentes variantes d'extension de requêtes utilisées pour l'expérimentation.....	164
3.11 (Suite) Les différentes variantes d'extension de requêtes utilisées pour l'expérimentation.....	165
3.12 Résultats d'évaluation des différents jeux d'extension de requêtes avec un paramétrage d'ICPU suivant : Longueur de marche =6, nombre de marche = 30, seuil de visite = 12, seuil de cooccurrence=6 ; Rela : RO Vocabulaires : MTH, MeSH, SNOMED_CT, NCI.....	169
3.13 Résultats d'évaluation des différents ensembles d'extension, Rela : RO, avec extension supplémentaire de « bo1 ».....	171
3.14 Mesures détaillées et impact sur les « sujets » pour « Qinit + ICPU (p) » avec rela = RO et InL2 .....	172
3.15 Résultats d'évaluation des 15 ensembles d'extension de requêtes avec un paramétrage d'ICPU suivant : Longueur de marche =6, nombre de marche = 30, seuil de visite = 12, seuil de cooccurrence=6 ; Rela : CHD, PAR, RB, RN, Vocabulaires : MTH, MeSH, SNOMED .....	175
3.16 Résultats d'évaluation des 15 ensembles d'extension avec extension supplémentaire de « bo1 ». Rela : CHD, PAR, RB, RN, .....	176
3.17 Mesures détaillées et impact sur les « sujets » pour « Qinit +ICPU (p) + PA (sp) +bo1 » avec rela = Rela : CHD, PAR, RB, RN et InL2 .....	177
3.18 Comparaison aux résultats statistiques par sujet (29 groupes et plus de 125 exécutions) pour 30 sujets pour la recherche des articles dans la littérature. ....	179
4.1 Résumé du protocole.....	184
4.2 Objectifs de l'étude.....	201
4.3 Envoi et réception des questionnaires.....	209
4.4 Caractéristiques des répondants par groupe de randomisation.....	212

4.5	Seuil d'hémoglobine (Hb) dans chaque scénario. ....	214
4.6	Seuil d'hémoglobine (Hb) sélectionné dans chaque scénario par les deux groupes (Khi-2 ou Test Exact de Fisher). ....	215
4.7	Seuil d'hémoglobine choisi en fonction du seuil recommandé par TAXI pour les affections des différents scénarios et en fonction de la randomisation (groupe 1 / groupe 2) ....	217
4.8	Seuil Hb (choisi) vs (non choisi) par groupe. ....	217
4.9	Hémoglobine moyenne et écart-type par groupe. ....	218
4.10	Caractéristiques démographiques des répondants de chaque groupe par rapport au seuil d'hémoglobine moyen. ....	219
4.11	Seuil d'hémoglobine moyen par rapport au rang académique des répondants de chaque groupe. ....	220
4.12	Suivi du cours sur la transfusion et seuil d'hémoglobine moyen. ....	221
4.13	Analyse entre les catégories de réponses au questionnaire de suivi et le seuil d'hémoglobine moyen du groupe d'intervention. ....	225



## LISTE DES ABRÉVIATIONS, DES SIGLES ET DES ACRONYMES

MBE : La médecine basée sur les évidences

SADC : Système d'aide à la décision clinique

UMLS: Unified Medical Language System.

ICPU : Identificateur de concepts pertinents à partir d'UMLS.

CUI : *Concept unique Identifier.*

TAXI : *Transfusion and Anemia eXpertise Initiative.*

PICO : Problème, Intervention, Comparateur et Outcome.

RI : Recherche d'information.

SRI : Systèmes de recherche d'information.

CDSS: *Clinical Decision Support systems*

MAP: *Mean Average Precision*

TREC : *Text REtrieval Conference*

## RÉSUMÉ

La médecine basée sur les évidences (MBE) cherche à améliorer les soins et les traitements dispensés aux patients. Faciliter son application demanderait aux praticiens de se tenir au courant des acquis les plus récents en recherche clinique. Cette nécessité se heurte cependant à plusieurs obstacles tels le manque de temps ou le volume croissant des publications scientifiques. Pour améliorer l'accessibilité aux résultats récents pour les cliniciens lors de l'exercice de leurs activités cliniques, les systèmes d'aide à la décision clinique (SADC) peuvent devenir des vecteurs clés et privilégiés pour instaurer la MBE. Notre modèle permettra à ces systèmes d'opérer une recherche d'information ciblée au moment de la prise de décision clinique et ainsi devenir un instrument clé favorisant cette actualisation des connaissances. Cela permettra aux médecins d'avoir un complément d'information et d'évidences au moment de leur prise de décision. Pour y parvenir, nous présentons une méthodologie qui, à partir des termes qui se trouvent dans les règles de décision d'un SADC, utilise un processus de marche aléatoire afin d'identifier de nouveaux concepts pertinents dans une terminologie médicale (UMLS). Cette marche aléatoire est biaisée par la fréquence de co-occurrences entre les termes de départ et les concepts candidats. Cette méthode est matérialisée par un logiciel (appelé ICPU) générique, lequel est paramétrable pour d'autres domaines médicaux et d'autres utilisations par exemple dans le domaine cognitif. En second lieu, le modèle incorpore une recherche d'information des publications scientifiques plus efficace grâce à l'utilisation des concepts pertinents ainsi retrouvés. Pour y parvenir, une nouvelle approche est proposée pour étendre une requête initiale qui exprime une règle de décision. Cette extension de requête ontologique est renforcée par une pondération qui reflète l'importance des concepts au sein de leur voisinage terminologique. Pour affiner cette recherche d'information, l'extension de requête est soumise à une discrimination des concepts candidats suivant le paradigme PICO qui catégorise les concepts cliniques. Finalement et pour étudier l'impact d'un point de vue décisionnel et cognitif de ce modèle auprès des praticiens, une étude clinique randomisée a été menée au CHU de Saint Justine (Montréal). Les résultats de cette enquête démontrent un intérêt effectif pour ce modèle et un impact réel sur certaines décisions liées à la transfusion. De plus, certains des éléments cognitifs en jeu ont été identifiés pour améliorer ce modèle.

Mots clés : Terminologie, marche aléatoire, cooccurrences, recherche d'information, aide à la décision, extension de requête, étude clinique randomisée, PICO.

## ABSTRACT

Evidence-Based Medicine (MBE) seeks to improve patient's provided care and treatments. Facilitating its application would require keeping practitioners updated regarding recent clinical research publication. This necessity faces several obstacles such as the lack of time or the increasing volume of scientific publications. To improve accessibility to recent clinical results during clinical activities, clinical decision support systems (CDSS) may become a privileged vector for establishing MBE. Our model will allow CDSSs to perform an information retrieval operation at the time of clinical decision-making and thus become an instrument to ensure knowledge updating. Therefore, physicians will have access to additional information and evidence from scientific publications at the time of their decision-making. To achieve this, we present a methodology which, from the terms found in the decision rules of a CDSS, uses a random walk process to identify new relevant concepts in a medical termino-ontology (UMLS). This random walk is biased by the frequency of the co-occurrence between the starting terms and the candidate concepts. This method is implemented in a generic and customized web-based software (called ICPU), which can be customized to serve other medical fields or cognitive process analysis. Furthermore, the model incorporates a more efficient information retrieval for scientific publications that uses the relevant concepts previously found. To achieve this, a new approach is proposed to extend an initial query that expresses a CDSS decision rule. This ontological query extension is reinforced by a weighting that reflects the importance of concepts within their termino-ontological neighborhood. To refine this informational search, the candidate concepts for query extension are discriminated by the PICO paradigm, which helps to optimize information retrieval. Finally, the model's impact on decision-making and cognitive aspects was studied with a randomized clinical study conducted on practitioners at the Sainte-Justine Hospital in Montreal. The results of this survey demonstrate an effective interest in this model and a real impact on certain decisions related to transfusion. Moreover, cognitive aspects were identified to help improve further instances of the model.

**Keywords:** Termino-ontology, graph exploration, random walk, co-occurrences, information research, clinical decision support system, clinical research, query extension, PICO, decision making, randomized clinical study.

## INTRODUCTION

L'intérêt pour la médecine basée sur les évidences (MBE) a connu une croissance fulgurante depuis la création du terme au début des années 90 (Straus et McAlister, 2000). La MBE propose un nouveau paradigme selon lequel un médecin ou un autre soignant appelé à prendre une décision médicale tient compte non seulement de son expérience clinique, de ses connaissances de base et des préférences du patient, mais aussi (en priorité) des évidences disponibles dans la littérature scientifique médicale (Greenhalgh *et al.*, 2014; Sackett, David L *et al.*, 1996). L'application de la MBE s'est traduite par plusieurs succès (Fineout-Overholt *et al.*, 2005; Greenhalgh *et al.*, 2014; Straus et McAlister, 2000) ; le guide des directives pour le traitement de l'asthme établi par la société thoracique britannique en est un bel exemple. Ce document a été mis au point en 1990 par consensus et il se base principalement sur les résultats d'une combinaison d'essais randomisés et d'études observationnelles. Le suivi de ces directives a démontré qu'une utilisation accrue des stéroïdes inhalés (traitement très utilisé actuellement) pour l'asthme, ce qui a fait chuter le taux de mortalité et la comorbidité.

Cependant, l'intégration d'une MBE au moment de la prise de décision clinique n'est pas chose simple. En effet, le praticien qui fait face à un problème/dilemme qui nécessite une alternative thérapeutique à la pratique habituelle ou une meilleure approche pour le bien-être du patient, doit être capable de trouver le temps nécessaire (s'il le peut) pour chercher de nouvelles évidences publiées, identifier les évidences les plus pertinentes pour le cas de son patient et analyser et juger la probable application de ces évidences en tenant compte de l'affection et des préférences du patient (Naik *et al.*, 2002). De plus, cette démarche peut être stressante dans certains milieux cliniques

comme celui des soins intensifs ou des urgences où le temps compte, ce qui limite le temps que l'on pourrait consacrer à une recherche des évidences dans la littérature médicale.

Avec l'avènement de l'informatique dans le domaine médical, les outils d'aide à la décision clinique sont devenus des instruments incontournables pour faciliter l'accomplissement de plusieurs pratiques médicales. Ces outils utilisent des algorithmes qui assistent le médecin dans une ou plusieurs étapes de la décision clinique (Berner, 2007). Plusieurs de ces systèmes se spécialisent dans un domaine en particulier comme la radiologie ou l'interprétation des tracés d'ECG (électrocardiogramme) et n'interviennent que dans une étape précise du raisonnement médical (Berner, 2007). Ces systèmes peuvent devenir des outils facilitant l'intégration des nouveaux résultats de recherche. Cependant, la plupart des SADC contemporains sont majoritairement basés sur des guides de pratique et des évidences formalisées au moment de la création du système (Afzal *et al.*, 2015). Ceci implique que ces systèmes restent figés dans le temps en attendant une mise à jour, laquelle est souvent coûteuse, laborieuse et viendrait probablement en retard pour bon nombre de patients.

Notre recherche vise à doter les SADC d'une capacité de recherche d'information intégrée, et ce, par la proposition de nouvelles approches dans le domaine termino-ontologique et de la recherche d'information par extension de requête. L'objectif étant d'ouvrir ces systèmes sur les nouvelles publications scientifiques en recherche clinique. Ce modèle permettra d'offrir à l'utilisateur l'aide nécessaire pour trouver en temps réel les résultats de recherche clinique les plus pertinents en lien avec le cas du patient traité, mais avec une application spécifique à la prescription transfusionnelle.

Le premier chapitre est une introduction générale à la problématique de l'application de la MBE qui reprend le contexte dans lequel évolue notre recherche, une description

de la question de recherche. Aussi, nous présenterons notre modèle global afin de soutenir les SADC pour une meilleure application de la MBE au quotidien. Les chapitres II et III se structurent d'une façon identique et traitent des sous-questions de recherche chacune dans son domaine scientifique respectif. Chaque chapitre contient en général les éléments suivants : motivation, objectif, question de recherche, contribution, revue de littérature, méthodologie, expérimentation, résultats et discussion. Au chapitre II, nous tisserons un lien entre la contribution proposée et les neurosciences cognitives par réseau afin d'ouvrir la possibilité de développement futurs pour notre application. Le dernier chapitre IV est présenté comme le protocole de recherche qui a été soumis au comité d'étude et de recherche pour validation du modèle que nous proposons. Nous concluons enfin, par l'énumération des différentes contributions que nous espérons avoir apportées en réalisant notre projet, ainsi qu'une discussion sur les défis et les obstacles auxquels nous sommes confrontés.

## CHAPITRE I

UN MODÈLE D'AIDE À LA DÉCISION CLINIQUE POUR PROMOUVOIR  
LA MÉDECINE BASÉE SUR LES ÉVIDENCES. APPLICATION À LA  
PRESCRIPTION TRANSFUSIONNELLE.

### 1.1 Défi de l'application de la MBE

Comme évoquée précédemment dans l'introduction, une MBE permet une utilisation consciencieuse, explicite et judicieuse des meilleures preuves (trouvailles scientifiques émanant surtout de la recherche clinique rigoureuse) disponibles susceptibles de favoriser une meilleure prise de décision et une prise en charge clinique de cas individuels mieux éclairée (Sackett, David L *et al.*, 1996). L'utilisation des nouvelles évidences permet d'aborder un problème médical avec une approche mieux adaptée que celle habituellement appliquée, et ce, pour le bien-être des patients.

### 1.2 Impact de la MBE sur la pratique clinique

Selon Fineout-Overholt *et al.* (2005) des études ont démontré que des patients ayant profité de soins basés sur des évidences ont présenté (au moins) 28 % de meilleurs résultats que ceux dont les soins furent prodigués de façon habituelle. Pour Tonelli, M. R. (2011), la MBE assure une pratique clinique plus uniforme et plus scientifique. Selon Haynes, B. et Haines (1998), l'incorporation des meilleurs résultats de recherche actuels dans la prise de décision clinique permet de raccourcir le délai traditionnel entre la production de données probantes et leur application et d'augmenter la proportion de patients auxquels le meilleur traitement actuel est offert. Toutefois et en dépit de cette belle perspective, plusieurs réserves et obstacles à l'application de la MBE ont été recensés (Fava *et al.*, 2015; Greenhalgh *et al.*, 2014; Haynes, R. B. *et al.*, 2002; Sackett, David L *et al.*, 1996) à l'égard de l'application d'une MBE dans la pratique.



### 1.3 Les obstacles entravant le recours à la MBE

- Délai entre la publication des résultats de recherche scientifique et leur application

La première critique concerne le délai entre la publication et l'application « clinique » de ces résultats de recherche scientifique et leur application. Cet écart qui a été à l'origine du concept de la MBE persiste encore de nos jours, malgré l'amélioration de l'accessibilité à la documentation par voie électronique. Selon Brownson *et al.* (2006), il a fallu en moyenne 17 ans avant d'intégrer dans la pratique médicale seulement 14 % des résultats de recherches originaux. Selon Dr Lacroix, même au bout de ce délai, on constate généralement que l'adoption des informations nouvelles fournies par la recherche ne dépasse pas 50 %, et ce, dans les meilleures circonstances (Lacroix *et al.*, 2007). Néanmoins certaines données suggèrent que ce délai a diminué depuis quelques années, et ce, grâce notamment à l'évolution du web qui a démultiplié l'offre de « preuves » numériquement accessibles. En plus des données probantes et à la littérature bibliographique de base, cette offre inclut aussi des critiques scientifiques, des méta-analyses systématiques, des guides de pratique clinique, des normes de soins, des bases de données comparatives et des bases de connaissances (Georgiou, 2001; Lisa, 2007; Rodrigues, 2000).

- L'énorme volume de preuves et la publication d'un nombre considérable de lignes directrices cliniques sont devenus ingérables.

Selon Greenhalgh *et al.* (2014), le volume de données disponibles est si grand que cela constitue en soi un problème majeur alors que cela devrait en réalité représenter une

opportunité formidable. En particulier, le nombre de lignes directrices cliniques<sup>1</sup> publiées est devenu à la fois ingérable et insondable. En 2005, un audit de 24 heures portant sur 18 patients hospitalisés avec 44 diagnostics a identifié 3679 pages de lignes directrices qu'il serait nécessaire de lire pour traiter ces cas selon les normes proposées, ce qui correspondrait à environ 122 heures de lecture. Dans la pratique, imposer une telle charge de travail aux praticiens n'a aucun sens, surtout quand la décision à prendre doit être déterminée rapidement au chevet d'un malade.

- Comportement des praticiens face à la MBE

Selon Wenger (2012), les médecins ont tendance à préférer de se fier à leurs connaissances tacites acquises par l'expérience qu'aux résultats d'essais cliniques randomisés. Ils estiment que les essais cliniques randomisés mesurent l'effet moyen d'un traitement sur une population sélectionnée ; les résultats et les conclusions de ces études peuvent ne pas correspondre à la situation complexe d'un patient particulier. Selon Dr Lacroix, le dilemme signalé par Claude Bernard (Bernard, 1865) continue de hanter les praticiens : est-ce que la réponse moyenne d'un échantillon de patients retenu dans un essai clinique représente vraiment la réponse d'un patient donné ? Il est évident que l'on pourrait affirmer sans crainte de se tromper que la réponse moyenne trouvée lors d'un essai clinique représente effectivement la réponse d'un patient donné si tous les patients partagent des caractéristiques identiques et présentent les mêmes facteurs de risque (Yusuf *et al.*, 1991). Malheureusement, de nombreux praticiens pensent encore que les résultats de la majorité des essais cliniques<sup>2</sup> ne s'appliquent pas parce que le patient

---

<sup>1</sup> Les lignes directrices de pratique clinique sont des énoncés rédigés systématiquement pour aider le personnel soignant et le patient à prendre des décisions portant sur les soins adéquats dans des circonstances cliniques précises.

<sup>2</sup> Un essai clinique, ou étude clinique, ou encore essai thérapeutique est une étude scientifique réalisée en thérapeutique médicale humaine pour évaluer l'efficacité et la tolérance d'une méthode diagnostique ou d'un traitement.

qu'ils traitent ne correspond pas assez bien aux patients inclus dans l'essai clinique qui pourrait s'appliquer ; en conséquence, ces praticiens refusent d'utiliser les données publiées pour affiner leur pratique clinique parce que d'après eux, leur patient est trop différent.

Aussi et durant leur formation clinique, les médecins sont surtout exposés à des modèles de pratique qui sont basés sur les croyances individuelles de leurs enseignants et sur la pratique populaire à l'époque. Ces modèles peuvent influencer les médecins consciemment ou inconsciemment d'une manière qui peut ne pas concorder avec les preuves contemporaines.

D'autre part, les médecins peuvent ne pas bien se servir des meilleures données disponibles. En effet, certaines publications scientifiques sont meilleures que d'autres et les médecins peuvent adopter une conduite clinique préconisée dans une « preuve » qu'ils trouvent dans les revues qu'ils suivent plus volontiers bien qu'il puisse y avoir une étude de meilleure qualité dans une source différente (Newell et Shanks, 2014). De plus, la quantité de revues médicales est si grande (au-delà de 26 000) qu'il est impossible de se tenir à jour et de connaître tout ce qui est publié.

- Les systèmes d'aide et la MBE

Les progrès des technologies de l'information ont grandement contribué au développement de la MBE. L'informatique est même devenue l'outil par excellence de sa réalisation (Bakken, 2001; Georgiou, 2001). Plusieurs auteurs et institutions scientifiques (Bakken, 2001; Bakken *et al.*, 2004; Lisa, 2007) abordent le rôle de l'informatique dans la promotion et la diffusion de la médecine basée sur les évidences (MBE) et mettent l'accent sur l'aspect crucial et critique de ce rôle, sans lequel la MBE serait toujours à la traîne.

Les systèmes d'aide à la décision font partie de ces outils technologiques. Ils peuvent aider les médecins dans l'exercice de leurs fonctions (Jaspers *et al.*, 2011), car ils analysent les données des patients et automatisent les aspects de la prise de décision clinique. En général, ces systèmes contiennent des règles explicites et reproductibles construites et entretenues par l'entremise de bases de données, de lignes directrices ou de guides médicaux (Rodrigues, 2000). Toutefois, convaincre les médecins de l'efficacité de ces systèmes peut constituer en lui-même un défi, car bon nombre de ces médecins sont réticents à changer leur pratique (Sim *et al.*, 2001).

Malgré l'apport indéniable de ces systèmes à la pratique médicale, l'utilisation « mécanique » de ces outils peut perturber le raisonnement clinique d'un praticien (Lisa, 2007). En effet, un clinicien inexpérimenté peut s'aligner mécaniquement sur les suggestions d'un outil d'aide à la décision simplement afin d'éviter tout risque de litige subséquent.

Pour ne pas étouffer le développement d'une expertise clinique plus nuancée qui exploiterait le cumul d'une expérience pratique éclairée, la tolérance à l'incertitude et la capacité d'appliquer un jugement pratique, des systèmes plus « ouverts » à la MBE peuvent devenir des solutions plus adaptées à la réalité clinique et à la prise de décision. Nous proposerons notre vision du modèle adéquat dans la section 1.5 plus loin dans ce chapitre.

- La MBE dans la pratique transfusionnelle

La transfusion de globules rouges vise à prendre en charge des patients anémiques ou souffrant d'une perte de sang aiguë. Par exemple, la réanimation d'un patient souffrant d'une hémorragie importante comprendra non seulement l'apport liquide de remplacement intraveineux, d'oxygène et d'autres médicaments, mais aussi la transfusion de globules rouges, accompagnée de transfusion de plasma et de plaquettes si l'hémorra-

gie est particulièrement grave. La décision de transfuser un ou plusieurs produits sanguins doit être étayée par la nécessité de soulager les signes et symptômes cliniques du patient et de prévenir une morbidité ou une mortalité significative. Cette transfusion doit être prescrite uniquement lorsque les avantages pour le patient sont susceptibles de l'emporter sur les risques.

La médecine transfusionnelle souffre, comme d'autres domaines de la médecine, d'un retard important dans l'application de la MBE. Le bénéfice de la MBE n'est pas à remettre en cause, car certaines données ont déjà démontré son efficacité et son intérêt dans certaines circonstances particulières. En effet, l'implantation d'un système d'aide à la décision qui réagissait à toute prescription d'une transfusion de globules rouges en proposant au besoin une stratégie restrictive (MBE) quand cela était approprié a permis de réduire les transfusions de globules rouges de 69,0 % à 43,4 % ( $p \leq 0.005$ ) et les transfusions de plaquettes de 41,9 % à 31,2 % ( $p = 0,16$ ) chez des adultes hospitalisés (Butler *et al.*, 2015). La conclusion de cet article fut que l'introduction d'un SADC portant sur les transfusions, appuyée par une éducation avisée et une rétroaction en temps réel aux médecins prescripteurs a eu un impact immédiat sur l'amélioration de la conformité aux lignes directrices soutenant une pratique transfusionnelle restrictive.

## 1.4 Questions de recherche

Nous avons entamé nos travaux de recherche avec la question suivante :

Dans un contexte de soins complexes caractérisé par la surcharge de travail, les praticiens ne disposent que de peu de temps pour prendre connaissance des nouvelles publications scientifiques applicables aux situations de leurs patients. Dès lors, appliquer de la MBE telle qu'énoncée paraît difficile. La question primordiale serait de savoir comment peut-on améliorer l'intégration de cette MBE au plus près de ces praticiens et au moment même de la prise de décision clinique en lien avec leurs patients ?

## 1.5 Un modèle d'aide à la décision intégrant la recherche d'information pour favoriser la MBE

Le modèle que nous préconisons pour répondre à la question de recherche s'inscrit dans le cadre des nouveaux systèmes d'aide à la décision dits « Système d'apprentissage rapide » (*Rapid Learning Health System*) » (Lambin *et al.*, 2013). Ces nouveaux modèles essaient de rapprocher la recherche clinique et les soins médicaux prodigués aux patients. Cet échange se traduit par l'accroissement des transferts (rapidité d'apprentissage) entre le milieu de la recherche et celui de la pratique, et vice versa. L'information sur les nouvelles découvertes parvient rapidement à l'utilisateur (le praticien dans notre cas) et l'application des nouvelles recommandations peut faire l'objet d'un retour vers la recherche pour évaluation. Nous proposons d'ouvrir les systèmes d'aide à la décision (à base de règles) sur les publications scientifiques au moment de la prise de décision en respectant le modèle ci-dessous (Figure 1.1).

La réalisation d'un tel modèle passe par le développement des capacités du système d'aide à la décision pour faire une recherche d'information ciblée à partir des termes qui apparaissent dans ses règles de décisions. Aussi, l'accès à des résumés d'articles de publication scientifique au moment de la prise de décision fera l'objet d'investigations.

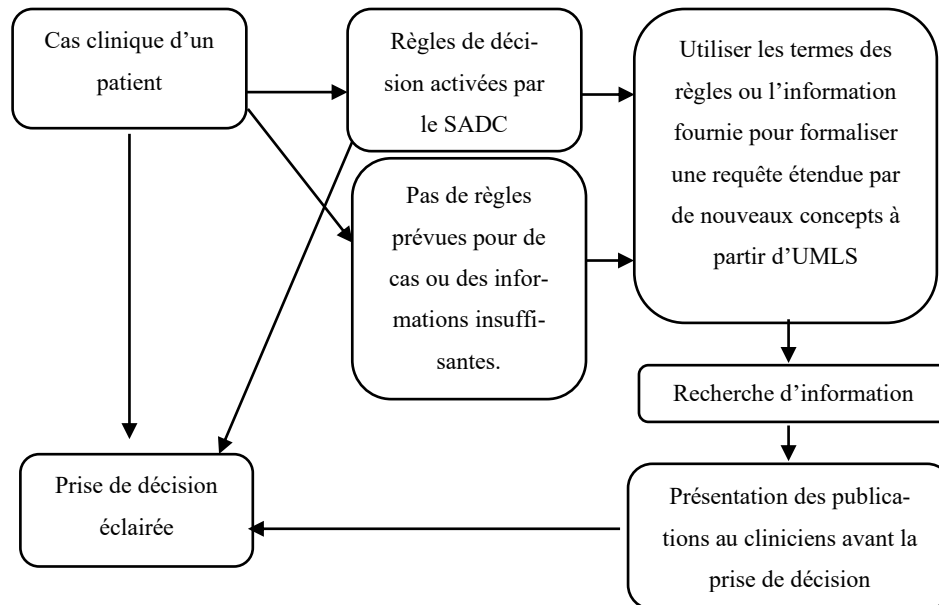


Figure 1.1 Modèle d'aide à la décision pour améliorer l'application de la MBE

Notre travail de recherche doit pouvoir répondre aux sous-questionnements suivants :

1. Dans la perspective d'améliorer la recherche d'information à partir d'un ensemble restreint de termes qui apparaissent dans les règles de décision pour une tâche médicale, serait-il possible d'identifier de nouveaux concepts pertinents qui appartiennent au domaine d'application, et ce, en utilisant des ressources termino-ontologiques comme UMLS (système de vocabulaire médical uniformisé) ?
2. Peut-on améliorer la pertinence des résultats de la recherche d'information avec une extension de requête qui exploite les concepts identifiés dans l'étape précédente

et une pondération reflétant l'importance de ces concepts dans leur entourage termino-ontologique ?

3. Mener cette extension de requête en choisissant les nouveaux concepts candidats suivant leur appartenance à l'une ou l'autre des catégories du modèle PICO (Problème, Intervention, Comparateur, Outcome — voir section 3.7.4), pourrait-il améliorer cette recherche d'information ?
4. Se pourrait-il qu'un accès aux résumés d'articles scientifiques particulièrement pertinents, au moment de la prise de décision et en situation d'incertitude, permettrait d'améliorer l'adhérence des praticiens aux nouvelles recommandations dans la médecine transfusionnelle pour des patients traités dans un service de soins intensifs pédiatriques ?

Chacune de ces sous-questions a fait l'objet de travaux de recherche dont les résultats sont rapportés dans les chapitres qui suivent.



## CHAPITRE II

MARCHE ALÉATOIRE SUR DES RESSOURCES TERMINO-ONTOLOGIQUES D'UMLS POUR L'IDENTIFICATION DES CONCEPTS PERTINENTS AU DOMAINE D'APPLICATION D'UN SADC.

## 2.1 Motivation

Comme il a été évoqué dans le chapitre précédent, les systèmes informatiques, et notamment les outils d'aide à la décision, peuvent s'avérer des vecteurs influant la qualité des soins offerts aux patients, en particulier si ces systèmes sont capables d'être proactifs et assez rapides pour accélérer en ce qui concerne la dissémination des nouvelles publications scientifiques de qualité au moment même de la prise de décision clinique par les médecins traitants.

Doter ces systèmes d'aide à la décision d'une assise conceptuelle plus étendue en utilisant des concepts issus des ontologies, en lien avec leur logique décisionnelle de départ, permettra à ces systèmes de passer d'un cadre de programmation rigide vers un cadre où la recherche d'articles scientifiques se fait de façon autonome et intégrée, ce qui devrait améliorer l'application de la médecine basée sur les évidences MBE. Nous présenterons dans ce chapitre les premières démarches que nous avons accomplies afin de parvenir à ces fins.

## 2.2 Objectif du chapitre

L'objectif de ce chapitre est de présenter une approche qui permettra d'étendre l'assise conceptuelle (concepts) qui entoure une tâche médicale particulière, et ce, par l'identification des nouveaux concepts ontologiques pertinents pour le domaine d'application (des connaissances encadrées dans UMLS). Il en résultera un ensemble de « grappes de concepts candidats » susceptibles de permettre une utilisation polyvalente comme l'extension de requête pour la recherche d'information sur des articles scientifiques.

L'approche adoptée explore le métathésaurus UMLS en utilisant une marche aléatoire modifiée et orientée par la cooccurrence entre les concepts de départs et les concepts candidats pour l'exploration. Ce processus stochastique permet de dégager après un

certain nombre d'itérations, les concepts les plus centraux entourant les règles de décision d'un SADC à partir des ressources termino-ontologiques d'UMLS.

### 2.3 Question de recherche

*Dans la perspective d'améliorer la recherche d'information (publications scientifiques) à partir d'un ensemble restreint de termes qui apparaissent dans les règles de décision pour une tâche médicale, serait-il possible d'identifier de nouveaux concepts pertinents qui appartiennent au domaine d'application, et ce, en utilisant des ressources termino-ontologiques comme UMLS (système de vocabulaire médical uniformisé) ?*

### 2.4 Contributions

Notre principale contribution dans ce chapitre est une méthode pour l'identification des concepts importants entourant le contexte terminologique et ontologique d'une tâche médicale spécifique, ainsi que le développement d'un outil logiciel pour la matérialiser. Cet outil a été pensé d'une façon générique et applicable à d'autres types d'utilisation ; pas forcément pour un seul domaine médical ou uniquement pour des systèmes d'aide à la décision, car l'exploration ontologique de ce logiciel ne se limite pas aux termes de départ des règles de décision, du texte libre peut être utilisé comme point de départ de l'exploration sans aucun prétraitement de ce dernier.

Cet outil d'exploration ontologique pourrait servir à d'autres tâches comme la possibilité pour les ontologistes d'extraire des sous-ensembles ontologiques de départ à moindre effort pour bâtir des ontologies plus grandes, juste en fournissant des concepts clés (la transfusion, la génétique, une maladie particulière, une partie anatomique en particulier...).

## 2.5 Structure du chapitre

La première partie du chapitre présente une introduction aux notions importantes comme la terminologie, l'ontologie et de ressources termino-ontologiques (vocabulaire contrôlé, taxonomie, thésaurus et ontologie). Puis, une description détaillée du méta-thésaurus UMLS qui représente la ressource termino-ontologique majeure utilisée dans le contexte de nos recherches. Suivra une revue de littérature qui abordera les thèmes de l'identification des concepts potentiellement pertinents dans une ontologie, les processus et les méthodes d'exploration sur des graphes de représentation de connaissances, notamment la marche aléatoire, la notion de concept pertinent. Cette revue permettra de mettre en lumière quelques travaux d'intérêt ayant été réalisés dans des domaines limitrophes. Après cette partie, nous introduirons notre méthodologie pour l'exploration et l'extraction des concepts pertinents associés aux termes figurant dans les clauses des règles de décisions d'un SADC. Dans cette section, toutes les étapes de raisonnement ainsi que l'algorithme issu de cette approche seront détaillées (quelques parties se trouveront dans les annexes pour alléger le texte). Les résultats des expérimentations et les différents constats seront présentés dans la dernière partie et commentés en conclusion. L'analyse de l'apport cognitif de ce module d'identification des concepts pertinents sera discutée à la fin de ce chapitre.

## 2.6 Introductions et notions générales

### 2.6.1 Terminologie et ontologie

La terminologie s'intéresse aux mots qui désignent des concepts (notions) permettant d'appréhender les objets du monde (Roche, 2005b, 2006). On désigne également par terminologie, l'ensemble des termes rigoureusement définis et qui sont spécifiques à une science, une technique ou un domaine particulier de l'activité humaine. La construction de terminologies vise généralement à combler un vide dans un domaine spécialisé (Zweigenbaum, 2004). Au sens classique, une terminologie liste les termes d'un domaine, et pour chacun d'eux, propose une fiche qui en décrit les usages, la (ou les) signification(s), ainsi que les relations entretenues avec des termes sémantiquement et/ou syntaxiquement proches (Reymonet et al., 2007). Dans le contexte médical (Zweigenbaum, 2004), les terminologies sont généralement organisées selon une approche terminologique classique avec à la clé pour chaque terminologie un inventaire de concepts ou de classes dans le domaine considéré. Puis, à chaque concept est associé un libellé ou terme qui l'exprime dans une langue donnée (ici, le terme est envisagé comme une unité lexicale<sup>3</sup> [une étiquette linguistique] servant à matérialiser le véritable objet de la terminologie, à savoir le concept (L'homme, 2005)). À ce terme normatif (terme préféré, ou terme vedette), il peut s'ajouter des dénominations supplémentaires, considérées comme synonymes et qui participent à la normalisation du sens des termes qui expriment le concept. Finalement, un code ou un identifiant unique est associé à chaque concept pour le désigner indépendamment. Cette approche conceptuelle de la terminologie place le concept au cœur de la terminologie comme un élément entrant

---

<sup>3</sup> Unité lexicale (Linguistique) : Unité de sens et de son figée dans une langue, comme un mot, un groupe de mots, ou un proverbe, y compris toutes les formes flexionnelles. Par exemple : « jeune », « jeune homme » et « les voyages forment la jeunesse » sont des unités lexicales.

dans une structure de connaissances. Une définition servira à le fixer à l'intérieur du domaine de connaissances et à le baliser par rapport aux autres concepts avec lesquels il entretient un ensemble de relations logiques ou ontologiques (L'homme, 2004; L'homme, 2005).

Comme les terminologies, les ontologies en ingénierie des connaissances sont une réponse à un besoin relativement similaire, celui de la normalisation des communications et des échanges, mais dans un langage formel. En effet, Gruber (1995) définit l'ontologie comme étant une spécification formelle et explicite d'une conceptualisation partagée d'un domaine d'intérêt. Pour Guarino *et al.* (2009) les ontologies sont un moyen de modéliser formellement la structure d'un système, c'est-à-dire les entités et relations pertinentes qui émergent de l'observation [ce qui « existe »] et qui sont utiles à des fins spécifiques. L'ontologie est donc, pour l'ingénierie des connaissances, un objet informatique exploité par les terminologistes (Roche, 2006). Pour Després et Szulman (2008), les concepts de l'ontologie et les relations conceptuelles les associant sont construits à partir d'unités termino-ontologiques figurant dans les réseaux termino-ontologiques<sup>4</sup>. Ces concepts ontologiques sont décrits dans un langage formel, organisés dans une structure hiérarchique, liés par des relations conceptuelles et contraintes par des règles et des axiomes.

---

<sup>4</sup> Un réseau termino-ontologique est construit à partir de l'interprétation, dans un cadre applicatif, des unités linguistiques (termes et relations lexicales les liant) constituant le réseau terminologique. Il est constitué d'unités dites termino-ontologiques (concepts terminologiques et relations sémantiques les reliant).

Ci-dessous (Figure 2.1) un exemple d'ontologie représentée en graphe et montrant certaines des classes en chimie médicinale ainsi que les relations hiérarchiques de type « is\_a » qui les relie (Gómez-Pérez *et al.*, 2013). En résumé, la terminologie est à la fois une pratique et une science. Elle relève de la linguistique par la prise en compte des langues d'usage et s'en distingue par la modélisation scientifique de la signification des mots; alors que l'ontologie en tant qu'artéfact et son ingénierie en tant que méthode formelle de modélisation, constituent des voies pour la construction et la représentation du système notionnel des terminologies (Roche, 2005a).

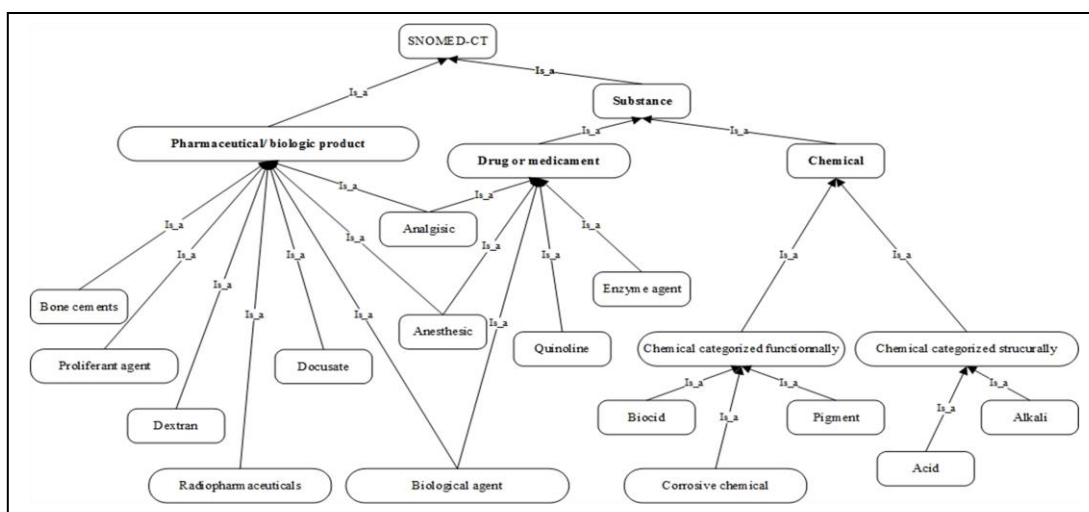


Figure 2.1 Exemple d'une ontologie, représentation en graphe.

## 2.6.2 Ressources termino-ontologiques

Nous utilisons dans ce chapitre la notion de ressources termino-ontologiques pour englober toutes les ressources telles que les vocabulaires contrôlés, les taxonomies et les thésaurus (*Ontology-like Vocabulaires*) ainsi que les ontologies (Dramé *et al.*, 2014; Mougin, 2019; Roussey *et al.*, 2011). Ces ressources sont les principales composantes d'UMLS objet de notre exploration pour l'identification des concepts pertinents (voir la section 2.9) et voici quelques définitions importantes pour expliquer et souligner la différence entre ces ressources.

Un vocabulaire contrôlé fait référence à un ensemble de termes établi par un groupe (une communauté de pratiques) afin de pouvoir labelliser des contenus (Arsenault, 2006; Harpring, 2010). La signification des termes n'est pas forcément définie et il n'y a pas nécessairement d'organisations logiques des termes entre eux. « *Pediatric Terminology* » est un exemple de vocabulaire contrôlé qui peut être consulté sur le lien suivant : (<https://www.cancer.gov/research/resources/terminology/pediatric>).

Avec la taxonomie, on passe d'un vocabulaire contrôlé vers un vocabulaire organisé. En effet, dans une taxonomie, tous les termes d'un vocabulaire contrôlé sont organisés sous une forme hiérarchique unique. Cette hiérarchisation correspond souvent à une spécialisation avec des relations « parent/enfant » ou « sens plus larges/sens plus étroits » avec d'autres termes (Hedden, 2010). Il existe donc un lien précis entre un terme du vocabulaire et ses enfants. Ces liens donnent un sens supplémentaire, une signification au vocabulaire et la structure est parfois appelée « arbre ». L'ajout de synonymes peut ou non faire partie d'une taxonomie. Exemple : La taxonomie du « National Center for Biotechnology Information (NCBI) » comprend des noms d'organismes et des classifications pour chaque séquence de la séquence nucléotidique et protéique (<https://www.ncbi.nlm.nih.gov/taxonomy>).

Un thésaurus est plus une liste organisée de manière formelle afin de rendre explicites les relations préalables entre les termes normalisés (appelés descripteurs) représentant les concepts d'un domaine de connaissance (Aitchison *et al.*, 2003). Ces relations sont de type sémantique comme des relations d'équivalence intralinguistique (synonymie) ou des relations d'équivalence interlinguistique (traduction) ainsi que des relations hiérarchiques et des relations d'association (Zacklad, 2007). Le thésaurus est usuellement utilisé par un agent humain (documentaliste, spécialiste) pour indexer des documents ou comme vocabulaire contrôlé. Un exemple de thésaurus inclus dans UMLS est celui du « *Medical Subject Headings* » produit par « The National Library of Medicine ». (<https://www.nlm.nih.gov/mesh/meshhome.html>)



Un thésaurus ou même une taxonomie sont des formes d'ontologie dont la grammaire n'a pas été formalisée. Lorsque l'on établit une catégorie et une hiérarchisation de cette catégorisation, on établit des dépendances entre ces termes. Ces hiérarchisations ont un sens en dehors du vocabulaire lui-même. Pour finir, une ontologie formelle correspond à un vocabulaire contrôlé et organisé avec une formalisation explicite des relations décrites dans un langage de représentation des connaissances. Voir le tableau 2.1 comparatif de Thomas (2013).

Tableau 2.1 Ontologie, Thésaurus et Taxonomie.

	Ontologie	Thésaurus	Taxonomie
Définition	L'ontologie cherche à décrire de façon formelle un domaine de connaissance, en identifiant les types d'objets de ce domaine, leurs propriétés et leurs relations.	« Liste organisée de termes contrôlés et normalisés (descripteurs et non-descripteurs) servant à l'indexation des documents et des questions dans un système documentaire »	« Taxonomie » est utilisée pour désigner la « science de la classification » et par extension tout système de classification/catégorisation.
Contient	Des classes, des propriétés, et des règles logiques formelles. Éventuellement des instances de classe.	Des concepts et des termes, organisés entre eux, avec leurs libellés, leurs traductions, leurs synonymes, et leurs descriptions/définitions.	Des catégories organisées hiérarchiquement
Utilisation	Sers à instancier et à raisonner	Sers à indexer des contenus ou des ressources avec des mots-clés et à les rechercher (avec ces mêmes mots-clés)	Sers à classer, à ranger des contenus ou des ressources
Niveau de formalisme logique	Très formel (formalisme mathématique)	Peu formel	Moyennement formel (peut-être plus formel qu'un thésaurus, dans le sens où la hiérarchie des entrées peut suivre certaines contraintes.)
Niveau de proximité avec la langue naturelle	Très éloigné de la langue naturelle (utilise des identifiants techniques pour s'abstraire du langage naturel)	Proche de la langue naturelle (donne des équivalents linguistiques de chaque entrée, des traductions dans d'autres langues)	Pas particulièrement en lien avec la langue naturelle.
Types de relations utilisées	Inclusion (classe/sous-classe) ; Opérations ensemblistes : union, intersection, exclusion ; Caractéristiques des propriétés : domaine, ensemble d'arrivée, transitivité, propriétés inverses, etc.	Hiérarchiques et associatives ; éventuellement relations d'alignement	Hiérarchiques seulement

## 2.7 UMLS : le métathésaurus

« UMLS » est l’acronyme pour « *Unified Medical Language System* » qui renvoie vers un effort initié par la NLM<sup>5</sup> afin de regrouper sous un même espace de nombreuses sources vocabulaires et terminologiques spécialisés dans le domaine médical. L’objectif de cette entreprise est de permettre une interopérabilité des concepts médicaux entre les systèmes informatiques (Neoh et Na, 2005). Cette interopérabilité d’UMLS permet de mettre en lien des informations de santé, des termes médicaux, des noms de médicaments et des codes de facturation pour des systèmes d’information biomédicaux hétérogènes (hôpital, cabinet médical, pharmacie, compagnie d’assurance...). Plus précisément, UMLS est une interconnexion de systèmes termino-ontologiques (Zweigenbaum, 2004). Comme définition, un système (ou plus exactement un réseau) termino-ontologique est construit à partir de l’interprétation, dans un cadre applicatif, des unités linguistiques (termes et relations lexicales les liant) formant des concepts terminologiques et des relations sémantiques les liant (Després et Szulman, 2008).

Cette initiative d’alignement termino-ontologique à grande échelle permet à UMLS d’être considéré comme la plus large ressource actuellement disponible pour la médecine. Elle est le résultat de la fusion de plus d’une centaine de thésaurus, dont elle préserve les réseaux de relations entre les termes (Tchechmedjiev et Jonquet, 2017). UMLS peut ainsi jouer dans le domaine biomédical un rôle que l’on peut comparer à celui de WordNet pour la langue générale (Fellbaum, 1998). Parlons de rôle comparable, mais non de possible remplacement (Burgun et Bodenreider, 2001).

---

<sup>5</sup> Cet ensemble de connaissances terminologique a été développé et maintenu par la Bibliothèque nationale de médecine (NLM) depuis 1986 aux États-Unis.

Il est important de mentionner qu'UMLS n'est pas une ontologie formelle, mais c'est un métathésaurus qui s'en approche avec son réseau sémantique (Kashyap et Borgida, 2003). Pour être considéré comme une ontologie formelle, il doit être mis dans un système de représentation des connaissances (logiques de description, graphes conceptuels...). Ce passage vers une ontologie formelle demande un nouvel effort de modélisation (Zweigenbaum, 2004). Au vu de nos objectifs de recherche, nous exploitons la structure (UMLS) sous sa forme native sans aucune modification et sans utiliser un langage formel pour la manipulation.

### 2.7.1 Principales sources vocabulaires et ontologiques dans UMLS

UMLS contient environ 3,67 millions de concepts ainsi que 14 millions de noms de concepts uniques issus de 203 vocabulaires sources (version 2018AA). Ces sources incluent de nombreux types vocabulaires comme des vocabulaires cliniques (Figure 2.2)

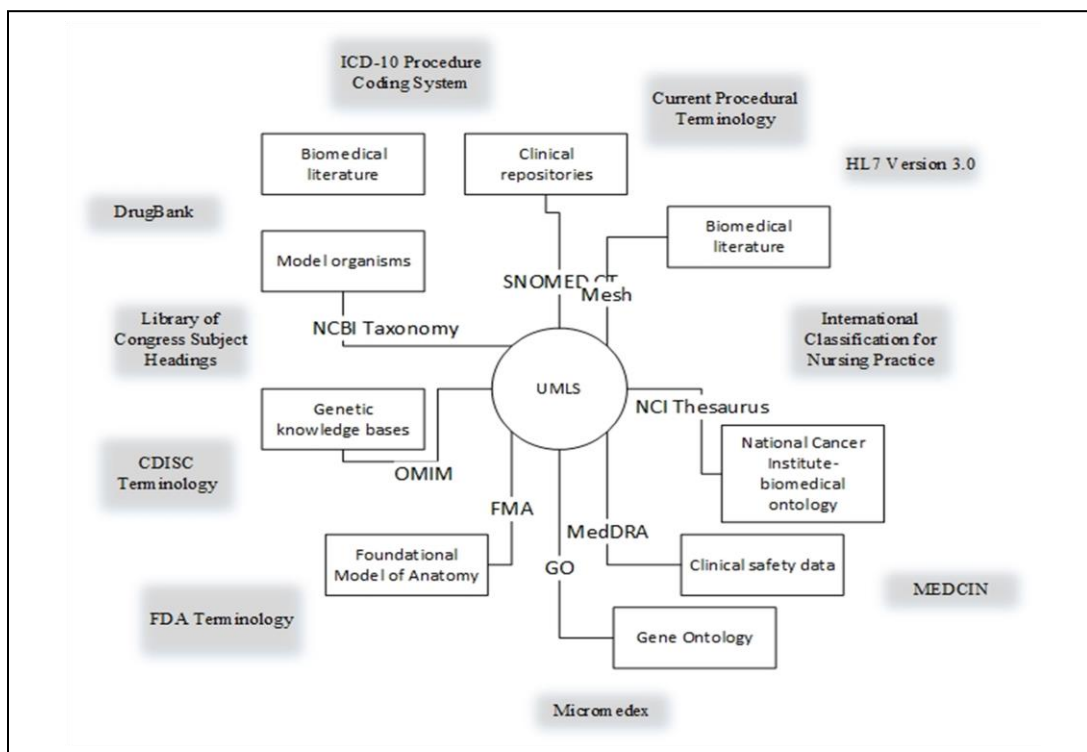


Figure 2.2 Les différentes sources vocabulaires qui composent UMLS.

, des classifications, des ensembles de codes administratifs et des thésaurus utilisés pour indexer et récupérer des publications scientifiques (NIH, 2019) ainsi que des ontologies. Voici des descriptions de quelques sources vocabulaires qui font partie d'UMLS (à titre non exhaustif) :

**MeSH** (*Medical Subject Headings*) est un important vocabulaire contrôlé. Ce dernier est maintenu par la « U.S. National Library of Medicine <sup>6</sup> ». Ce thésaurus est généralement utilisé pour donner une uniformité et une cohérence à l'indexation de la littérature biomédicale, et ce en utilisant une catégorisation par un entête sujet (semblables aux mots-clés dans d'autres systèmes) (Merabti *et al.*, 2009). Cette catégorisation permettra aux systèmes de recherche documentaire, tels que PubMed<sup>7</sup>, de fournir une recherche par sujet (<https://www.nlm.nih.gov/mesh/introduction.html>).

Tous les termes dans MeSH sont classés dans des catégories hiérarchisées en arborescence et qui se composent de 16 grandes branches (Anatomie, organismes, maladies, produits chimiques et médicaments, voir la (Figure 2.3). UMLS contient l'ensemble du vocabulaire MeSH avec ces 29.351 descripteurs et 161 qualificatifs (sous-titres) (<https://www.nlm.nih.gov/research/umls/sourcereleasedocs/current/MSH/stats.html>).

---

<sup>6</sup> La « United States National Library of Medicine » (NLM, Bibliothèque américaine de médecine) est une bibliothèque spécialisée en médecine et dans les sciences et techniques associées. Elle est rattachée aux Instituts américains de la santé. Les principaux produits fournis et maintenus sont : PubMed/Medline, MeSH, UMLS et MedlinePlus.

<sup>7</sup> PubMed : est une ressource documentaire gratuite qui soutient la recherche et la récupération de la littérature biomédicales et des sciences de la vie évaluée par des pairs dans le but d'améliorer la santé, à la fois au niveau mondial et personnel. La base de données PubMed contient plus de 30 millions de citations et résumés de littérature biomédicale évaluée par des pairs.

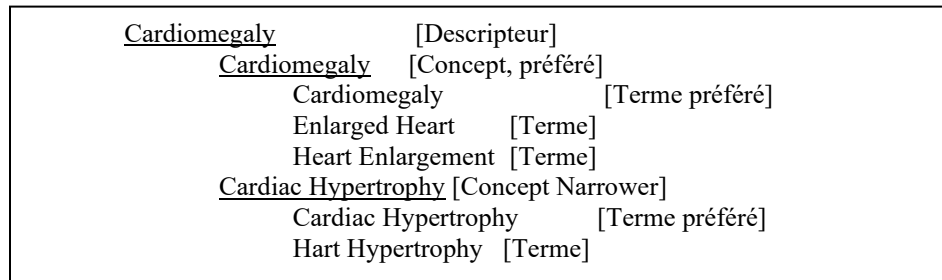


Figure 2.3 Exemple de hiérarchie de concepts dans MeSH.

**US\_SNOMED CT** (*Systematized Nomenclature of MEDicine*). Ce vocabulaire contrôlé est une terminologie clinique normalisée et multilingue qui vise à améliorer les soins aux patients grâce à la mise au point de systèmes permettant de consigner et de concilier avec précision les consultations de soins entre les cliniciens et autres fournisseurs dans le domaine de la santé (<http://www.snomed.org/>). En tant que terminologie la plus complète au monde, SNOMED CT contient plus de trois cents milles concepts cliniques actifs, chacun représentant une signification clinique unique. Ces concepts sont organisés en hiérarchies suivant le découpage : procédure, anatomie du corps, produit pharmaceutique ou biologique (<http://www.snomed.org/>).

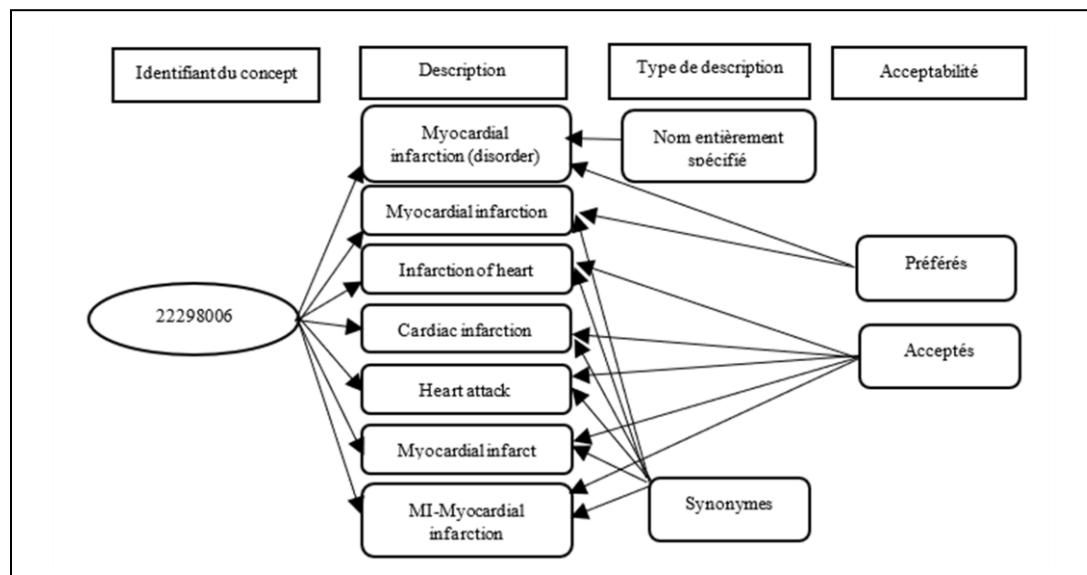


Figure 2.4 Exemple de concept et ses descriptions dans US\_SNOMED.

La figure 2.4 présente un concept « infarctus du myocarde » dans SNOWMED CT avec les différentes descriptions et leurs acceptabilités (description considérée comme acceptée ou préférée).

**NCI Thesaurus (NCIt)** est la terminologie et l'ontologie de référence de NCI (National Cancer Institute<sup>8</sup>). Le contenu de « NCI Thesaurus » provient du NCI, des NIH<sup>9</sup>, d'agences et d'organisations de développement de normes telles que la FDA<sup>10</sup> et le CDISC<sup>11</sup>. NCI couvre la terminologie pour les soins cliniques, la recherche fondamentale, ainsi que l'information publique et les activités administratives. « NCI thésaurus » est considéré comme un standard largement reconnu pour le codage et le référencement biomédical. Son contenu est organisé dans un environnement logique de description richement structuré, avec plus de 400 000 relations entre concepts pour aider à les définir et à les connecter. Les concepts individuels incluent des codes, des termes, des abréviations, des synonymes, des définitions, des liens vers des sources extérieures et une grande variété d'autres informations utiles.

---

<sup>8</sup> Le National Cancer Institute (NCI) est le principal organisme fédéral de recherche et de formation sur le cancer. Établi en vertu de la « National Cancer Institute Act » de 1937, le NCI fait partie des « National Institute of Health » (NIH), l'une des 11 agences qui composent le ministère de la Santé et des Services sociaux (HHS).

<sup>9</sup> National Institutes of Health (NIH) sont aujourd'hui l'un des centres de recherche médicale les plus importants au monde. Une agence du ministère de la Santé et des Services sociaux, le NIH est le point focal fédéral pour la recherche en santé aux États-Unis.

<sup>10</sup> La Food and Drug Administration (FDA, « Agence américaine des produits alimentaires et médicamenteux » 1) est l'administration américaine des denrées alimentaires et des médicaments.

<sup>11</sup> Le CDISC (*Clinical Data Interchange Standards Consortium*) est une organisation de normalisation (SDO) traitant des données de recherche médicale liées aux soins de santé, afin de « permettre l'interopérabilité des systèmes d'information afin d'améliorer la recherche médicale et les domaines de soins de santé connexes ».

UMLS contient aussi d'autres vocabulaires contrôlés, des taxonomies et des ontologies consultables sur le lien suivant : (<https://www.nlm.nih.gov/research/umls/sourcereleasedocs/index.html>).

### 2.7.2 La structure d'UMLS

UMLS comprend trois principales composantes (Figure 2.5), à savoir : le Métathésaurus, le réseau sémantique, le lexique du spécialiste.

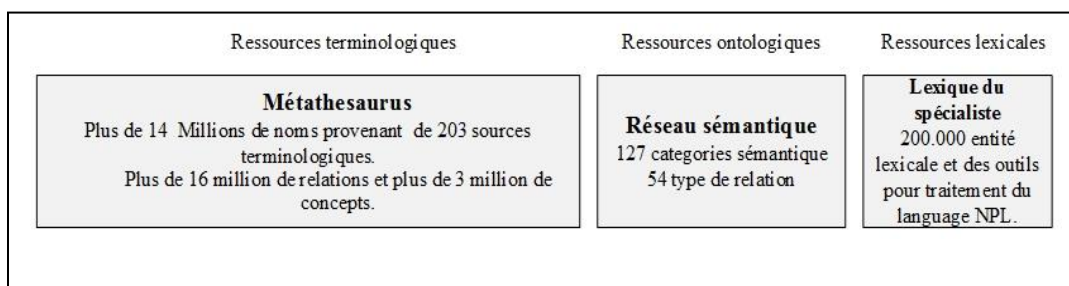


Figure 2.5 Les trois principales composantes d'UMLS.

Le Métathésaurus ou « MTH » est un grand thésaurus biomédical organisé par concept et reliant les termes similaires provenant de sources vocabulaires différents et de classifications contrôlées (SNOMED, MeSH, NCBI...) pour désigner un seul et unique concept. L'objectif étant de permettre une interportabilité des concepts entre différentes terminologies.

Le métathésaurus identifie également les relations utiles entre les différents concepts qui le composent et préserve leurs significations d'origine. Comme évoqué, le « concept » est l'unité de base de l'information dans Métathésaurus UMLS. Ce dernier se voit attribuer un identifiant unique appelé (CUI – *Concept unique Identifier*). Tous les termes qui ont une signification équivalente (Synonymes) sont regroupés sous ce même concept unique (ces concepts peuvent provenir de différentes sources terminologiques) (Figure 2.6). Un seul nom de concept est identifié comme « terme préféré » pour un vocabulaire source donné (C001403) pour la maladie d'Addison (Figure 2.6).



De plus, les concepts de l'UMLS au niveau du métathésaurus sont connectés par un ensemble de relations. Ces relations sont identifiées dans UMLS par une combinaison de « REL » et « RELA » pour exprimer la relation entre un concept CUI1 et un autre concept CUI2, la dénomination « REL » désigne un type de relation hiérarchique et la dénomination « RELA » pour exprimer une relation sémantique additionnelle de niveau thésaurus (Tran *et al.*, 2015).

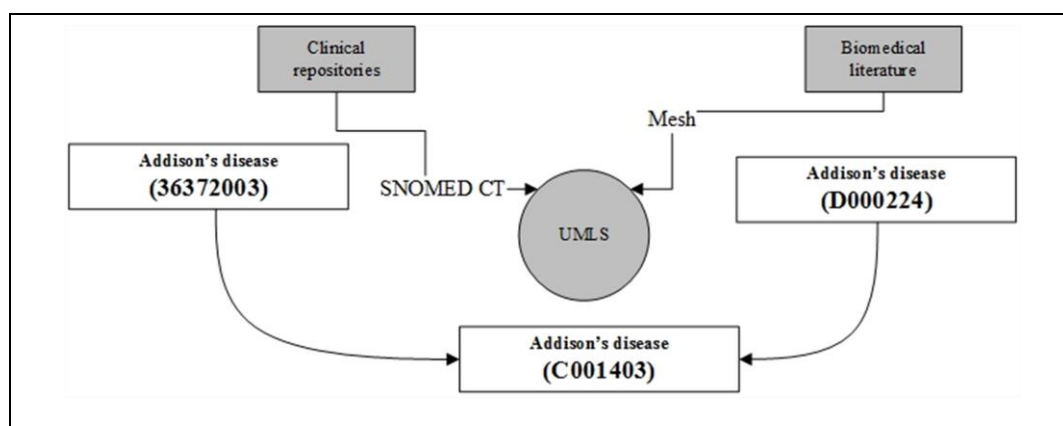


Figure 2.6 Intégration des espaces noms entre différentes terminologies.

Le réseau sémantique est une autre couche sémantique supplémentaire dite de « haut niveau » au-dessus des concepts de base dans les vocabulaires (McCray *et al.*, 2001), ce réseau permet une catégorisation par « type sémantique » des concepts médicaux représentés dans le métathésaurus. Chaque concept Métathésaurus se voit attribuer un ou plusieurs types sémantiques (catégories), qui sont reliés entre eux par des relations sémantiques (Bethesda, 2009 -a; McCray, 1989; Sony et Sureshkumar, 2019).

Le lexique spécialisé se compose d'un ensemble d'entrées lexicales avec une entrée pour chaque orthographe ou ensemble de variantes orthographiques dans une partie particulière du discours. L'entrée de lexique pour chaque mot ou terme enregistre les informations syntaxiques, morphologiques et orthographiques. Il est destiné à être un lexique anglais général qui comprend de nombreux termes biomédicaux. La couverture

comprend à la fois les mots anglais courants et le vocabulaire biomédical (Bethesda, 2009 -b).

### 2.7.3 Les relations entre concepts dans UMLS

Comme nous l'avons déjà mentionné dans la section précédente, les concepts dans UMLS appartiennent au métathésaurus et ils sont chapeautés par un réseau sémantique. Les relations entre les concepts sont alors utilisées pour représenter les liens hiérarchiques et sémantiques dans ces deux couches.

#### 2.7.3.1 Les relations au niveau du métathésaurus

Les principales relations hiérarchiques entre concepts dans le métathésaurus sont les relations « parent/enfant » (PAR/CHD) et les relations « élargies/spécifiques » (RB/RN). En général, les relations « parent/enfant » sont déjà définies dans les sources vocabulaires avant qu'ils ne soient intégrés à l'UMLS, alors que les relations « élargies/spécifiques » sont ajoutées par les éditeurs UMLS (McInnes et al., 2013). D'autres types de relations subsistent dans UMLS, comme illustré dans la (Figure 2.7) et décrites dans le (Tableau 2.2).

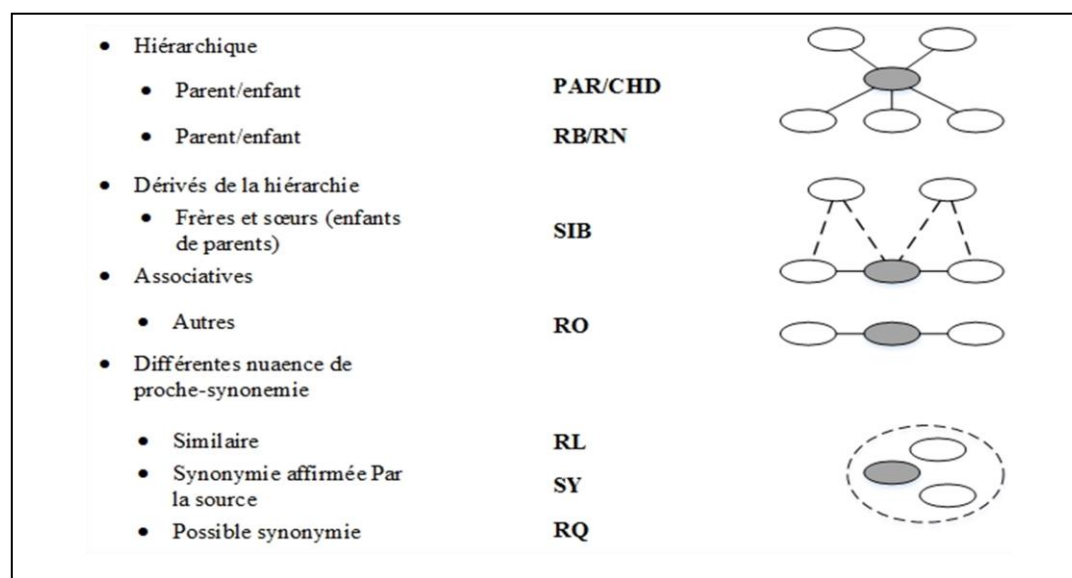


Figure 2.7 Types de relations dans UMLS.

Tableau 2.2 Relations dans le métathésaurus UMLS.

Abréviation	Description	Exemple
CHD	A une relation d'enfant	C1 est enfant de C2 (is_a)
PAR	A une relation de parent	C1 est parent de C2 (inverse_isa)
RB	A une relation plus large (élargie)	C1 est parent de C2 (inverse_isa)
RN	A une relation plus spécifique (étroite)	C1 est enfant de C2 (is_a)
RL	La relation est similaire ou ressemble à	C1 est comme C2
RO	Relations autres que CHD, PAR, RB, RN OU SY	C1 est associative avec C2
SIB	A une relation frères et sœur	C1 est C2 enfants du même parent
SY	Relation de synonyme	C1 est synonyme de C2
RQ	Liés et éventuellement synonyme.	C1 et C2 sont liées.

Ces relations de niveau métathésaurus peuvent aussi être accompagnées par une relation supplémentaire de nature sémantique entre les concepts. Ces relations sont en général le fruit d'un héritage provenant du réseau sémantique de « haut niveau ». Le tableau 2.3 présente un exemple de relations reliant le concept « Blood transfusion » à d'autres concepts dans UMLS.

Tableau 2.3 Exemple de relations avec le concept « Blood transfusion ».

CUI1	STR	CUI2	STR	RELA	REL
C0005841	Blood Transfusion	C0086252	Erythrocyte Transfusion	CHD	is_a
C0005841	Blood Transfusion	C0363457	Transfusion duration	RO	has_component
C0005841	Blood Transfusion	C0150150	Administration of blood	PAR	inverse_isa
C0005841	Blood Transfusion	C0398313	Other blood transfusion	RN	was_a
C0005841	Blood Transfusion	C0005841	Transfusion	SY	has_permuted_term
C0005841	Blood Transfusion	C0015236	Exchange Transfusion	RQ	use

Voici ci-dessous des exemples (en représentation graphique) d'extraction de concepts selon certains types de relation (réalisée avec l'outil ICPU à partir d'UMLS ; (<https://www.icpu-umls.ml/>)).

- Exemple de relations CHD (is\_a) pour le concept : C0005841 Blood Transfusion (Figure 2.8).

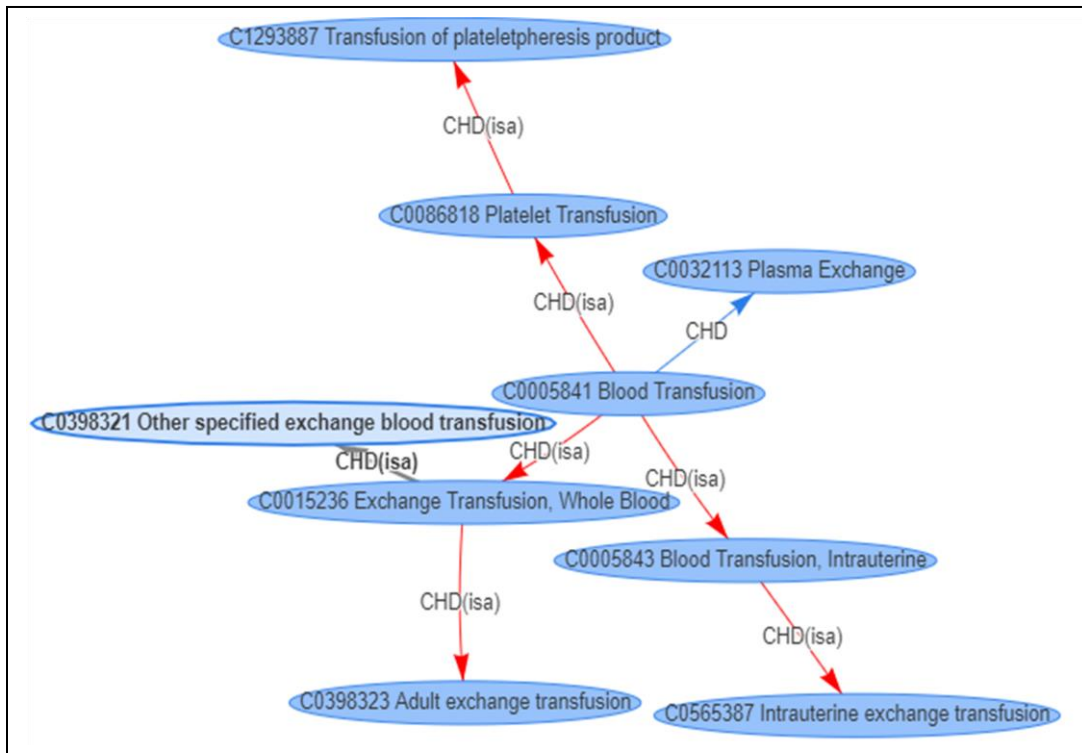


Figure 2.8 Graphe qui représente les relations CHD (is\_a) entourant le concept C0005841 : Blood Transfusion.

Dans l'exemple ci-dessus, le concept « Blood Transfusion » a plusieurs relations enfant comme le concept « *Exchange Transfusion, whole blood* ». Ce dernier est lié aussi avec une autre relation de ce type au concept « *Adult exchange transfusion* » et ainsi de suite. L'algorithme explore l'entourage termino-ontologique de chaque concept avec les paramètres choisis par l'utilisateur.

- Exemples d'un ensemble de relations hiérarchiques (CHD, PAR, RN et RB) pour le concept : C0005841 : Blood Transfusion (Figure 2.9)

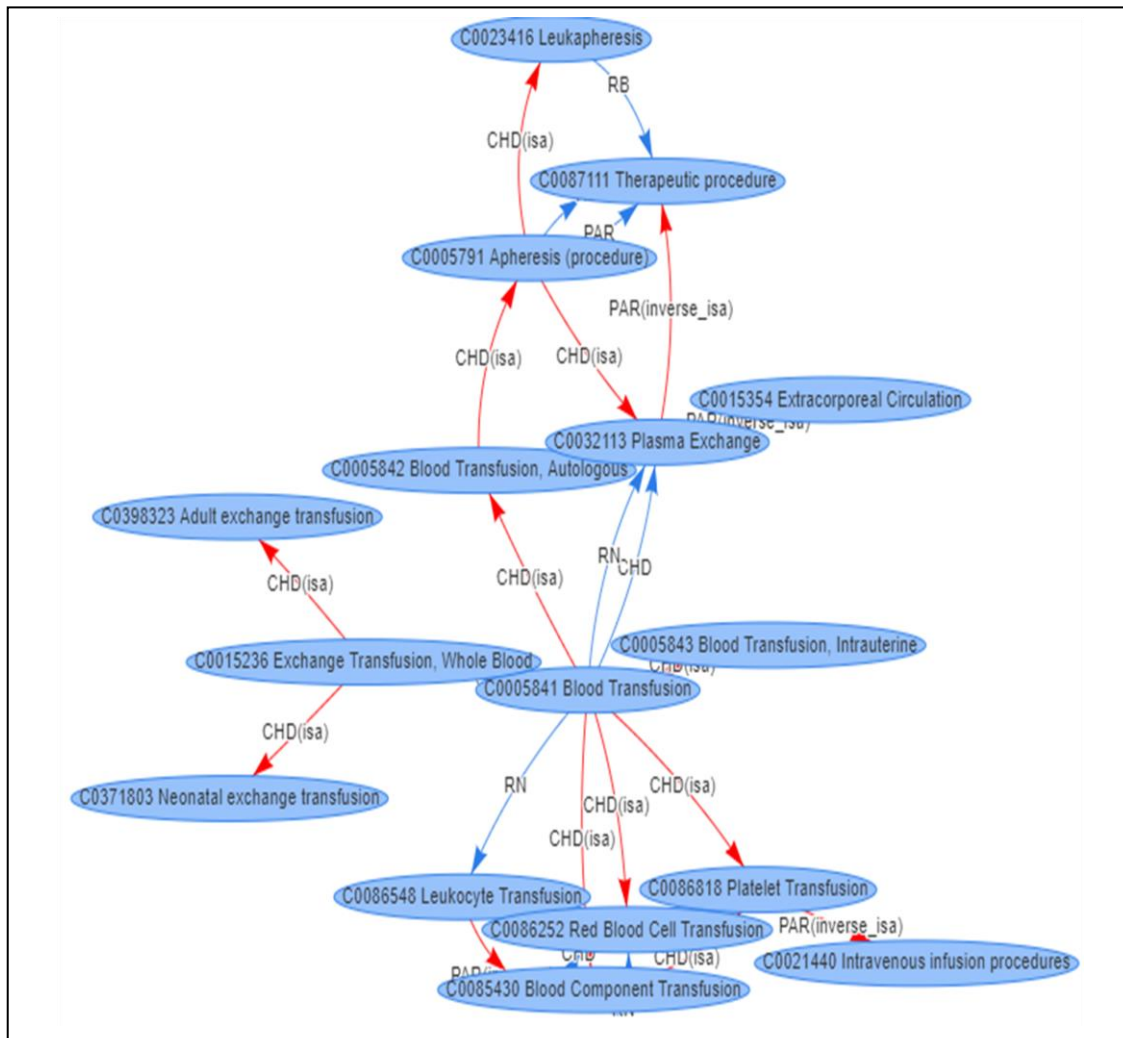


Figure 2.9 Relations hiérarchiques pour le concept : C0005841 Blood Transfusion

Avec plus de relations prises en compte, nous pouvons voir dans l'exemple ci-dessus que le graphe est plus étendu et inclus d'autres concepts avec des relations comme PAR et RN. Ce n'est pas tous les concepts qui existent autour de « Blood Transfusion », mais juste une partie discriminée par les paramètres 'd'exploration.

- Exemple de la relation RO pour le C0005841 : Blood Transfusion (Figure 2.10)

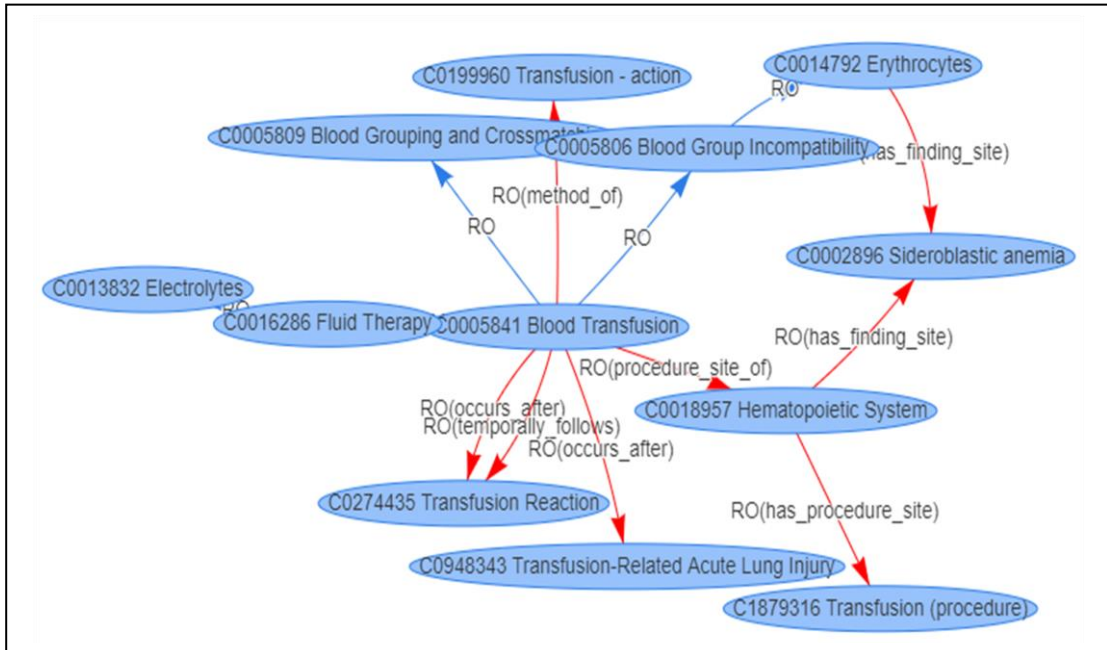


Figure 2.10 Relation de type RO entourant le concept C0005841 : Blood Transfusion.

Dans l'exemple ci-dessus, ce sont uniquement les relations associatives qui entourent le concept « *Blood Transfusion* » qui font l'objet d'exploration. Cela donne de nouveaux concepts qui diffèrent des relations hiérarchiques et de la synonymie. Exemple la relation « *occurs-after* » qui subsiste entre notre concept exemple et le concept « *Transfusion-Related Acute Lung Injury* » et qui décrit une relation de type événement et qui peut être pertinents pour un système d'aide à la décision par exemple (voir section suivante).

### 2.7.3.2 Réseau sémantique

Les relations sémantiques qui existent au niveau du métathésaurus (principalement dans les relations de type RO) sont héritées de la couche supplémentaire de haut niveau appelée « réseaux sémantiques ». Ce réseau est constitué de 127 types sémantiques et 54 types de relations représentant deux grandes familles de concepts en arborescence. Une branche pour les entités (Objet physique, entité conceptuelle) et une autre branche pour les événements (Activité, phénomène ou processus). Un exemple est donné dans la figure 2.11 (Bodenreider, 2001), où le type sémantique « *Body Part, Organ or Organ Component* » chapeaute les concepts du métathésaurus représentant des organes comme « *Heart, Heart valve...* ». Ces types sémantiques ne se retrouvent pas comme des concepts au niveau du métathésaurus, mais ils servent comme des catégories de haut niveau assigné à chaque concept du métathésaurus. Ces types sémantiques existent dans un fichier<sup>12</sup> de définition séparée du métathésaurus UMLS.

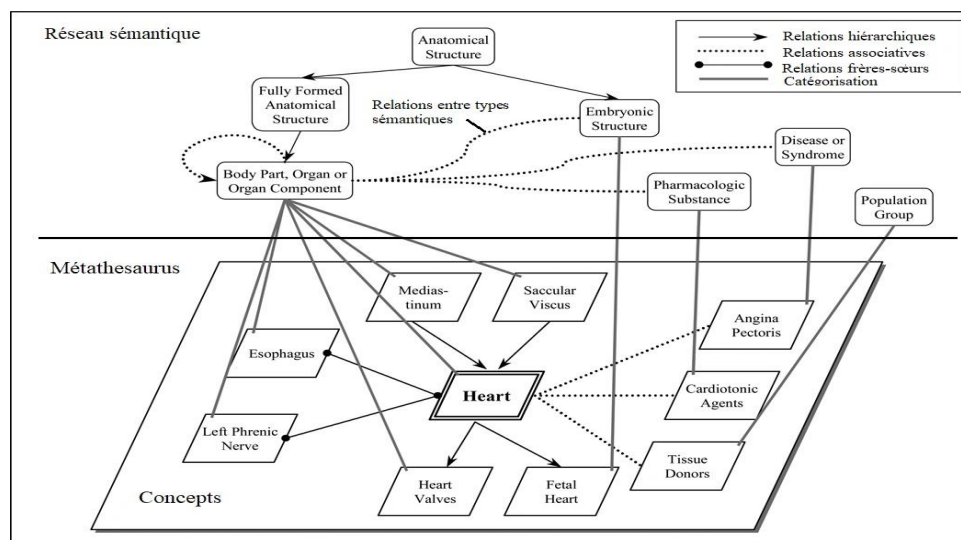


Figure 2.11 Réseau sémantique d'UMLS (Bodenreider, 2001).

<sup>12</sup> Pour plus d'information sur le réseau sémantique, voir le lien suivant : <https://www.ncbi.nlm.nih.gov/books/NBK9679/>.

## 2.8 Revue de littérature

### 2.8.1 Identification des concepts potentiellement pertinents

Comme mentionné, les ressources termino-ontologiques d'UMLS ainsi que son réseau sémantique s'apparentent grandement à une ontologie et de ce fait, nous avons exploré l'identification des concepts potentiellement pertinents dans le contexte des ontologies. Dans ce domaine, le besoin d'identification de concepts pertinents existe même avant la construction d'une ontologie où cette identification est considérée comme une étape importante du cycle « *d'apprentissage ontologique* » et qui permettra de définir les concepts qui feront partie de cette dernière et dont la qualité de l'ontologie en dépendra grandement (Asim *et al.*, 2018; Kim, Y. *et al.*, 2017; Maedche et Staab, 2004). Ce processus d'identification des concepts se fait à partir de nombreuses sources d'information en lien avec le domaine cible de l'ontologie (texte non structuré, sources sémantiques sur le web, ontologies existantes...). Pour permettre une certaine automatisation du processus d'acquisition de ces concepts, plusieurs techniques ont été utilisées comme l'apprentissage automatique, la fouille de textes, la représentation et le raisonnement sur les connaissances ainsi que le traitement du langage naturel (Asim *et al.*, 2018; Hacene *et al.*, 2008; Maedche et Staab, 2001).

Dans le cadre de notre travail, c'est plus l'identification des concepts pertinents dans une ontologie existante qui nous intéresse. En effet, plusieurs études ont exploré ce type d'identification que ce soit pour : améliorer la compréhension des ontologies (Graves *et al.*, 2008; Wu, G. *et al.*, 2008) ; pour améliorer l'indexation bibliographique (Willis et Losee, 2013), 3) pour classer et/ou sélectionner les ontologies à utiliser suivant la centralité des concepts (Alani et Brewster, 2005; Alani *et al.*, 2006), 4) ou pour améliorer la recherche d'ontologies sur le web pour la réutilisation (Butt *et al.*, 2016) et bien d'autres utilisations. Wu, G. *et al.* (2008) ont défini l'importance d'un concept dans une ontologie par rapport au nombre des relations sortantes de ce dernier



vers d'autres concepts (en utilisant l'algorithme de « reverse PageRank »<sup>13</sup>). Dans le (Tableau 2.4) ci-dessous, un exemple d'une extraction est illustré avec les concepts retrouvés en gras par rapport à la colonne de référence. Cette importance est accrue si un lien se trouve étant un lien avec un autre concept plus important.

Tableau 2.4 Extraction des concepts pertinents pour l'ontologie du logiciel (Wu, G. et al. (2008))

Classement	Référence	Wu, Li et al. 2008	Utilisateurs
1	Projet	<b>Projet</b>	<b>Projet</b>
2	Membre	<b>Statistiques d'utilisation</b>	<b>Catégorie</b>
3	Développeur	Enregistrement	<b>Message</b>
4	Catégorie	<b>Développeur</b>	Discussion
5	Forum public	<b>Catégorie</b>	Aide
6	Dernières nouvelles	Versionnage	Personne
7	Message	<b>Membre</b>	<b>Membre</b>
8	Version	<b>Message</b>	<b>Développeur</b>
9	Page d'accueil	Aide	Administration projet
10	Statistiques d'utilisation	<b>Forum public</b>	<b>Forum public</b>

Comme l'étude de Wu, G. *et al.* (2008) l'approche de Butt *et al.* (2016) prend en considération la centralité du concept au sein de l'ontologie en se basant sur la connectivité et le voisinage de ce dernier (Figure 2.12).

- Connectivité : un concept est plus central à une ontologie, s'il existe davantage de relations intra-ontologiques à partir de ce concept.

---

<sup>13</sup> Reverse PageRank : à l'inverse du PageRank, cette approche donne de l'importance au concept si de nombreuses relations (arcs) sortent de ce dernier.

- Voisinage : un concept est plus central pour une ontologie, s'il existe une relation intra ontologique à partir du concept jusqu'à l'autre concept central.

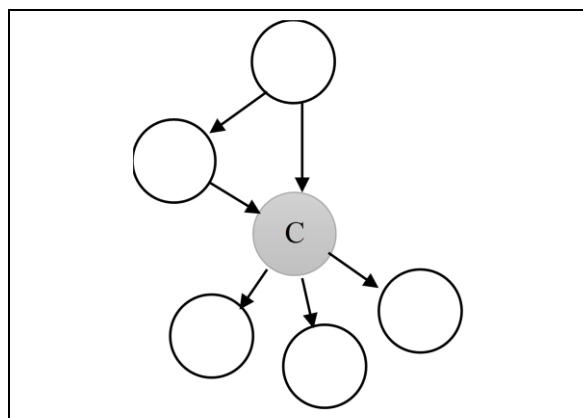


Figure 2.12 Centralité d'un concept (Butt et al., 2016)

Graves et al. (2008) ont considéré, quant à eux, que les concepts les plus centraux comme « importants », car ils peuvent accéder relativement facilement aux autres concepts dans le graphe de l'ontologie. Ce classement de concepts appelé « centralité de proximité » qui mesure combien coûte d'atteindre tous les nœuds du graphe à partir d'un nœud spécifique. Moins ce coût est élevé plus les concepts est centrale.

Plus récemment, l'étude de Kim, Y. *et al.* (2017) estiment l'importance des concepts en utilisant des indices supplémentaires provenant de source documentaire du domaine qui contiennent les concepts ontologiques. Ils utilisent la méthode dite de « synthèse multi-documents » qui vise à la sélection de phrases clés (ou mots) pour représenter (résumer) un document cible. Par la suite, ils exploitent le résultat de ce processus de sélection de mots-clés pour identifier (potentiellement) les concepts les plus importants dans l'ontologie.

Nombre de ces méthodes sont influencées par l'algorithme PageRank (Brin et Page, 1998) ou des variantes de ce dernier pour analyser la structure de l'ontologie. Pour

rappel, PageRank est un type d'algorithme de « marche aléatoire » (avec ou sans téléportation<sup>14</sup>) qui explore un graphe pour différents objectifs, notamment l'identification des nœuds importants. L'idée derrière le PageRank est que les nœuds visités le plus souvent dans cette marche aléatoire sont les plus importants. Le PageRank d'un nœud dépendra donc de la structure de ces liens au sein du graphe. Néanmoins, d'autres études comme celle de (Kim, Y. *et al.*, 2017) utilisent d'autres approches pour identifier les concepts potentiellement importants à l'ontologie comme des sources documentaires supplémentaires et externes à cette dernière (voir en début de cette section). L'identification des concepts importants dans ontologies cherchent à classer les concepts par rapport à la totalité des concepts de cette dernière.

L'une des études réalisées sur une termino-ontologie est celle de Willis et Losee (2013) pour améliorer l'indexation bibliographique par « sujet »<sup>15</sup>. Ces derniers utilisent la structure du métathésaurus (AGROVOC<sup>16</sup>, HEP<sup>17</sup>, NALT<sup>18</sup>, MESH) pour déterminer les concepts importants qui peuvent être utilisés pour cette indexation de documents.

---

<sup>14</sup> Téléportation : Dans les algorithmes de type « PageRank », la téléportation est utilisée lorsque l'algorithme se trouve sur un nœud sans liens sortants, alors il se téléporte (saute à un autre nœud) aléatoirement sur un autre nœud du graphe.

<sup>15</sup> Indexation par sujet : est un processus qui permet l'identification des sujets centraux d'un document et de leur représentation à l'aide d'un langage d'indexation.

<sup>16</sup> AGROVOC a été créé dans les années 1980 comme un thésaurus structuré multilingue dans tous les domaines ayant trait à l'agriculture, à la foresterie, à la pêche, à l'alimentation et autres domaines apparentés

<sup>17</sup> La taxonomie HEP pour *high Energy physics* Le thésaurus est actuellement utilisé comme vocabulaire contrôlé par les instituts de la physique des hautes énergies du monde entier.

<sup>18</sup> NALT pour *The National Agricultural Library's Agricultural Thesaurus*, le thésaurus et le glossaire sont des outils de vocabulaire en ligne de termes agricoles en anglais et en espagnol et produits par la National Agricultural Library, États-Unis.

L'objectif principal est d'analyser la contribution de la structure du thésaurus au processus d'indexation. Durant ce processus cognitif d'indexation et pour retrouver des concepts pertinents, les indexeurs de documents s'appuient en général sur leurs connaissances, leurs expériences antérieures, ainsi que les thésaurus. En feuilletant ces sources, l'indexeur prend nécessairement en compte les différentes relations entre les concepts pour décider si un terme particulier devrait ou ne devrait pas être utilisé. Pour les auteurs, les relations (associative, hiérarchiques) qui existent dans la structure du métathésaurus et peuvent servir pour extraire des concepts importants.

Pour modéliser l'effet de la structure du thésaurus sur le processus d'indexation par sujet, la méthode s'appuie sur un algorithme de marche aléatoire qui utilise comme points de départ un ou plusieurs termes initiaux issus d'une phase précédente appelée « correspondance orientée-thésaurus ». Cette étude considère UMLS comme un graphe non orienté avec des sommets pour chaque terme « préférés » dans le thésaurus. Les arcs représentent des relations entre concepts de type : « sens plus large/sens plus étroit ou spécifique » (RB pour « *broader* » /RN pour « *narrower* ») et RO pour les liens « associatifs » (voir aussi la section 2.7.3). L'étude analyse l'impact de certains paramètres sur la performance de l'algorithme de marche aléatoire pondéré : comme le nombre d'itérations de la marche, la longueur et les probabilités pour choisir chaque type de relation. La pondération étant une probabilité qui sera attribuée aux différentes relations.

Selon cette étude, l'augmentation du nombre de marches sert à réduire le « bruit » du processus aléatoire (variabilité du résultat). Avec un plus grand nombre de marches, les concepts les plus importants ou les plus fréquemment visités se démarqueront des concepts moins fréquemment visités. La longueur de la marche représente la profondeur sur laquelle l'indexeur navigue depuis le terme de « départ » jusqu'à sélectionner le dernier terme.

La probabilité attribuée pour chaque relation suggère que l'indexeur favorise un type particulier de relation par rapport à d'autres. (Exemple : une relation avec une probabilité de 0.6 permettra de favoriser certains sommets par rapport à des relations avec 0.4 de probabilité.)

Les résultats de cette étude mettent en lumière l'apport de la structure du thésaurus à la sélection des termes pour l'indexation des documents. Ainsi, l'indexation par sujet comme processus peut être modélisée en partie en tant qu'une navigation entre concepts. La marche aléatoire pondérée améliore les performances d'identification des concepts importants avec une augmentation de la précision moyenne entre 9 % et 37 % (suivant le vocabulaire utilisé) par rapport à indexation manuelle. La mesure de la précision et du rappel a été utilisée pour valider ces résultats avec des définitions adaptées au contexte de l'indexation des documents, avec une précision qui reflète le rapport du nombre de termes pertinents, extraits sur l'ensemble des termes du thésaurus alors que le rappel est défini comme le rapport des termes pertinents sur tous les termes extraits manuellement.

Dans notre travail de recherche, nous nous sommes inspirés de cette étude. Comme les auteurs de cette étude, nous pensons que l'identification des concepts pertinents pourrait être améliorée par l'utilisation de la structure termino-ontologique d'UMLS (avec l'utilisation de plusieurs sources vocabulaires en même temps ainsi que d'autres paramètres) pour des besoins autres que l'indexation (ex. recherche documentaire, extraction de sous-ensembles ontologiques...). Toutefois, nous pensons que cette exploration se fera d'une façon plus efficace en utilisant la cooccurrence entre concepts pour influencer l'orientation de la marche aléatoire. Nous présenterons notre approche dans la section 2.9 avec plus de détail.

### 2.8.2 Processus pour l'identification des concepts potentiellement pertinents

Dans la section précédente 2.8.1, nombreuses sont les études (Willis et Losee, 2013; Wu, G. *et al.*, 2008; Wu, W. *et al.*, 2012) qui utilisent des algorithmes d'exploration apparentée à PageRank et qui découlent du processus stochastique appelé « Marche aléatoire » afin de déterminer l'importance d'un concept au sein d'un graphe. Ces études utilisent généralement un modèle de représentation en graphe des ontologies ou de termino-ontologies (Kim, Y. *et al.*, 2017; Wu, G. *et al.*, 2008). Partant du fait que ce modèle convient à la structure relationnelle d'une ontologie, et ce, d'une façon simple, propre, élégante et sous un format utilisable et réutilisable (Mitra, P. *et al.*, 1999).

Le processus de « marche aléatoire » a été fortement utilisé pour les analyses structurales sur des graphes, car elle peut rendre compte de la structure topologique locale et globale au sein de ces graphes (réseaux) et elle est aussi très utile pour identifier les nœuds les plus importants/centraux (Komurov *et al.*, 2010).

En substance, la marche aléatoire est une méthode qui permet d'explorer un graphe en choisissant aléatoirement le prochain sommet à visiter (Wu, B. et Chellapilla, 2007). Cette chaîne de Markov<sup>19</sup> est caractérisée par une suite d'étapes indépendantes avec un ensemble de probabilités de passer d'un état à un autre à l'étape suivante, appelées probabilités de transition (Saeed, 2005). La marche aléatoire a été largement appliquée dans une variété de disciplines, telles que l'informatique, la biologie et l'économie pour modéliser le comportement des internautes, des cellules vivantes et de la bourse

---

<sup>19</sup> Un processus de Markov est un processus stochastique possédant la propriété de Markov : l'information utile pour la prédiction du futur est entièrement contenue dans l'état présent du processus et n'est pas dépendante des états antérieurs (le système n'a pas de « mémoire »). Les processus de Markov portent le nom de leur inventeur, Andreï Markov.

(Codling *et al.*, 2008; Godfrey *et al.*, 1964; Safaeifard *et al.*, 2018; Willis et Losee, 2013; Wu, B. et Chellapilla, 2007).

En informatique et parmi les applications les plus célèbres d'une marche aléatoire, on retrouve l'algorithme de PageRank utilisé par Google pour classer les pages web (Brin et Page, 1998). PageRank utilise une marche aléatoire pour modéliser le comportement d'un « surfeur aléatoire » qui se déplace de page à une autre page sur le web, et ce, en cliquant sur des liens. En termes plus simples, un score est attribué à chaque page en fonction de la probabilité que le surfeur arrive sur celle-ci. Depuis, la marche aléatoire a été utilisée dans diverses applications comme la désambiguïsation du sens des mots (Agirre *et al.*, 2014) la collaboration (Li *et al.*, 2014) et la recommandation (Zhang *et al.*, 2013).

Voici en résumé les étapes de la marche aléatoire :

- À partir d'un graphe et d'un point de départ (nœud), on sélectionne un voisin au hasard et on se déplace vers ce nœud ;
- On répète le processus à partir de ce nœud vers d'autres nœuds voisins et ainsi de suite ;
- La séquence (aléatoire) de nœuds sélectionnés de cette façon est une marche aléatoire sur le graphe ;
- Le poids final (probabilité) du nœud correspond à l'importance structurelle de ce nœud dans le graphique et représente la proportion de temps que la marche aléatoire a visité ce nœud après un temps suffisamment long (Yeh *et al.*, 2009).

Selon Weigert (2014) : « *Il peut sembler contre-intuitif d'attendre une réponse déterministe (l'importance du nœud) avec un processus aléatoire, mais le hasard devient déterministe lorsqu'il est appliqué à l'infini : si vous lancez une pièce de monnaie, personne ne peut prédire de quel côté la pièce va tomber, mais si vous lancez une pièce*

de monnaie pour un nombre infini de fois, nous savons que 50 % du temps, elle sera pile ». Appliquer cette idée sur des graphes, cela revient à marcher le long des arrêtes (arcs) et ce, en commençant sur un sommet particulier et en choisissant une arête au hasard pour aller au sommet de l'autre côté de l'arête. Si cette expérience est répétée un nombre infini de fois, une distribution déterministe sera obtenue pour cette marche aléatoire (Figure 2.13). Dans cette figure, le nombre dans les nœuds correspond au nombre de visites de chaque nœud après un processus de marche aléatoire avec nombre de marche égale à 100.

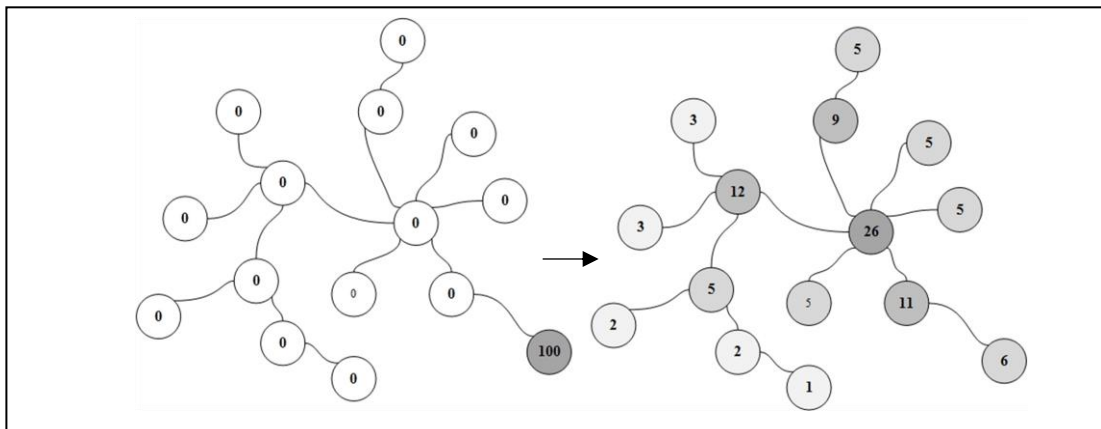


Figure 2.13 Diffusion d'information sur un graphe par marche aléatoire (Weert, 2014).

L'avantage d'utiliser la marche aléatoire est de pouvoir explorer un graphe totalement inconnu au départ (découvrir les nœuds au fur et à mesure). Aussi, son implémentation est relativement simple à mettre en place et elle requiert moins de ressources informatiques, de plus, ce processus peut s'adapter si le graphe change dans le temps (Gąsieniec et Radzik, 2008; Kosowski, 2013). Les motivations pour utiliser la marche aléatoire sont nombreuses : l'exploration des sites web, l'échantillonnage des nœuds et la collecte de statistiques, le classement des nœuds, exécution de la maintenance du réseau, etc. Plusieurs variantes de la marche aléatoire dans les algorithmes PageRank et PageRank personnalisés, cherchent à atteindre une distribution stationnaire pour



l'entièreté du graphe avec une probabilité de départ prédéfinie de passer d'un sommet à un autre.

Cette distribution exprime la probabilité de se retrouver sur un nœud donnée à partir d'un autre nœud dans le graphe. La marche aléatoire comme concept pourrait être transposée à un ensemble termino-ontologie comme celui d'UMLS. Ce dernier, peut être modélisé comme un arbre ou un graphe avec des concepts reliés par de nombreuses relations où chaque concept représente un nœud et chacun relation est représenté par arc entre deux nœuds (Willis et Losee 2013).

La marche aléatoire n'est pas l'unique méthode utilisée pour extraire des concepts à partir d'une ontologie, d'autres approches et processus permettent aussi l'extraction partir de graphe de représentation des connaissances pour différents besoins. C'est le cas de la segmentation et de la modularisation des ontologies (Abbes *et al.*, 2012; Stuckenschmidt *et al.*, 2009).

Cette dernière vise à extraire des sous-ensembles de concepts pour redimensionner de grandes ontologies — monolithiques — (en sélectionnant et en n'utilisant que la partie appropriée), la réutilisation ou pour accélérer le raisonnement sur les ontologies (Doran *et al.*, 2007; Konev *et al.*, 2009; Pathak *et al.*, 2009). Certaines de ces méthodes de modularisation sont des approches dites structurelles qui utilisent des algorithmes de parcours de graphe avec des caractéristiques d'extraction comme la similarité, la distance où la dépendance entre concepts pour choisir et délimiter les modules (Abadi et Zamanifar, 2011; Ghafourian *et al.*, 2013; Stuckenschmidt et Klein, 2004; Stuckenschmidt *et al.*, 2009).

Pour l'identification des concepts pertinents dans le contexte de nos travaux sur UMLS, une modularisation n'est pas envisagée pour plusieurs raisons. La principale raison est

que l'objectif de ces techniques est d'extraire un certain nombre de concepts pour former un module ontologique (sous-ensemble) correct (Un module  $M [O]$  d'une ontologie  $O$  ne doit contenir que les informations présentes dans  $O$ ), complet et avec une taille minimale et qui appartient à un domaine donné (Pathak *et al.*, 2009), ce qui ne serait pas le cas avec des termes différents provenant des règles de décision est qui peuvent ne pas appartenir au même domaine ontologique. Aussi, nombre de ces algorithmes (Pathak *et al.*, 2009; Stuckenschmidt et Klein, 2004) sont destinés à des ontologies formelles qui ne sont pas facilement adaptables pour un ensemble complexe de termino-ontologies comme UMLS. En effet, une extraction de concepts à partir de cette ressource ne permet pas d'obtenir une sous-ontologie formelle du domaine (Bourdé *et al.*, 2012).

### 2.8.3 Notion de concepts pertinents

Comme nous l'avons constaté au travers des exemples d'études présentés dans la section 2.8.2), la pertinence des concepts est une notion subjective dont l'appréciation dépend principalement des besoins d'identification des utilisateurs. En effet, une ontologie peut inclure des milliers de concepts, ce n'est pas toujours facile à comprendre si certains d'entre eux sont plus pertinents que d'autres, représentant ainsi une sorte de connaissance de base du domaine (une information qui n'est généralement pas explicitement encodée dans des ontologies). Les utilisateurs peuvent trouver ces informations très importantes pour avoir une meilleure compréhension du domaine (Tonelli, S. *et al.*, 2011). Toutefois, cette recherche de pertinence peut s'articuler sur des approches et des unités de mesure qui aident à cette identification au sein des représentations des connaissances en graphes pour les ontologies. Parmi celles-ci, la centralité locale du nœud (concept) au sein d'un graphe non orienté et qui peut être calculé par la somme de toutes les arêtes qui relient ce concept à d'autres concepts. Ce degré de centralité est divisé en deux catégories pour les graphes orientés : la somme des arêtes entrantes

(centralité d'entrée) et la somme des arêtes sortantes (centralité de sortie) (Pouriyeh *et al.*, 2018). Les nœuds avec un degré de centralité supérieur (centralité d'entrée/centralité de sortie) sont généralement considérés comme des nœuds plus importants.

Il existe aussi une centralité liée au positionnement du concept par rapport à sa structure hiérarchique dans l'ontologie (Alani et Brewster, 2005). Cette dernière utilise comme paramètre le rapport entre le plus long chemin à partir de la racine (est qui traverse ce concept) jusqu'au « concept enfant » le plus profond du concept et la longueur entre le sommet et le concept lui-même (Alani et Brewster, 2005).

D'autres mesures peuvent être utilisées pour déterminer l'importance d'un concept au sein d'une ontologie, comme la mesure de la densité qui inclut le nombre de concepts plus spécifiques (enfants), les relations avec des concepts « frère » issus de la même parenté... etc. Centralité intermédiaire qui se concentre principalement sur la position d'un sommet dans un graphe est définie comme le nombre de chemins les plus courts de tous les nœuds d'un graphe vers tous les autres nœuds et qui passent par un nœud particulier.

La fréquence de visite d'un concept par un processus d'exploration aléatoire sur un graphe peut aussi être utilisée comme une unité de mesure qui reflète directement les centralités relatives des nœuds et leur importance dans le graphe. Cette dernière est fortement influencée par la topologie du graphe local (Komurov *et al.*, 2010; Willis et Losee, 2013; Yeh *et al.*, 2009).

De plus, la notion de pertinence ne se résume pas uniquement à l'importance du nœud (concepts) au sein de sa structure par ses composantes relationnelles et son positionnement hiérarchique, la proximité sémantique « *semantic relatedness* » entre concepts est une autre information susceptible de refléter la pertinence des concepts pour un domaine donné. Des concepts différents peuvent également être liés sémantiquement par

certaines relations telles que les relations associatives (Exemple : étudiant-école ou voiture - moteur) (Dallmann *et al.*, 2016; Liu, Y. *et al.*, 2012; Pedersen *et al.*, 2007; Ruotsalo et Mäkelä, 2009). Selon Wu, W. *et al.* (2012), les concepts pertinents pour un domaine incluent tous les concepts spécifiques au domaine ainsi que les concepts qui n'appartiennent pas directement au domaine, mais qui sont sémantiquement très pertinents pour le domaine. Nous pouvons donner l'exemple du terme « admission », ce dernier n'est pas un concept spécifique au domaine de la transfusion, mais plutôt un concept de la gestion hospitalière ; toutefois, il est fréquemment utilisé, car sémantiquement pertinent ; il s'agit donc d'un concept pertinent pour le domaine de la transfusion. Pour calculer cette proximité sémantique plusieurs approches ont été utilisées, certaines de ces approches s'appuient sur la co-occurrence Schuetze & Pedersen (1997). L'idée de base est de représenter chaque concept comme un vecteur capturant le nombre de cooccurrences avec tous les autres concepts dans un espace multidimensionnel. D'autres études comme celles de Singer *et al.* (2013) et (Niebler *et al.*, 2016), qui ont adapté le concept pour déterminer cette proximité sémantique sur des graphes (représentation graphique de Wikipédia), utilisent un historique de navigation entre les pages effectuées par les utilisateurs pour extraire des co-occurrences des concepts entre eux sur chaque « chemin » de navigation. Les vecteurs de co-occurrences résultants pour chaque concept permettront de calculer cette proximité (utilisant le cosinus). Ce qui suppose que les mots sont sémantiquement liés s'ils partagent des voisins similaires.

Pour la suite de notre travail, l'identification des « concepts » pertinents englobera les concepts qui se trouveront importants par leur « centralité » liée à la fréquence de visite du nœud par un processus d'exploration aléatoire orienté par la co-occurrence avec les concepts de départs (qui reflètera la proximité sémantique entre concepts). Cette exploration se fera dans un rayon défini par un nombre de marches, une longueur de la marche (fenêtre), un choix de relations inter concepts et un seuil de co-occurrence des concepts dans les publications.

#### 2.8.4 Termes de départs pour l'identification des concepts potentiellement pertinents

Au moment de rédiger ce travail, nous n'avons pas trouvé d'études qui utilisent les termes issus d'un arbre de décision ou d'un ensemble de règles pour amorcer la recherche de concepts pertinents entourant ces termes dans des ontologies ou des termino-ontologies ; ceci est probablement due à la spécificité de nos besoins. Les études comme celle de (Willis et Losee, 2013) ou de (Wu, W. *et al.*, 2012), utilisent comme termes de départs des concepts fournis par des experts en ontologie ou en indexation. Notre besoin d'identification de concepts est spécifique à l'extension de la base des connaissances entourant les règles de décisions d'un SADC.

L'une des rares études que nous avons retrouvées et qui utilisent les termes issus d'un arbre de décision est celle de Kharbat et El-Ghalayini (2008). L'objectif de cette étude n'est pas le classement des concepts selon leur importance dans une ontologie, mais plutôt la génération d'une ontologie à partir d'un arbre de décision. L'approche propose, en premier, un processus automatisé pour extraire des règles de classification d'un ensemble d'images numérisées (*Wisconsin Breast Cancer Dataset*) en lien avec des cas de cancer du sein. D'autres informations et indications sur la malignité ou la bénignité des cancers sur ces images ont été exploitées. Après l'extraction, 25 règles ont été définies. À partir de ces règles, ils ont pu produire un sous-ensemble ontologique du domaine du cancer du sein sans interaction humaine.

Cette approche illustre la possibilité d'utiliser des termes inclus dans les règles de décision comme des termes clés pour élargir les connaissances et même la construction d'une ontologie de domaine sans intervention humaine (Figure 2.14). Avec la même intuition, nous explorons l'identification des concepts importants pour un SADC à partir de ses propres règles de décisions.

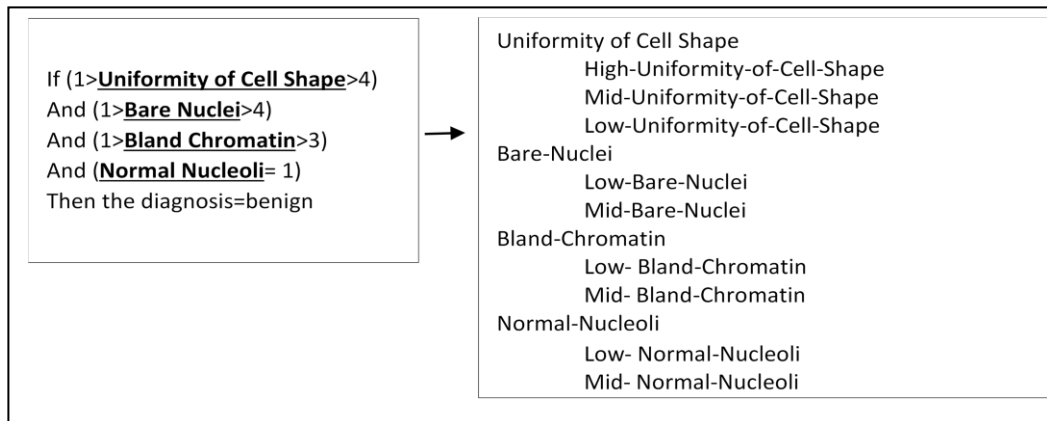


Figure 2.14 Exemple de construction d'une ontologie à partir des termes inclus dans les règles de décisions.

## 2.9 Méthodologie pour identification des concepts potentiellement pertinents

La méthodologie proposée pour identifier les concepts pertinents en explorant UMLS pour un SADC s'appuie sur un algorithme de marche aléatoire simulée et limitée (Nombre d'itérations et longueur de marche définie) qui utilise les termes provenant des règles de décision du SADC, des parties d'UMLS (concepts, relations sémantiques et/ou des relations thésaurus) ainsi que la fréquence de cooccurrence entre concepts dans des sources documentaires (PubMed). Nous avons adopté ce processus dans le cadre de nos travaux au vu de la nature et la grandeur de la structure à explorer (UMLS), des concepts de départ qui peuvent être variés et différents selon les règles d'aide à la décision et qu'il faut étendre les connaissances, la simplicité d'implémentation aux regards de nos ressources. Nous avons nommé ce processus *ICPU pour Identification des concepts pertinents dans UMLS* (Section 2.8.2).

La portion de concepts ainsi sélectionnée pourrait permettre d'élargir l'assise conceptuelle du SADC, d'extraire un sous-ensemble ontologique ou l'extension ontologique des requêtes de recherche d'information qui sera l'objectif de notre prochain chapitre.

En considérant UMLS comme un graphe orienté avec des sommets reflétant des concepts préférés et des arcs reflétant les relations entre concepts, voici l'approche préconisée (Figure 2.15) :

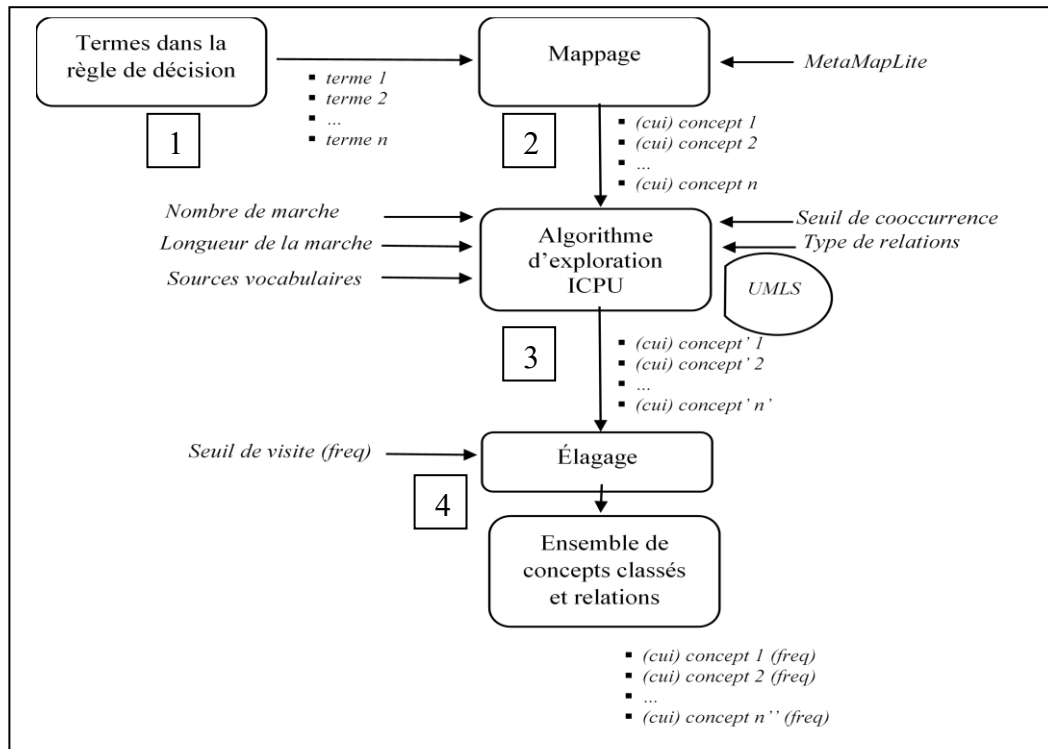


Figure 2.15 Processus d'identification des concepts pertinents ICPU.

**Étape 1 :** Comme point de départ, utiliser les termes qui existent dans les clauses conditionnelles et décisionnelles des règles de décision du SADC. (Arbre de décision). Nous considérons dans notre méthodologie que les règles de décision sont les vecteurs qui peuvent dessiner le contour d'une activité ou une tâche médicale. Toutefois, les termes de départs peuvent provenir d'autres sources d'information comme du texte libre.

**Étape 2 :** Mapper ces termes en utilisant l’outil « MetaMapLite » pour identifier les concepts en lien avec les termes issus de l’étape précédente. (Définir les *cui* : *Concept unique Identifier* pour pouvoir explorer UMLS).

**Étape 3 :** Exécuter l’algorithme avec un choix de paramétrer suivant les besoins de l’exploration :

- Le nombre d’itérations de l’algorithme à partir du concept de départ (Nombre de marches) ;
- La longueur du chemin d’exploration à partir de chaque terme de départ.
- Les sources vocabulaires utilisés ;
- Le type de relations (hiérarchiques, associatives... etc.) ;
- Le nombre de cooccurrences (avec le terme du départ) à partir duquel les nœuds voisins seront considérés.

L’algorithme récupère les chemins ainsi parcourus, additionne le nombre de visites pour chaque concept ainsi retrouvé dans ces chemins.

**Étape 4 :** La sélection ou l’élagage des concepts sera décidé selon un seuil de visite de ce nœud.

Le résultat de ce processus est un ensemble de clusters de concepts et de relations entourant chaque concept de départ et qui peuvent être possiblement interconnectés. Avant de développer les contours en détail de la méthodologie utilisée et l’algorithme, il serait pertinent de décrire les tables d’UMLS qui seront exploitées (structure) et les outils incorporés (en occurrence : MetaMapLite).



### 2.9.1 Quelles parties d'UMLS (la structure)

L'exploration termino-ontologique de concepts se fera en parcourant les tables suivantes MRCONSO.RRF (pour les concepts), MRREL.RRF (pour les relations) et la table MROCR (pour la cooccurrence de concepts). Ces tables seront au cœur de notre algorithme et elles représentent les trois axes suivants : concepts, relations et cooccurrences. Ces tables seront stockées dans une base de données MySQL. Les deux premières tables proviennent de « *MetamorphoSys* » et la dernière table est importée séparément du site de NIH (Figure 2.16).

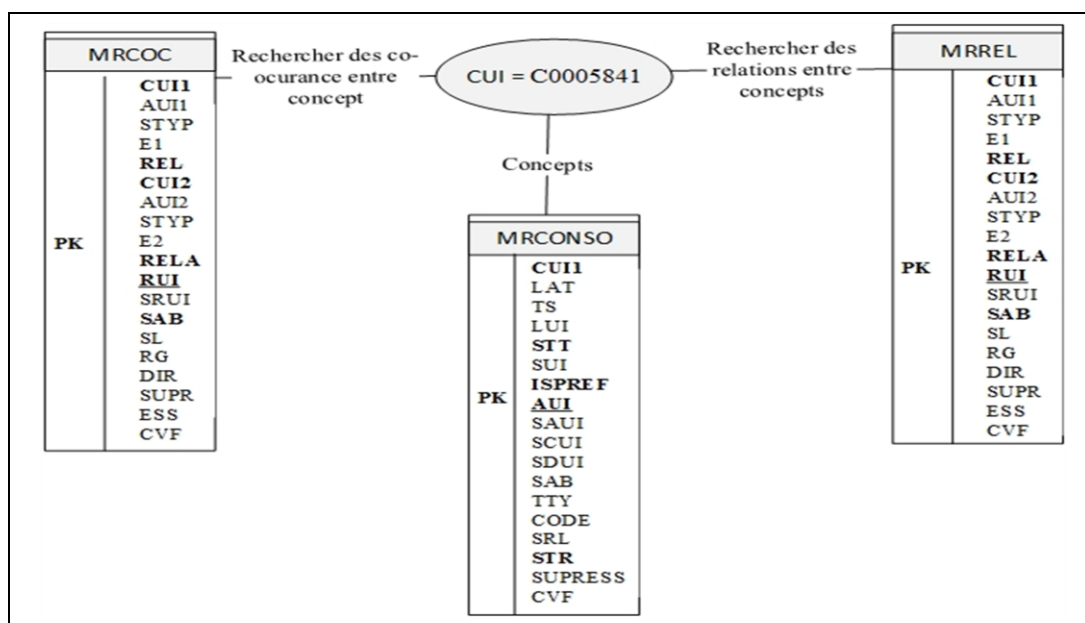


Figure 2.16 Principales tables d'UMLS utilisées par ICPU.

1. Concepts, noms de concepts et sources vocabulaires de la table : « MRCONSO.RRF »

Le Métathésaurus MTH est organisé suivant la notion de « concept » qui regroupe sous le même concept différents termes/noms provenant de différentes sources vocabulaires. La table MRCONSO (<https://www.ncbi.nlm.nih.gov/books/NBK9685/>) qui définit ces

concepts est au cœur du Métathésaurus et recèle plus de 8 582 934 lignes (version 2018 A). Ci-dessous les principales colonnes d'informations structurant cette table (Tableau 2.5).

Tableau 2.5 Description des colonnes de MRCONSO et un exemple pour le concept « Transfusion ».

Colone	Description	Ex. tem : Transfusion	
CUI	Identifiant unique pour le concept.	C1879316	C005841
LAT	Langue du terme	ENG	ENG
LUI*	Identifiant unique du terme	L0242446	L0242446
TS	Statut du terme	P	S
STT	Langue du terme	PF	SY
ISPREF	Statut Atome — préféré (Y) ou non (N) pour cette chaîne de caractères dans ce concept.	Y	Y
SUI*	Identifiant unique de la chaîne	S0281931	S0281931
AUI*	Identifiant unique pour l'atome	A2895502	A12800326
SAB	Nom de la source vocabulaire abrégé	SNOMEDCT US	NCI

\* LUI, SUI et AUI sont des codes attribués aux différentes déclinaisons terminologiques du concept. Note : d'autres colonnes n'ont pas été reportées dans le ce tableau, car elles ne sont pas pertinentes pour notre travail. Toutefois, une documentation peut être consultée sur le site de *National Center for Biotechnology Information* (NCBI).

Voici ci-dessous un exemple d'énumération pour le concept « Transfusion » désigné par le CUI= C1879316' (Tableau 2.6).

Tableau 2.6 Exemple des différentes descriptions lié aux concepts CUI= C1879316.

CUI	TS	LUI	STT	ISP.	AUI	SAB	STR
C1879316	S	L0242446	PF	Y	A2895502	SNOMEDCT	Transfusion
C1879316	P	L1027179	PF	N	A3760821	SNOMEDCT	Transfusion (procedure)
C1879316	S	L0242446	VO	Y	A3760904	SNOMEDCT	Transfusions
C1879316	S	L0242446	VO	N	A4735725	SNOMEDCT	Transfusion, NOS
C1879316	S	L0242446	PF	Y	A0314840	LCH	Transfusion
C1879316	S	L0242446	VC	Y	A11916242	PDQ	Transfusion
C1879316	P	L1027179	PF	Y	A13073423	MTH	Transfusion (procedure)
C1879316	S	L0242446	VO	Y	A8352013	MTHICD9	Transfusion, NOS

Remarque : Tous les concepts qui peuvent potentiellement être visités durant l’exploration de l’algorithme doivent avoir la valeur « P » dans la colonne TS (Statut du terme) et la valeur PF de la colonne STT pour toutes les terminologies utilisées. Les valeurs P et PF permettent de désigner un concept médical unique dans chacune des terminologies comme étant la terminologie la plus préférée, moins ambiguë, plus spécifique et autodéscriptive que possible.

## 2. Relation entre concepts : la table « MRRELA.RRF »

Le métathésaurus répertorie de nombreuses relations entre concepts (en plus des relations synonymes dans la structure de concepts du métathésaurus). Ces relations sont sauvegardées sous la table MRRELA.RRF. Cette dernière contient toutes les relations entre concepts qu’ils appartiennent au métathésaurus ou au réseau sémantique d’UMLS. Le tableau ci-dessous présente la description des colonnes nécessaires pour identifier la nature de la relation reliant deux concepts CUI1 et CUI2 (Tableau 2.7).

Tableau 2.7 Exemple de relation entre deux concepts dans la table MRRELA.RRF.

Colonne	Description	Exemple
CUI1	Identifiant unique pour le premier concept.	C1879316 (Transfusion)
CUI2	Identifiant unique pour le deuxième concept.	C0005841 (Blood transfusion)
REL	Relation du deuxième concept ou atome avec le premier concept ou atome	CHD
RELA	Relations additionnelles (plus spécifique) “sémantique”	Is_a
SAB	Nom de la source vocabulaire abrégé	SNOMEDCT_US

Remarque : si l’information dans les colonnes REL et RELA est fournie, cela implique qu’elles contiennent respectivement la relation au niveau du métathésaurus et la relation au niveau du réseau sémantique entre les deux concepts CUI1 et CUI2.

### 3. Cooccurrences entre concepts : la table extra-UMLS « MRCOC.RRF »

La table-fichier « MRCOC.RRF<sup>20</sup> » (Tableau 2.8) ne fait plus partie de la panoplie de tables produites et maintenues par les UMLS. Ce fichier est pris en charge par le projet « *Indexing Initiative* » du NIH. Ce fichier résume l'information sur l'analyse de cooccurrences entre les descripteurs MeSH qui apparaissent ensemble dans les citations MEDLINE de MEDLINE/PubMed Baseline<sup>21</sup>. Le tableau ci-dessous est un exemple des informations de cooccurrence qui relie deux concepts CUI1 et CUI2.

Tableau 2.8 Colonnes de la table/fichier MRCOC.

Col.	Description	Exemple terme = « anemia »		
CUI1	Identifiant unique pour le concept.	C0002871		
CUI2	Langue du terme	C0000778 (ABO Blood group System)	C2347023 (Absorption)	C0001075 (Achlorhydria)
Freq	Fréquence	15	20	33
Year	Année de publication	1996	1996	1996

La colonne « Freq » nous informe du nombre de cooccurrences entre le concept « Anemia » et les concepts « ABO Blood group System », « Absorption » et « Achlorhydria » durant l'année de publication 1996 citations « MEDLINE/PubMed Baseline ». Cette table contient 250 millions de lignes que nous avons intégrées et exploitées dans notre logiciel. Dans le cadre de notre travail de recherche, nous nous sommes limités uniquement aux relations figurant dans le métathésaurus incluant les relations sémantiques héritées du réseau sémantique. Le parcours par notre algorithme du réseau sémantique

<sup>20</sup> [https://www.nlm.nih.gov/research/umls/implementation\\_resources/query\\_diagrams/er1.html](https://www.nlm.nih.gov/research/umls/implementation_resources/query_diagrams/er1.html)

<sup>21</sup> MEDLINE/PubMed Baseline est un instantané de citation créée au début de chaque nouvelle année d'indexation MeSH contenant les enregistrements MEDLINE, PubMed.

entre types sémantiques « de haut niveau » n'a pas été considéré et pourrait faire l'objet de nouvelles recherches. Pour plus d'information, suivre le lien (<https://semanticnet-work.nlm.nih.gov/>).

### 2.9.2 MetaMapLite

L'objectif principal d'avoir intégré « MetaMapLite » dans notre algorithme c'est de pouvoir fournir un identifiant d'entité nommée<sup>22</sup> en quasi-temps réel. Cette version allégée de MetaMap n'est pas aussi rigoureuse que la version complète, mais beaucoup plus rapide tout en permettant aux utilisateurs de personnaliser et d'améliorer son comportement pour des objectifs spécifiques. Selon Demner-Fushman *et al.* (2017), les performances de MetaMapLite sont comparables à celles d'un MetaMap avec un paramétrage de base, sans un réglage spécifique pour une tâche donnée.

Comme évoqué, MetaMapLite est issu de MetaMap<sup>23</sup> qui est un outil de reconnaissance des entités nommées. Ce dernier s'est imposé comme l'une des premières applications pour l'identification des concepts de métathésaurus (UMLS) dans les textes biomédicaux (Demner-Fushman *et al.*, 2017). Le programme permet d'établir un lien entre le texte biomédical libre et les concepts y compris la relation de synonymie<sup>24</sup> intégrée au métathésaurus (Aronson et Lang, 2010). MetaMap a été développé en prenant en con-

---

<sup>22</sup> Les mentions « d'entités nommées » sont des unités textuelles qui renvoient à des « entités » du domaine, mais qui peuvent relever de différentes catégories linguistiques : noms propres (« Air France »), pronoms (« elle »), et plus largement descriptions définies (« cette compagnie », « la principale compagnie aérienne française »). Nous mettons cependant l'accent ici sur les seuls noms propres, plus faciles à détecter dans le flux textuel (Omrane, Nazarenko et al. 2010).

<sup>23</sup> Pour plus d'information sur l'outil : <https://metamap.nlm.nih.gov/>

<sup>24</sup> Synonymie est un rapport de similarité sémantique entre des mots ou des expressions d'une même langue. La similarité sémantique indique qu'ils ont des significations très semblables. Des termes liés par synonymie sont des synonymes.

sidération les principes linguistiques qui fourniront à la fois la base rigoureuse et l'architecture flexible dans laquelle les stratégies de mise en correspondance « Mapping » sont exécutées (Aronson et Lang, 2010). Le fonctionnement de MetaMap consiste à appliquer une série d'analyses lexicales successives et syntaxiques afin de proposer un ensemble de concepts (UMLS) en lien avec le texte initial. Voici les étapes d'analyses successives que MetaMap applique sur le texte biomédical :

- 1) Tokenisation<sup>25</sup>, détermination de la limite des phrases et identification des acronymes/abréviation ;
- 2) Étiquetage morphosyntaxique <sup>26</sup>;
- 3) Recherche lexicale des mots d'entrée dans le lexique SPECIALIST <sup>27</sup>;
- 4) Une analyse syntaxique superficielle dans laquelle les phrases et leurs têtes lexicales sont identifiées par l'analyseur d'engagement minimal de SPECIALIST. Chaque phrase trouvée par cette analyse est ensuite analysée selon les processus suivants :
  - a. Génération de variantes, dans laquelle les variantes de tous les mots de la phrase sont déterminées (normalement par une table) ;
  - b. Identification du candidat après évaluation en fonction de leur correspondance avec le texte saisi.

---

<sup>25</sup> Tokenisation : une tâche courante dans le traitement automatique du langage naturel elle consiste à prendre un texte ou un ensemble de texte et à le diviser en mots individuels. Ces mots ou « jetons » sont ensuite utilisés comme entrée pour d'autres types d'analyse ou de tâches.

<sup>26</sup> Étiquetage morphosyntaxique : le processus consiste à associer aux mots d'un texte, les informations grammaticales correspondantes comme la partie du discours, le genre, le nombre, etc. à l'aide d'un outil informatique.

<sup>27</sup> Le lexique SPECIALIST contient des informations sur le vocabulaire anglais courant, les termes biomédicaux, les termes trouvés dans MEDLINE et les termes trouvés dans le métathésaurus UMLS. Chaque entrée contient des informations syntaxiques (comment les mots sont assemblés pour créer un sens), morphologiques (forme et structure) et orthographiques (orthographe).

### 2.9.3 Algorithme

#### 2.9.3.1 Description

Comme déjà mentionnée dans la section 2.9, notre méthode d'exploration s'appuie sur l'utilisation des termes qui apparaissent dans les clauses (conditions, décisions) des règles de décisions propres à un domaine ou une tâche médicale particulière. Cet ensemble de départ permettra l'extraction d'un ensemble plus grand de concepts « pertinents » pour cette pratique médicale ainsi que les relations (hiérarchiques et/ou sémantiques) qui subsistent entre ces concepts.

Ces concepts et ces relations conceptuelles seront retrouvés par l'application d'un processus stochastique (marche aléatoire) avec la modulation de certains paramètres (comme le choix des sources ontologique et vocabulaire, la longueur de la marche, le nombre d'itérations... - voir section 2.9). L'idée est de réaliser un certain nombre de marches (ou itération) à partir d'un point de départ avec une certaine profondeur (longueur de marche) pour explorer les alentours ontologiques d'un concept. Toutefois, nous avons biaisé l'orientation de la marche en introduisant la cooccurrence (si elle existe) comme facteur pour choisir le prochain nœud (concept) à visiter. Si cette cooccurrence n'a pas été reportée alors un choix aléatoirement sera opéré. On se base sur la présomption que la priorisation des concepts co-occurents (dans des publications) favorisera la découverte des concepts pertinents pour le domaine plus que d'autres nœuds. Cette marche fera ressortir (après cumul du nombre de visites) qu'elles seront les sommets (concepts) les plus visités. Cela signifiera qu'ils sont les plus interconnectés avec d'autres nœuds du graphe et qu'ils peuvent apporter une information pertinente. Certains paramétrages seront aussi utilisés pour limiter et élaguer des concepts superficiels et non utiles durant cette exploration. À noter que la pertinence d'un concept pour notre algorithme est représentée par la somme des fréquences de visite de ce concept durant la marche aléatoire dans l'ensemble des clusters générés autour des concepts de départs. Cette fréquence représente une « centralité locale » des concepts

(le nombre de relations sortantes ou rentrantes vers le concept à travers les grappes de concepts).

Pour chaque concept de départ ( $c$ )

Marche aléatoire ( $c, NM, LM, rela, sab, fr, freq$ )

Sommet actuel := ( $c$ );

Pour ( $n$ ) allons de  $1$  à  $NM$

Pour ( $k$ ) allons de  $1$  à  $LM$

Faire cette marche uniquement pour les relations qui existent dans ( $rela$ ) et uniquement pour les termino-ontologies dans ( $sab$ ).

À partir du sommet actuel, sélectionner au hasard le prochain concept à visiter ( $c'$ ) parmi les sommets voisins ayant une cooccurrence  $\geq$  ( $freq$ ) avec le concept de départ ( $c$ );

Si aucun concept voisin n'a de co-occurrence avec le concept ( $c$ ) alors, choisir au hasard le prochain sommet à visiter ( $c'$ );

Sommet actuel := ( $c'$ );

Si pas de voisins alors quitter.

CumulerleNombreVisites [ $c'$ ] ++

Fin pour  $k$

Fin pour  $n$

Classer *CumulerNombreVisites* de toutes les ( $c'$ ) et sélectionner les ( $c'$ )  $\geq fr$ .

*NM* : nombre d'itérations (ou de marches) à partir de chaque sommet de départ ;

*LM* : la profondeur ou la longueur de la marche à explorer depuis le sommet de départ ;

*rela* : les types de relations ;

*sab*: sources vocabulaire à tenir en compte) ;

*freq* : le nombre de cooccurrences (minimal) entre le concept candidat à la visite et le concept de départ ;

*fr* : seuil de visite ou de fréquentation à partir duquel le sommet sera considéré comme pertinent et sera sélectionné à la fin du processus.



Voici ci-dessous, l'algorithme détaillé avec les différentes procédures et fonctions qui le composent :

---

**Algorithm 1** Algorithme de marche aléatoire orienté par la cooccurrence pour l'identification des concepts pertinents à partir d'UMLS (ICPU)

---

**Données :** Graph  $G (V, E)$  avec :  $C$  : un ensemble de concepts de départ provenant des règles de décision extrait par metamap;  
 $NM$  : nombre d'itérations;  
 $LM$  : longueur de la marche;  
 $Rela$  : relation du thesaurus cibles;  
 $Sab$  : source de vocabulaires cibles;  
 $fr$  : la fréquence de visite "seuil" à partir de laquelle le concept est retenue;  
 $frq$  : la fréquence de cooccurrence seuil entre le concept ( $c'$ ) et le concept ( $c''$ ) pour que le concept  $c''$  soit retenue comme prochain nœud à visiter.

**Resultats :**

$List$  : liste des concepts les plus pertinents avec fréquence de visite cumulée et fréquence de cooccurrence avec le concept de départ issue des règles de décisions du SADC;

$ListRelation$  : Un ensemble de tuples (concepts, relation, concepts) permettant de produire un graphe relationnel.

```

procedure CLUSTEREXTRACTOR( $C, NM, LN, Rela, Sab, fr, frq, output$ 
 $ListOfConcepts, ListOfRelation$ )
   $ListTemp[ ]$  an empty array
   $ListConcepts[ ]$  an empty array
  for all  $c \in Crule$  do
     $ListTemp \leftarrow findcluster(c, NM, LM, Rela, Sab, fr, frq);$ 
     $ListConcepts \leftarrow ListConcepts \cup ListTemp$ 
  end for
   $List \leftarrow UniqueConcept(ListConcepts)$ 
   $ListRelation \leftarrow ExtractRelation(ListConcepts)$ 
end procedure

```

**UniqueConcept(list)** : fonction qui permet d'éliminer les doublons dans la liste des concepts de la 'list' (en paramètre);

**ExtractRelation(list)** : fonction qui permet de retrouver toutes les relations entre différents concepts de la même 'list' (en paramètre);

---

---

```

function FINDCLUSTER(cui,NM,LM Rel,sab,fr,frq)
  Path[ ] an empty array
  ListConceptsWithFrequencies[ ] an empty array
  for i ← 1, N do
    Path := RandomWalk(cui,LM,Rel,Sab,frq);
    for j ← 0, Lenght(Path) do
      if (Path(j) not exist in ListConceptsWithFrequencies[]) then
        ListConceptsWithFrequencies[ ] := Path(j);
      end if
      ListConceptsWithFrequencies['Path(j)'] ++;
    end for
  end for
  ListConceptsWithFrequencies[] := letOnlyHighFreq(ListConceptsWithFrequencies[],fr);
  Return(ListConceptsWithFrequencies[]);
LetOnlyHighFreq(list) : fonction qui permet d'éliminer les concepts
avec un seuil inferieur à fr (Si définie).

```

---



---

```

function RANDOMWALK(c,LM,rela,frq)
  Path[ ] an empty array
  InList = getRelatedConcepts(c,rela,frq);
  for i ← 1, LM do
    NextCui = InList[rondom(InList)];
    path[] := NextCui;
  end for
  return(path[]);

```

---



---

```

function GETRELATEDCONCEPTS(c,Rela,sab)
  InList[] := select all related cuix to cui with rela ;
  for i ← 1, len(InList) do
    InList[frq] := findoccurence(cui, cuix);
  end for
  return(InList);
findoccurence(cui, cuix) : fonction qui permet de trouver la fréquence de co-
occurence entre cui et cuix

```

---

Dans la page qui suit, la figure 2.17 présente est un exemple d'exploration de l'algorithme à partir du concept de départ « Blood Transfusion : C0005847 ». Cette représentation graphique montre deux cheminements différents de marche aléatoire. L'une biaisée par la co-occurrence et l'autre totalement aléatoire. Plusieurs concepts de la démarche se retrouvent d'emblée dans le dictionnaire de l'INST (voir section 2.10.2.1 pour le corpus de référence INST) tandis que d'autres n'y apparaissent pas. Il est à noter que les concepts figurant dans les résultats de la marche aléatoire orientée par la cooccurrence présentent des concepts très proches du domaine d'application qui est la transfusion comme les concepts (*Blood, Blood component transfusion... qui figurent dans le dictionnaire de l'INST*) ou les concepts ayant été visité d'une façon totalement aléatoire comme « Subject of record » ou « Centesis - action ».

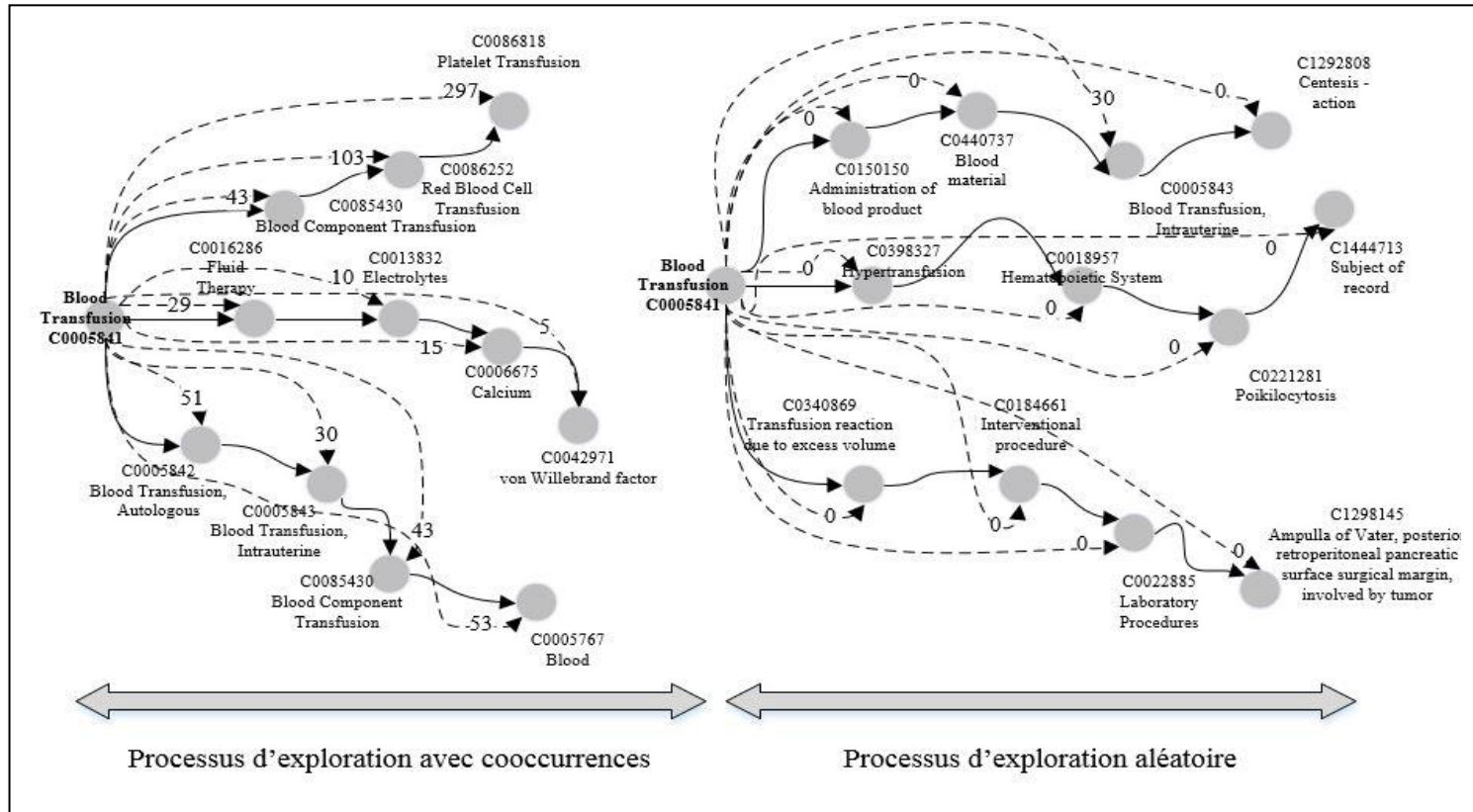


Figure 2.17 Deux ensembles de résultats issus d'un processus d'exploration avec co-occurrence (gauche) et un autre sans cooccurrence (droite).

### 2.9.3.2 Aspects relatifs au choix d'un sommet voisin suivant la cooccurrence

Considérons UMLS (termino-ontologies) comme étant un graphe orienté  $G = (V, E)$ , avec  $V$  l'ensemble des nœuds (concepts) et  $E$  est l'ensemble des arcs, une marche aléatoire naturelle dans un graphe (comme PageRank ou PageRank personnalisé) (Lovász, 1993; Weigert, 2014)) commence à partir d'un sommet donné, puis sélectionne l'un de ses voisins uniformément au hasard pour une visite. Ce passage entre nœuds est défini par la matrice de transition ( $M = (p(x, y)) \ x, y \in V$ ) suivante :

$$P(x, y) = \begin{cases} \frac{1}{\text{degré}(x)} & y \text{ est un des noeud voisin de } x \\ 0 & \text{sinon (si pas d'arc sortant)} \end{cases}$$

Ou degré ( $x$ ) est le nombre des arcs sortant du nœud  $x$ .

Dans notre approche, l'introduction de la cooccurrence comme biais fera en sorte que la marche aléatoire sera orientée pour choisir au hasard et prioritairement un nœud ( $y$ ) uniquement parmi les nœuds voisins qui ont au moins un degré (fréquence plus exactement) de cooccurrence avec le nœud de départ ( $x$ ). Si aucun nœud voisin ne répond à ce critère alors le prochain nœud ( $y$ ) sera choisi au hasard parmi les nœuds qui ont une relation avec le nœud ( $x$ ). Le type de relation et leur appartenance (au thésaurus et/ou au niveau sémantique) sont définis par un paramétrage au départ. Ceci nous donnera la présomption suivante :

$$P(x, y) = \begin{cases} \frac{1}{\text{degré}(x)} * (1 - c) + \frac{1}{\text{degré}'(x)} * c \\ 0 & \text{sinon (pas d'arc sortant)} \end{cases}$$

Avec  $\text{deg}'(x)$  est le nombre des arcs sortant du nœud  $x$  vers le nœud  $y$  et ayant un seuil de cooccurrence avec le concept initial. Ou  $\text{deg}(x)$  est le nombre des arcs sortant du nœud  $x$ .  $(c)$  est égale à 1 s'il y a cooccurrence entre  $x$  et  $y$  et 0 sinon.

### 2.9.3.3 Considérations et limites de ce processus d'exploration

- Dans nos expérimentations, nous n'avons pas eu recours au calcul de la matrice de transition pour atteindre un état stationnaire (une convergence globale comme dans PageRank pour l'ensemble du graphe - tous les concepts d'UMLS). Nous nous sommes limités à prioriser une marche aléatoire simulée et limitée à un voisinage restreint. Cette méthode d'exploration est adaptée pour évaluer l'apport de l'exploration d'UMLS en utilisant la cooccurrence pour s'orienter. Pour rappel, une matrice de transition est la matrice qui détermine la probabilité de passage entre tous les nœuds d'un graphe dans un processus de marche aléatoire.
- Une condition particulière qui se présente durant le processus de navigation sur les graphes orientés est celle de ne plus avoir de voisins à visiter. Pour remédier à cette situation, une implémentation comme celle de PageRank utilise la téléportation (saut aléatoire vers un autre nœud du graphe). Cependant et selon (Willis et Losee, 2013), ce n'est pas un modèle probable de navigation pendant la recherche de concepts pour une ontologie ou une base de connaissances dans notre implémentation, s'il n'y a pas de relation (arc) à partir du nœud présent, le comportement qui en résulte est de « quitter l'exportation du chemin » et initier une autre marche à partir du nœud de départ même si le nombre de marches n'est pas atteint.

- Pour la sélection aléatoire (que ce soit pour les nœuds avec cooccurrence ou totalement aléatoires) nous avons utilisé l’algorithme interne de génération aléatoire basé sur le générateur aléatoire de nombre appelé « Mersenne Twister <sup>28</sup>» fournis dans le langage de programmation PHP.
- Aussi, les résultats de notre méthode peuvent être légèrement différents entre les différentes exécutions, et ce, par le fait du caractère aléatoire de l’exploration et le choix des paramètres de départs comme la longueur de la marche et le nombre de marches à exécuter. Toutefois, l’augmentation du nombre de marches et l’utilisation de la cooccurrence contribueront grandement à stabiliser les résultats.

---

<sup>28</sup> Le Mersenne Twister est un générateur de nombres pseudo-aléatoires, réputé pour sa qualité, développé par Makoto Matsumoto et Takuji Nishimura en 1997. L’algorithme est basé sur un TGSFR et tient son nom d’un nombre premier de Mersenne. (Wikipedia)

#### 2.9.4 Exemple d'identification de concepts

Dans cet exemple nous présentons le résultat d'exploration d'UMLS pour l'une des règles de l'arbre de décision TAXI (voir la section 2.10.2.2). Le processus commence par l'utilisation des termes de la règle de décision ci-dessous.

Critically ill Child and Hemorrhagic shock and life-threatening => Transfusion RBC plasma: platelets in a ratio of 2:1:1 or 1:1:1 until bleeding is no longer life-threatening.

Le résultat de l'application de MetaMapLite sur cet ensemble de départ est présenté dans le tableau 2.9. Les concepts résultants (avec leur *cui* respectif) seront utilisés par l'algorithme d'exploration pour générer de nouveaux concepts susceptibles d'être pertinents pour la règle. Le tableau 2.10 présente les concepts ICPU résultants de l'exploration de l'algorithme avec les paramètres suivants NM : 10 ; LM : 5 ; Rela : CHD, PAR, RN, RB ; Sab: MTH, MeSH, SNOMED\_CT ; fr : seuil de visite = 4 ; freq >=1).

Tableau 2.9 Concepts issu du Mapping avec MetaMapLite.

Résultat de MetaMapLite			
1	C0032105 Plasma	11	C3537125 Life Threatening or Disabling Adverse Event
2	C4521445 Plasma Product	12	C1546953 Life threatening—Event Consequence
3	C0036982 Shock, Hemorrhagic	13	C0005841 Blood Transfusion
4	C0005821 Blood Platelets	14	C0014792 Erythrocytes
5	C1963076 Platelets Adverse Event	15	C0231218 Malaise
6	C0443116 Platelets Product	16	C0456603 Ratio
7	C0008059 Child	17	C1879316 Transfusion (procedure)
8	C0019080 Hemorrhage	18	C0199960 Transfusion — action
9	C0680063 Offspring	19	C1547037 data type — ratio
10	C2826244 Life Threatening		



Tableau 2.10 List de concepts résultants de l'exploration dans UMLS à partir d'un ensemble de termes de départs d'une règle d'un SADC

N	cui	concept	fréq	N	cui	concept	fréq
1	C0032105	Plasma	11	28	C3887242	National Cancer Institute common terminology criteria for adverse event grade finding	27
2	C0005889	Body Fluids	13	29	C1513302	Mild Adverse Event	4
3	C4521445	Plasma Product	11	30	C1513374	Moderate Adverse Event	5
4	C0036982	Shock, Hemorrhagic	11	31	C1546953	Life threatening—Event Consequence	11
5	C0019080	Hemorrhage	7	32	C0005841	Blood Transfusion	14
6	C0036974	Shock	8	33	C0032113	Plasma Exchange	4
7	C0005821	Blood Platelets	11	34	C0085430	Blood Component Transfusion	5
8	C0005773	Blood Cells	8	35	C0086548	Leukocyte Transfusion	4
9	C0007634	Cells	7	36	C0015236	Exchange Transfusion, Whole Blood	6
10	C0005767	Blood	4	37	C0014792	Erythrocytes	11
11	C1963076	Platelets Adverse Event	11	38	C0014802	Erythroid Precursor Cells	10
12	C0443116	Platelets Product	15	39	C0014762	Erythroblasts	7
13	C0301492	Human platelet product	7	40	C0231218	Malaise	20
14	C0443119	Platelets—CMV negative	4	41	C0024528	Malaise and fatigue	4
15	C0456388	Blood product	5	42	C0456603	Ratio	22
16	C0008059	Child	11	43	C1532109	Ratio value	5
17	C0259916	Disabled Children	4	44	C0237753	Number	5
18	C0018576	Disabled Persons	5	45	C1879316	Transfusion (procedure)	15
19	C0008101	Unwanted children	4	46	C0565532	Plasma/plasma substitute transfusion	4
20	C0027361	Persons	11	47	C0398314	Other intravenous transfusion NOS	6
21	C0018944	Hematoma	4	48	C0199960	Transfusion — action	11
22	C0680063	Offspring	24	49	C0574032	Infusion procedures	5
23	C0425383	Has stepmother	10	50	C0005791	Apheresis (procedure)	4
24	C0425384	Has stepfather	6	51	C1547037	data type—ratio	12
25	C1287235	Finding of personal status	11	52	C1265611	Quantity	24
26	C2732461	Conceived by in vitro fertilization	4	53	C0178602	Dosage	4
27	C2826244	Life Threatening	11	54	C3537125	Life Threatening or Disabling Adverse Event	14

## 2.10 Expérimentations et évaluation

Pour l'expérimentation et l'évaluation de notre méthodologie, nous avons jugé de la pertinence de l'approche, qui utilise une marche aléatoire modifiée et adaptée à UMLS, au nombre de concepts du domaine (du SADC) que la méthode pourrait identifier en variant différents paramètres et en comparant les résultats à une marche aléatoire d'exploration sans co-occurrence. L'importance de ces concepts sera définie par la « fréquence des visites » et sera utilisée ultérieurement (dans le chapitre suivant) pour vérifier l'amélioration de la recherche d'information.

### 2.10.1 Points importants et limites à prendre en considération pour l'évaluation

Il est difficile dans cette expérimentation de se comparer à d'autres méthodes d'exploration pour plusieurs raisons, notamment celles-ci :

- Ils n'existent pas d'approches qui utilisent des termes provenant des règles de décision pour retrouver les concepts importants entourant une tâche médicale et en utilisant des termino-ontologies d'UMLS.
- Les algorithmes qui identifient les concepts importants utilisent généralement et en globalité des ontologies existantes ou ils créent explicitement des ontologies pour valider leur approche. Dans notre cas, l'approche ne cherche pas à classer les concepts d'une ontologie, mais plutôt de découvrir et à classer les concepts au sein de grappes qui entourent des concepts de départ. Car nous pensons que de cette façon, nous bâtissons une base de connaissances plus grande pour le SADC suivant les concepts encapsulés dans ses règles de décisions. Cette extension de « concepts » pertinents sera testée dans le chapitre suivant pour améliorer la recherche d'information.

- La multitude de paramètres qui peuvent être utilisés pour l'exploration rend l'exercice d'évaluation très spécifique à la structure (UMLS) et au processus d'exploration utilisé (Marche aléatoire).

## 2.10.2 Ressources et prétraitement

### 2.10.2.1 Corpus de références et critères d'évaluations

Pour pouvoir évaluer les résultats de l'algorithme d'identification des concepts et leur appartenance au domaine d'application, nous avons utilisé l'unique dictionnaire terminologique (au moment de la rédaction) propre à la transfusion sanguine. Ce thésaurus regroupe plus de 2000 concepts, représentés par leurs termes en français et en anglais et par de nombreuses variantes (synonymes, acronymes...) utilisées dans la littérature du domaine.

Le « Thésaurus de la transfusion sanguine » réalisés par l'INIST<sup>29</sup> et publiés sur la plate-forme Loterre<sup>30</sup> qui a été retenus pour intégrer le projet européen *eTranslation TermBank* (eTTB)<sup>31</sup>. Ce projet, soutenu par la Commission européenne, vise à améliorer la qualité du système de traduction automatique de la communauté européenne pour les domaines de la santé, du droit des entreprises et de la protection des consommateurs. Il vise également à publier des ressources terminologiques sur la plate-forme lingu-

---

<sup>29</sup> Institut de l'information scientifique et technique INIST

<sup>30</sup> Développée par l'Inist-CNRS, **Loterre** (*Linked open terminology resources*) est une plate-forme de mise à disposition de ressources terminologiques scientifiques multidisciplinaires, fondée sur les technologies du web de données. (<https://www.loterre.fr/>)

<sup>31</sup> Le projet eTTB fournit des ressources terminologiques pour améliorer la qualité et la couverture de la traduction automatique. Le projet vise à identifier, collecter, traiter et fournir les ressources terminologiques appropriées aux secteurs (sans s'y limiter) : e-santé, système d'interconnexion des répertoires e-justice/commerce et résolution des conflits en ligne. [https://tilde.com/sites/default/files/downloads/CEF\\_eTranslation%20TermBank\\_poster.pdf](https://tilde.com/sites/default/files/downloads/CEF_eTranslation%20TermBank_poster.pdf)

tique de la Commission européenne. Le contenu du « Thésaurus de la transfusion sanguine » aborde en particulier les sous-domaines suivants : Acte transfusionnel ; produits sanguins ; appareillages de transfusion ; effets indésirables des transfusions.

Le dictionnaire a été publié en septembre 2016 par l'institut — Institut de l'information scientifique et technique INIST-CNRS (Centre national de recherche scientifique-France)<sup>32</sup>. Ce dictionnaire sera utilisé comme référence pour déterminer la pertinence des concepts résultants de l'extraction. Chaque concept découvert et mis en avant par l'algorithme sera confronté à ce dictionnaire pour le considérer comme appartenant ou pas au domaine de l'application médicale du SADC. Pour adapter ce dictionnaire terminologique à nos besoins, nous avons traité et codifié l'ensemble des entrées terminologiques du dictionnaire en utilisant MetaMapLite (Figure 2.18).

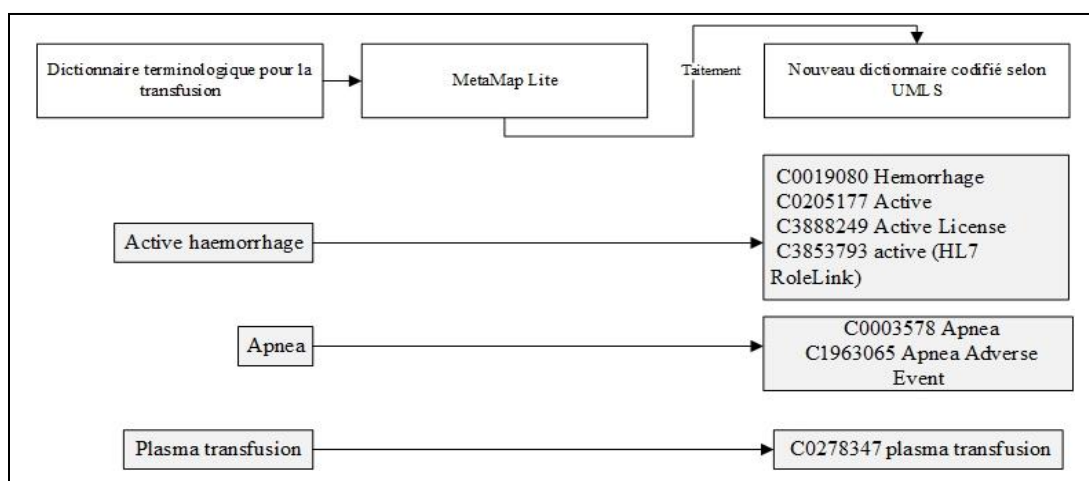


Figure 2.18 Codification selon UMLS du Thésaurus du dictionnaire terminologique de la transfusion sanguine.

<sup>32</sup> INIST a pour mission de faciliter l'analyse et la fouille de la production scientifique textuelle permet de répondre aux besoins de pilotage scientifique et d'extraction de connaissances. Pour cela elle offre un service terminologique avec l'objectif de constituer des ressources de représentation des connaissances facilitant la structuration des connaissances et de mutualiser les ressources terminologiques.

Le résultat est inséré dans une nouvelle table avec un champ d'information supplémentaire qui contient le CUI équivalent du terme issu de MetaMapLite L'objectif étant de pouvoir comparer les concepts résultant de l'exploration par l'algorithme et les concepts issus du dictionnaire.

#### 2.10.2.2 Concepts issus des règles de décisions et mappage MetaMapLite

Récemment, un groupe d'experts dénommé « *Transfusion and Anemia eXpertise Initiative* » (TAXI) (Bemba MM et al., 2018) a proposé un arbre de décision qui orienterait les praticiens sur la décision à prendre pour la transfusion de concentrés de globules rouges. Il est prévu de créer un système d'aide à la décision — un système expert — se basant sur cet arbre de décision afin d'améliorer l'adhérence des praticiens aux recommandations du groupe TAXI. Ces recommandations sont résumées dans les règles utilisées dans nos expérimentations (Figure 2.19). Pour plus d'information sur la transfusion restrictive et cet arbre de décision, voir le chapitre (section 4.8 du chapitre 4).

Remarque : Aucun prétraitement n'a été effectué sur les règles de décision, le texte a été utilisé tel que figurant dans l'arbre ni les abréviations ou autre contraction n'a fait l'objet de modification. Dans le tableau 2.11, l'ensemble des concepts retrouvés uniquement par le prétraitement avec MetaMapLite avant l'exploration avec notre algorithme. Ces concepts seront le point de départ de l'extraction.

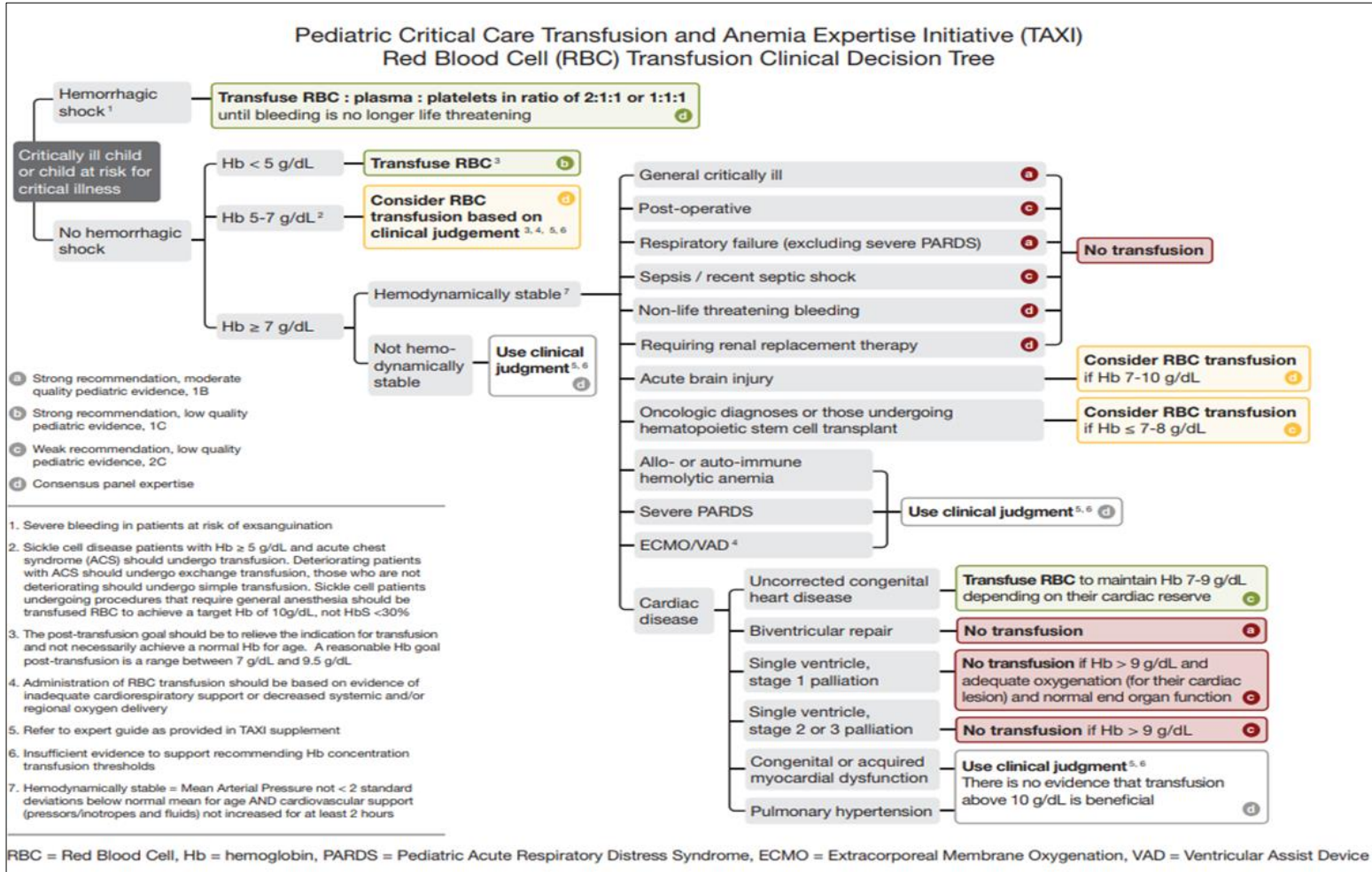


Figure 2.19 Arbre de décision TAXI pour la transfusion restrictive.

Les expériences que nous avons réalisées visent essentiellement à tester l'effet de diverses combinaisons de paramètres sur les performances de l'algorithme de marche aléatoire orienté par la cooccurrence et la marche aléatoire naturelle. La découverte de nouveaux concepts pertinents au domaine d'activité sera prise en considération. La cooccurrence des concepts est puisée dans la Table MROC et sera un élément déterminant pour les évaluations.

Les termes de dépôts par MetaMapLite sont ceux des règles de décision dans le tableau 2.11. Ces règles sont issues de l'arbre de décision TAXI pour la transfusion sanguine. Le dictionnaire utilisé pour identifier la pertinence de ces nouveaux concepts est celui de l'INIST. L'algorithme parcourt la table MRCONS et MRREL uniquement pour les trois termino-ontologies d'UMLS (MSH, MTH, US\_SNOWMED) que nous considérons que cela est suffisant pour couvrir le domaine de la transfusion. Avec le paramétrage des variables suivantes :

- Paramétrage du nombre de marches ;
- Paramétrage de la longueur de la marche ;
- Paramétrage du type de relations ;
- Paramétrage du seuil d'élagage des concepts suivant le seuil de visite ;
- Paramétrage de la marche aléatoire sans ou avec co-occurrence des concepts ;
- Paramètre du seuil de cooccurrence.

Note importante : beaucoup de paramètres peuvent être sollicités pour nos expérimentations, cependant nous nous sommes limités à quelques-uns pour ne pas perdre en vue nos objectifs.

Tableau 2.11 Règles et concepts de départs pour l'ICPU et les différentes expérimentations.

Règle	Conditions	Decisions	Concepts initiaux retrouvés par MetaMapLite
01	[Critically ill child] and [Hemorrhagic shock]	[Transfusion RBC plasma: platelets in a ratio of 2:1:1 or 1:1:1 until bleeding is no longer life threatening]	C0036982 Shock, Hemorrhagic C1515981 Critical Illness C0008059 Child C0680063 Offspring C0036983 Septic Shock C0270611 Brain Injuries C0032105 Plasma C4521445 Plasma Product C0018956 Hematopoietic stem cells C0439267 Gram per Deciliter C0018956 Hematopoietic stem cells C0439267 Gram per Deciliter
02	[Critically ill child] and [No Hemorrhagic shock] and [Hb <5 g/dl]	(Transfusion RBC)	C0231218 Malaise C0205478 Oncologic C0205390 Phase C0456603 Ratio C0332185 Recent C4319951 Repair C3891294 Safety Reporting Documentation C1306673 Stage C1504389 Stem cell transplant C0374711 Surgical repair C1300072 Tumor stage C0077973 VAD I protocol C0205178 acute C0699808 biventricular C1547037 data type—ratio C0340515 myocardial dysfunction
03	[Critically ill child] and [No Hemorrhagic shock] and [Hb >=5 and Hb <7 g/dl]	consider RBC transfusion based on clinical judgment	C0278627 oncology field C4072785 uncorrected C0120777 HB 7 C0032790 Postoperative Period C0441767 Stage level 2 C3537125 Life Threatening or Disabling Adverse Event C0852336 Immune Hemolytic Anemia C2826244 Life Threatening C0015357 Extracorporeal Membrane Oxygenation C3812897 General medical service
04	[Critically ill child] and [No Hemorrhagic shock] and [Hb >7 g/dl] and [Hemodynamically stable] and [[general critical ill] or [Postoperative] or [Respiratory Failure excluding PARDS] or [Sepsis/recent septic shock] or [no life-threatening bleeding]	No transfusion	C0005841 Blood Transfusion C1963076 Platelets Adverse Event C0443116 Platelets Product C0005841 Blood Transfusion C1547311 patient condition code Stable C0205360 Stable status C1879316 Transfusion (procedure) C0199960 Transfusion—action C0018799 Heart Diseases C0019080 Hemorrhage C0014792 Erythrocytes C0205210 Clinical C4523057 Clinical Batch C0022423 Judgement C0009678 Congénital MeSH qualifier C1744681 Congenital qualifier C0020542 Pulmonary Hypertension C1963220 Pulmonary Hypertension Adverse Event C1145670 Respiratory Failure C4050466 Borg Category-Ratio 10 Perceived Exertion Score 5 C0152424 Common ventricle C1546953 Life threatening—Event Consequence C0030231 Palliative Care C0205082 Severe (severity modifier) C4050465 Severe Extremity Pain C3274516 Single Ventricle Defect C0457083 Usage C0750591 consider C0042153 utilization qualifier C0439661 Acquired (qualifier value)
05	[Critically ill child] and [No Hemorrhagic shock] and [Hb >7 g/dl] and [Hemodynamically stable] and [acute brain injury] or [Oncologic diagnoses or those undergoing hematopoietic stem cell transplant]	Consider RBC transfusion if Hb >7 and Hb <=10	C0002144 Allopurinol C1441488 Auto C1705938 Base—General Qualifier C1527178 Basis—conceptual entity C1522601 Cardiac—anatomy qualifier C0242354 Congenital Disorders C1511545 Critical C0011900 Diagnosis
07	[Critically ill child] and [No Hemorrhagic shock] and [Hb >7 g/dl] and [Hemodynamically stable] and [[allo- or autoimmune hemolytic anemia] or [severe PARDS] or [ECMO/VAD]]	Use clinical judgment.	
08	[Critically ill child] and [No Hemorrhagic shock] and [Hb >7 g/dl] and [Hemodynamically stable] and cardiac disease] And [uncorrected congenital disease]	Transfuse RBC to maintain Hb 7–9 g/dl depending on their cardiac severe	
09	[Critically ill child] and [No Hemorrhagic shock] and [Hb >7 g/dl] and [Hemodynamically stable] and cardiac disease] and [[biventricular repair] or [single ventricle, stage one palliation] or [single ventricle, stages 2 and 3 palliation]	No Transfusion	
10	Critically ill child] and [No Hemorrhagic shock] and [Hb >7 g/dl] and [Hemodynamically stable] and cardiac disease] and [congenital or acquired myocardial dysfunction] or [pulmonary hypertension]]	use clinical judgment.	



Voici quelques annotations et acronymes importants pour la suite de l'expérimentation (Tableau 2.12) :

Tableau 2.12 Table des acronymes et annotation des variables de paramétrage de l'ICPU.

Annotation	Description
Sources termino-ontologiques	SAB
Nombre de marche	NM (nombre de marche à partir de chaque concept initial)
Longueur de la marche	LM
Groupes de relations	R1 : CHD, PAR, QB, RL, SY (Hiérarchiques — synonymie) R2 : CHD, PAR, RO (Hiérarchiques – associatives) R3 : CHD, PAR, RN, RB (hiérarchiques) R4 : RO (associative)
SV	Seuil de visite
CR	Nombre global de concepts retrouvés par l'algorithme
CP	Nombre de concepts pertinents au domaine.
ACO	Avec co-occurrence
SCO	Sans co-occurrence
ACO-4	Choix du prochain nœud avec cooccurrence en premier et aléatoirement sinon. Avec restitution uniquement des nœuds ayant été visités 4 fois et plus.
ACO-6	Item que la ACO-4, mais avec restitution de nœuds ayant été visité 6 fois et plus.
SCO-4	Choix du prochain nœud aléatoirement. Avec restitution uniquement des nœuds ayant été visités 4 fois et plus.
SCO-6	Item que la SCO-4, mais avec restitution de nœuds ayant été visité 6 fois et plus
R1-ACO-4	Relations utilisées est CHD, PAR, QB, RL, SY, Choix du prochain nœud avec cooccurrence en premier et aléatoirement sinon. Avec restitution uniquement des nœuds ayant été visités 4 fois et plus

Voici comme exemple le résultat de l'exécution de l'algorithme avec un certain paramétrage de départ et une variabilité sur le nombre d'itérations et la longueur de la marche (Tableau 2.13). Pour ce qui est de la longueur de la marche et le nombre de

marches, nous avons pris comme référence les paramétrages de départ qui ont été utilisés pour l'extraction de concepts sémantiquement reliés dans un graphe de pages de Wikipédia (Dallmann *et al.*, 2016). Une longueur de marche de NM=10 et une fenêtre de LM=3 (longueur de marche) sont nos points de références. Toutefois, nous avons varié ces paramètres durant l'expérimentation. Dans l'étude de Dallmann *et al.* (2016), les auteurs présente une méthode d'extraire les concepts clé à partir d'un graphe représentant une portion de Wikipédia avec un processus d'exploration aléatoire.

Tableau 2.13 Exemple d'exploration avec un NM égale à 10.

	NM =10																							
	LM =3								LM =4								LM=5							
	SV = 4				SV = 6				SV = 4				SV = 6				SV = 4				SV = 6			
	ACO		SCO		ACO		SCO		ACO		SCO		ACO		SCO		ACO		SCO		ACO		SCO	
	CR	CP	CR	CP	CR	CP	CR	CP	CR	CP	CR	CP	CR	CP	CR	CP	CR	CP	CR	CP	CR	CP	CR	CP
Itéra-1	163	70	149	63	129	64	111	63	175	74	163	68	131	65	124	60	203	79	191	66	146	66	122	62
Itéra-2	163	72	142	66	121	63	111	59	185	76	165	66	132	65	117	62	199	74	198	68	149	69	127	59
Itéra-3	165	69	155	67	121	63	107	58	186	75	166	68	134	68	122	63	200	73	179	68	146	68	130	63
Moy	164	70	149	65	124	63	110	60	182	75	165	67	132	66	121	62	201	75	189	67	147	68	126	61

Constantes : Source vocabulaires : MSH, MTH, US\_SNOWMED ; cooccurrence avec fréquence minimale égale à 1 (au moins une cooccurrence avec le terme initial) ; Groupes de relations = R4 (associative) variables : Nombre de marches (NM) =10 ; Longueur de la marche (LM) = 3, 4,5.

Nous avons opté pour trois itérations pour calculer la moyenne des concepts retrouvée et atténuer la variabilité des résultats. Cette variabilité est normale dans un contexte de choix aléatoire des nœuds à visiter par l'algorithme.

### 2.10.3 Variable d'évaluation

Pour pouvoir déterminer l'efficacité de l'algorithme, nous proposons l'utilisation de la variable d'évaluation (qui découle de la notion de précision en recherche d'information, mais avec une adaptation au contexte. Rappelons que la précision est le nombre de concepts pertinents/nombre de concepts retrouvés).

- (NCP) : Pourcentage des nouveaux concepts pertinents découverts avec l'algorithme et qui ne font pas partie des concepts initiaux pertinents après le mappage avec MetaMapLite.

$$NCP = \frac{\sum_{it=1}^n \frac{(CP_{it}) - CP_{init}}{CP_{init}}}{n}$$

- (CP<sub>init</sub>) : Concepts initiaux pertinents à la suite du mappage avec MetaMapLite. Nous avons exclu ces concepts du calcul, car nous voulions connaître l'effet réel de l'algorithme. Dans le cas des concepts issus des règles de l'arbre de décision utilisé, cela revient à soustraire 57 concepts retrouvés dans le dictionnaire de L'INST sur les 84 concepts produit par MetaMapLite.
- (CP<sub>it</sub>) : nombre de concepts pertinents retournés à chaque itération par l'exploration.
- (n) : Le nombre d'itérations.
- (Moy en %) : c'est le rapport CP/CR (CP : nombre de concepts pertinents sur l'ensemble des concepts retrouvés CR).

Le tableau 2.14 présente le résultat d'une exploration avec trois différentes exécutions de l'algorithme avec le même paramétrage.

Tableau 2.14 Exemple de résultats d'exploration avec les variables d'évaluation pour un NM = 10.

	NM =10																							
	LM =3								LM =4								LM=5							
	SV = 4				SV = 6				SV = 4				SV = 6				SV = 4				SV = 6			
	ACO		SCO		ACO		SCO		ACO		SCO		ACO		SCO		ACO		SCO		ACO		SCO	
	CR	CP	CR	CP	CR	CP	CR	CP	CR	CP	CR	CP	CR	CP	CR	CP	CR	CP	CR	CP	CR	CP	CR	CP
itération-1	163	70	149	63	129	64	111	63	175	74	163	68	131	65	124	60	203	79	191	66	146	66	122	62
itération-2	163	72	142	66	121	63	111	59	185	76	165	66	132	65	117	62	199	74	198	68	149	69	127	59
itération-3	165	69	155	67	121	63	107	58	186	75	166	68	134	68	122	63	200	73	179	68	146	68	130	63
%(ACO-SCO)	7.7%				5.6%				11.4%				7.0%				11.9%				10.3%			
NCP en %	23.4 %		14.6 %		11.1 %		5.3 %		31.6 %		18.1 %		15.8 %		8.2 %		32.2 %		18.1 %		18.7 %		7.6 %	
Moy en %	43.0 %		43.9 %		51.2 %		54.7 %		41.2 %		40.9 %		49.9 %		51.0 %		37.5 %		35.6 %		46.0 %		48.5 %	
	164	123	149	112	124	109	110	97	182	130	165	116	132	109	121	98	201	141	189	136	147	117	126	103

Sources termino-ontologies (MSH, MTH, US\_SNOMED) ; relation (R1) : CHD, PAR, QB, RL, SY ; cooccurrence >= 1.

Exemple de calcul : Dans la colonne CP d'ACO (LM=3 et SV=4) du tableau 2.14, nous avons trois valeurs 70, 72, 69 représentant le nombre de CP (Concepts pertinents) retrouvé dans chaque itération, la moyenne est de 70,33. Si nous calculons le pourcentage, de nouveaux concepts retrouvés par rapport à ceux retrouvés par MetaMapLite (57 concepts). Le calcul sera égal à  $((70.33-57)/57)$  et ceci, donne 23,4 %. Le tableau 2.15 présente, à titre indicatif, une partie du résultat d'une exploration de l'ICPU avec l'indication si le concept figure dans le dictionnaire de l'INST.

Tableau 2.15 Exemple du résultat ICPU avec vérification de l'existence des concepts dans le dictionnaire de l'INST. (LM=5; NM=10; SV=6; OC>=1; SAB=MTH, MSH, US\_SNOMED).

N	Concepts ICPU		D		N	Concepts ICPU		D	Dict. de L'INST	N	Concepts ICPU		D	Dict. de L'INST
1	C0014792	Erythrocytes	11	Concepts retrouvés dans le dictionnaire de L'INST	23	C0024501	Maintenance	11	Concepts retrouvés dans le dictionnaire de L'INST	45	C3812897	General medical service	21	Concepts retrouvés dans le dictionnaire de L'INST
2	C0005821	Blood Platelets	11		24	C0009678	Congenital MeSH qualifier	13		46	C0205246	Generalized	11	
3	C0005841	Blood Transfusion	12		25	C1744681	Congenital qualifier	14		47	C0018799	Heart Diseases	11	
4	C1879316	Transfusion (procedure)	14		26	C1145670	Respiratory Failure	15		48	C2826244	Life Threatening	11	
5	C0199960	Transfusion—action	12		27	C0002144	Allopurinol	11		49	C1546953	Life threatening—Event Consequence	11	
6	C0443116	Platelets Product	20		28	C1963220	Pulmonary Hypertension Adverse E.	11		50	C0374711	Surgical repair	12	
7	C0019080	Hemorrhage	11		29	C0439661	Acquired (qualifier value)	22		51	C0278627	oncology field	11	
8	C0030231	Palliative Care	11		30	C3245488	Acquired Name	11		52	C0042153	utilization qualifier	14	
9	C0005773	Blood Cells	8		31	C0015127	Etiology aspects	20		53	C0439267	Gram per Decilitre	12	
10	C0011900	Diagnosis	12		32	C1265611	Quantity	17		54	C0560150	Mass concentration	27	
11	C0243026	Sepsis	11		33	C0457083	Usage	12		55	C0021290	Neonatal disorder	6	
12	C1090821	Sepsis (Invertebrate)	11		34	C3537125	Life Threatening or Disabling Ad	14		56	C1547311	Patient Condition Code — Stable	11	
13	C0036690	Septicemia	15		35	C0032790	Postoperative Period	11		57	C1272746	Change patterns	12	
14	C0205178	acute	18		36	C0018787	Heart	11		58	C1533685	Injection procedure	7	
15	C4050466	Borg Category	11		37	C0038250	Stem cells	6		59	C0205159	General adjective	25	
16	C0237753	Number	8		38	C1705938	Base—General Qualifier	11		60	C0022423	Judgment	11	
17	C0205082	Severe (severity modifier)	21		39	C1527178	Basis—conceptual entity	11		61	C0033211	Problem Solving (mental process)	13	
18	C4050465	Severe Extremity Pain	11		40	C0007222	Cardiovascular Diseases	6		62	C0039869	Thinking, function	10	
19	C0559895	Periods of life	18		41	C0035239	Respiratory Therapy	18		63	C0699808	biventricular	11	
20	C0024115	Lung diseases	8	42	C0184633	Oxygen Therapy Care	7	64	C1547037	data type—ratio	13			
21	C0152424	Common ventricle	14	43	C3887621	Infectious disease service	9	65	C0340515	myocardial dysfunction	13			
22	C0039548	terminal patient care	8	44	C0587569	Physician service	10	66	C0025098	Medical oncology specialty	7			
				Concepts non retrouvés dans le dictionnaire				Concepts non retrouvés dans le dictionnaire						Concepts non retrouvés dans le dictionnaire

#### 2.10.4 Impact du biais de cooccurrence

Le biais de cooccurrence de la marche aléatoire que nous avons implémenté et expérimenté permet de choisir aléatoirement le prochain sommet à visiter parmi des concepts qui co-occurrent avec le concept de départ. Le cas échéant, choisir au hasard le prochain sommet à visiter parmi le voisinage du concept.

Les tableaux de la figure 2.20 sont le résumé de quatre tableaux comme celui du tableau 2.14 pour chaque relation R1, R2, R3 et R4 avec ou sans cooccurrence et par rapport à deux seuils de visite (4 et 6) pour retenir ou élaguer les concepts. (Voir le tableau 2.12).

ACO SV=4 et NM =10				SCO SV=4 et NM =10			
	LM=3	LM=4	LM=5		LM=3	LM=4	LM=5
R1-ACO-4	23.4%	31.6%	32.2%	R1-SCO-4	14.6%	18.1%	18.1%
R2-ACO-4	19.3%	20.5%	32.7%	R2-SCO-4	12.3%	12.3%	18.7%
R3-ACO-4	29.2%	30.4%	40.4%	R3-SCO-4	15.2%	24.6%	<b>26.9%</b>
R4-ACO-4	28.7%	33.9%	<b>46.8%</b>	R4-SCO-4	16.4%	22.8%	25.1%
Moyenne	25.1%	29.1%	<b>38.0%</b>	Moyenne	11.0%	13.7%	<b>18.1%</b>

ACO SV=6 et NM =10				SCO SV=6 et NM =10			
	LM=3	LM=4	LM=5		LM=3	LM=4	LM=5
R1-ACO-6	11.1%	15.8%	18.7%	R1-SCO-6	5.3%	8.2%	7.6%
R2-ACO-6	5.8%	5.8%	11.1%	R2-SCO-6	4.7%	2.9%	8.2%
R3-ACO-6	12.3%	17.5%	21.1%	R3-SCO-6	6.4%	9.9%	<b>14.0%</b>
R4-ACO-6	14.6%	15.8%	<b>21.6%</b>	R4-SCO-6	7.6%	6.4%	9.9%
Moyenne	14.6%	19.4%	<b>22.2%</b>	Moyenne	6.0%	6.9%	<b>9.9%</b>

Figure 2.20 Différents tableaux de l'impact de la cooccurrence avec marche aléatoire ; NM=10.

La figure 2.21 est la représentation graphique des résultats des tableaux énumérés dans la figure 2.20. Ces courbes indiquent que la cooccurrence avec un seuil de visite égale à 4 ou 6 a contribué à presque doubler le pourcentage de concepts pertinents retrouvés (9.9 à 18,1 % et 22.2 à 38 %) avec les mêmes conditions de départ et une longueur de marche de 5 pas. Noter que le chiffre à la fin de l'abréviation indique le nombre des visites d'un nœud. Par exemple, SCO/ACO-4 a été visité 4 fois tout comme SCO/ACO-6 qui a été visité 6 fois (en dessous desquels, les concepts seront élagués).

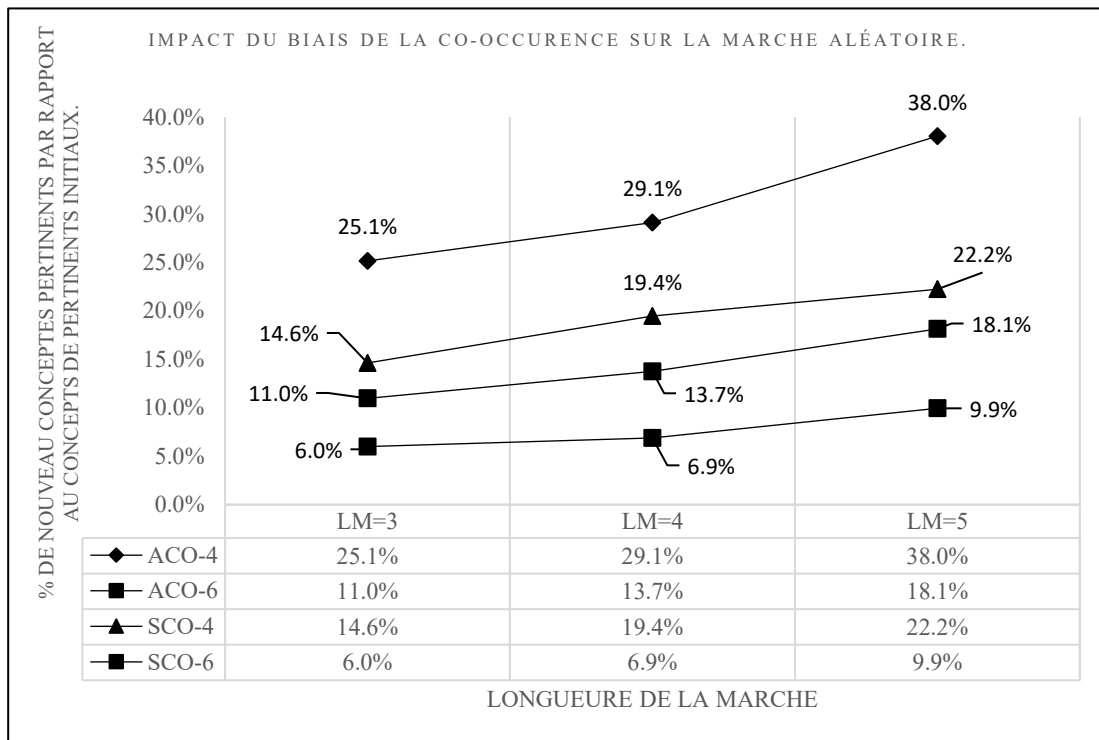


Figure 2.21 Impact du biais de la cooccurrence sur l'exploration de l'ICPU ; NM =10.

ACO SV=4 et NM =15

	LM=3	LM=4	LM=5
R1-ACO-4	29.8%	38.6%	52.6%
R2-ACO-4	31.6%	49.7%	53.8%
R3-ACO-4	41.5%	50.9%	56.1%
R4-ACO-4	45.6%	52.6%	63.2%
Moyenne	37.1%	48.0%	56.4%

SCO SV=4 et NM =15

	LM=3	LM=4	LM=5
R1-SCO-4	16.4%	18.1%	34.5%
R2-SCO-4	15.8%	25.1%	31.6%
R3-SCO-4	24.6%	33.9%	36.0%
R4-SCO-4	26.3%	34.5%	32.7%
Moyenne	20.8%	27.9%	33.7%

ACO SV=6 et NM =15

	LM=3	LM=4	LM=5
R1-ACO-6	17.5%	22.2%	28.7%
R2-ACO-6	8.0%	22.2%	24.6%
R3-ACO-6	26.9%	27.5%	39.8%
R4-ACO-6	22.8%	31.0%	36.8%
Moyenne	18.8%	25.7%	32.5%

SCO SV=6 et NM =15

	LM=3	LM=4	LM=5
R1-SCO-6	8.2%	7.6%	9.9%
R2-SCO-6	4.7%	11.1%	10.5%
R3-SCO-6	13.5%	14.0%	21.1%
R4-SCO-6	13.5%	22.2%	16.4%
Moyenne	10.0%	13.7%	14.5%

Figure 2.22 Différents tableaux de l'impact de la cooccurrence avec marche aléatoire de NM=15.

La même tendance est constatée pour un nombre de marches un peu plus important (Figure 2.22 et sa représentation graphique avec la figure 2.23). La cooccurrence a contribué à presque doubler le pourcentage de concepts pertinents retrouvés avec respectivement 14,5 % vs 32.5 % et 33.6 vs 56,4 %. Il en ressort clairement que le biais de cooccurrence lors de la marche aléatoire permet d'augmenter les chances de retrouver des concepts potentiellement pertinents par rapport, quelles que soient les relations choisies (hiérarchique, associative...).



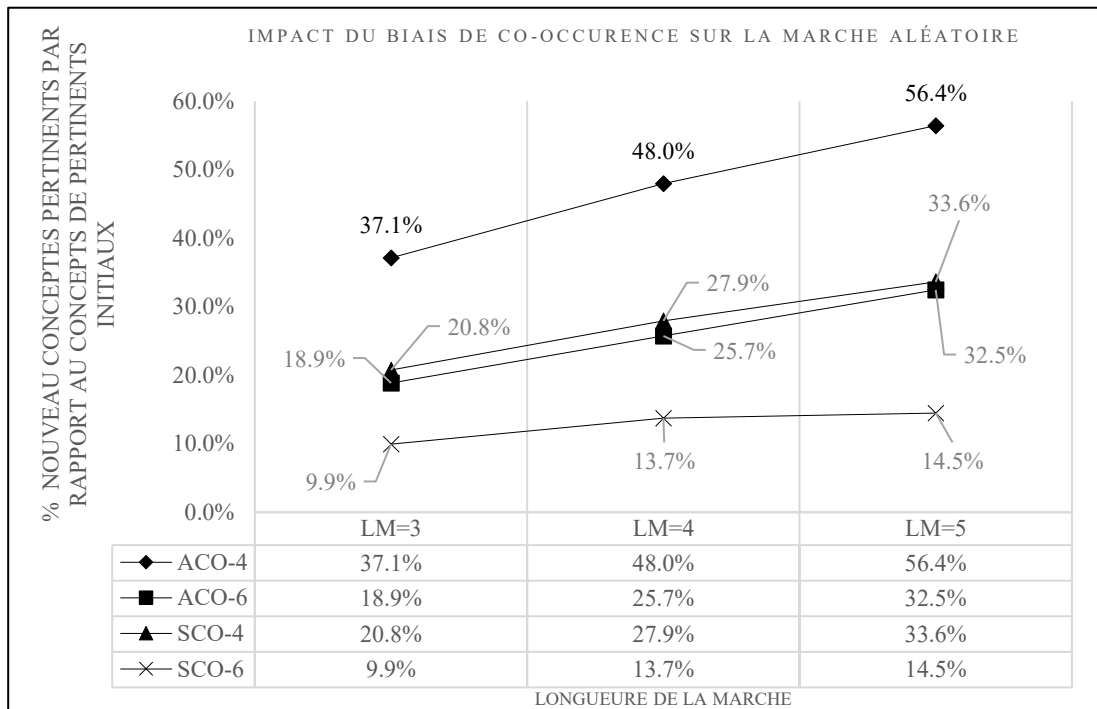


Figure 2.23 Impact du biais de la cooccurrence sur l'exploration de l'ICPU ; NM =15.

Dans le tableau 2.16, le résultat comparatif des concepts potentiellement pertinents (avec et sans le biais de la co-occurrence) d'une seule itération paramétrée comme suit : le nombre de marches NM=15 ; LM=5 ; SV =4 et co-occu >=1 ; SAB = MTH, MSH et US\_SNOMED. Le résultat statistique de l'itération est le suivant : 225 concepts retrouvés avec une marche biaisée par la co-occurrence et 201 concepts sans co-occurrences. Les concepts potentiellement pertinents sont 75 concepts et 63 respectivement. Le nombre de concepts potentiellement pertinents commun aux deux marches est de 59 concepts et 16 concepts supplémentaires retrouvés par la co-occurrence. Le nombre de concepts pertinents est plus important par la recherche biaisée que sans cette dernière. Ceci démontre qu'une orientation plus judicieuse de la marche aléatoire, en utilisant la co-occurrence des concepts dans notre cas, est meilleure qu'une marche totalement aléatoire.

Tableau 2.16 Exemple de concepts supplémentaires potentiellement pertinents issus de l'exploration ICPU avec cooccurrences.

Concepts retrouvés avec co-occurrence		Concepts retrouvés sans co-occurrence		Concepts retrouvés avec co-occurrence		Concepts retrouvés sans co-occurrence	
C0014792	Erythrocytes	C0014792	Erythrocytes	C1265611	Quantity	C1265611	Quantity
C0005821	Blood Platelets	C0005821	Blood Platelets	C0457083	Usage	C0457083	Usage
C0005841	Blood Transfusion	C0005841	Blood Transfusion	C3537125	Disabling Adverse Event	C3537125	Disabling Adverse Event
C1879316	Transfusion (procedure)	C1879316	Transfusion (procedure)	C0032790	Postoperative Period	C0032790	Postoperative Period
C0199960	Transfusion—action	C0199960	Transfusion—action	C0015354	Extracorporeal Circulation	C0015354	Extracorporeal Circulation
C0456388	Blood product	C0456388	Blood product	C0018787	Heart	C0018787	Heart
C0018956	Hematopoietic stem cells	C0018956	Hematopoietic stem cells	C1705938	Base—General Qualifier	C1705938	Base—General Qualifier
C0005791	Apheresis (procedure)	C0005791	Apheresis (procedure)	C1527178	Basis—conceptual entity	C1527178	Basis—conceptual entity
C1504389	Stem cell transplant	C1504389	Stem cell transplant	C3812897	General medical service	C3812897	General medical service
C1963076	Platelets Adverse Event	C1963076	Platelets Adverse Event	C0205246	Generalized	C0205246	Generalized
C0443116	Platelets Product	C0443116	Platelets Product	C0018799	Heart Diseases	C0018799	Heart Diseases
C0019080	Hemorrhage	C0019080	Hemorrhage	C1444637	In the past	C1444637	In the past
C0030231	Palliative Care	C0030231	Palliative Care	C2826244	Life Threatening	C2826244	Life Threatening
C0005773	Blood Cells	C0005773	Blood Cells	C1546953	Life threatening	C1546953	Life threatening
C0011900	Diagnosis	C0011900	Diagnosis	C0205390	Phase	C0205390	Phase
C0036690	Septicemia	C0036690	Septicemia	C3891294	Safety Reporting Doc	C3891294	Safety Reporting Doc
C0205178	acute	C0205178	acute	C0374711	Surgical repair	C0374711	Surgical repair
C0014508	epidemiologic	C0014508	epidemiologic	C0278627	oncology field	C0278627	oncology field
C4050466	Borg Category-Ratio 10	C4050466	Borg Category-Ratio 10	C0042153	utilization qualifier	C0042153	utilization qualifier
C0205082	Severe (severity modifier)	C0205082	Severe (severity modifier)	C0005767	Blood	Concepts additionnels retrouvés par la co-occurrence	
C4050465	Severe Extremity Pain	C4050465	Severe Extremity Pain	C0015236	Exchange Transfusion, Whole Blood		
C0024501	Maintenance	C0024501	Maintenance	C0035286	Reticulocytes		
C0009678	Congenital MeSH qualifier	C0009678	Congenital MeSH qualifier	C0007634	Cells		
C1744681	Congenital qualifier	C1744681	Congenital qualifier	C0237753	Numbers		
C1145670	Respiratory Failure	C1145670	Respiratory Failure	C0020538	Hypertensive disease		
	Allopurinol	C0002144	Allopurinol	C0086818	Platelet Transfusion		
C0205210	Clinical	C0205210	Clinical	C0040732	Transplantation		
C4523057	Clinical Batch	C4523057	Clinical Batch	C0085430	Blood Component Transfusion		
C0574032	Infusion procedures	C0574032	Infusion procedures	C2939425	Neonatal (qualifier value)		
C0205360	Stable status	C0205360	Stable status	C0205191	chronic		
C0005953	Bone Marrow	C0005953	Bone Marrow	C0018747	Health Services		
C0020542	Pulmonary Hypertension	C0020542	Pulmonary Hypertension	C0038250	Stem cells		
C1963220	Pulmonary Hypertension	C1963220	Pulmonary Hypertension	C0443172	Changed status		
C0439661	Acquired (qualifier value)	C0439661	Acquired (qualifier value)	C0086548	Leukocyte Transfusion		
C3245488	Acquired Name	C3245488	Acquired Name	C0205374	Transitory		

### 2.10.5 Impact du choix des groupes de relations

Pour cette marche aléatoire avec cooccurrence, certains groupes de relations peuvent être plus avantageux à utiliser pour retrouver des concepts pertinents au domaine que d'autres types de relations. Les résultats (Figure 2.24 et figure 2.25) de notre expérimentation nous montrent que les relations de type « uniquement associatives » désignées sous l'appellation « RO » dans UMLS (relations associatives) ainsi que les relations de types hiérarchiques (CHD, PAR, RN, RB) se démarquent des autres types. Ces dernières sont représentées par la R3 et la R4 avec respectivement 40.4 et 46.8 % de nouveaux concepts pertinents identifié par le processus pour un NM de 10. 56,1 et 63.2 % pour un NM de 15. Ces résultats pourraient être un bon indicateur pour le choix des relations pour mieux explorer les ontologies d'UMLS.

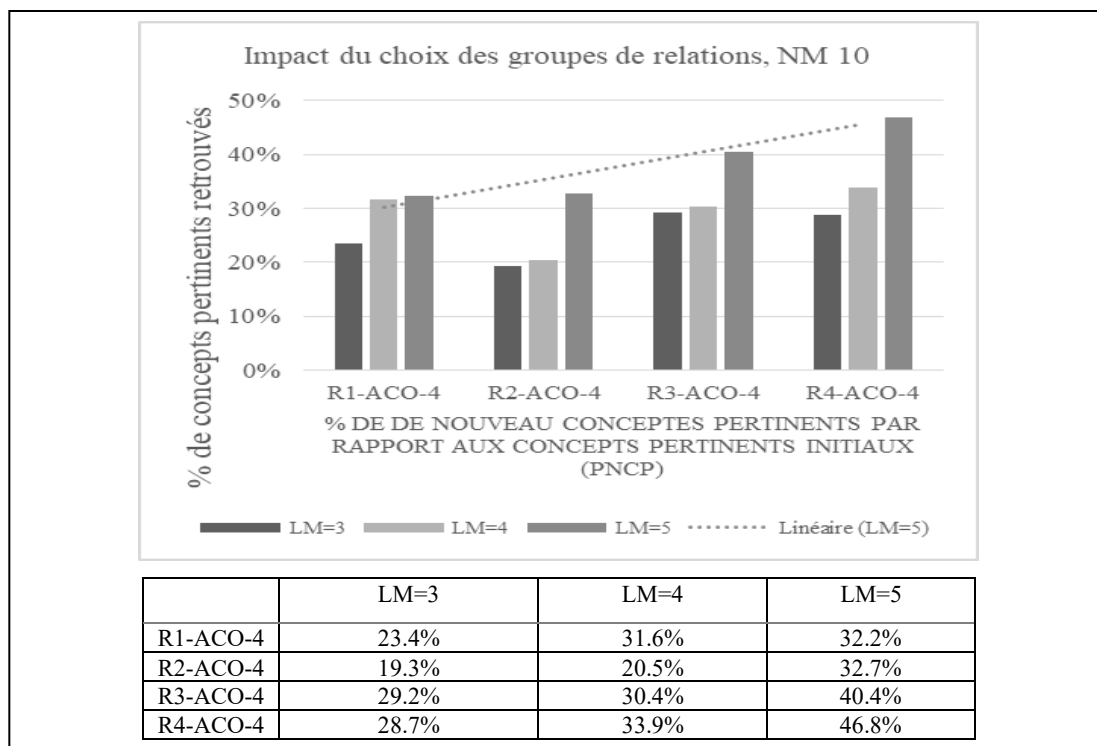


Figure 2.24 Impact du choix des groupes de relation, NM =10.

Pour notre chapitre suivant (recherche d'information par extension ontologique), nous avons choisi l'une des deux relations R3 (hiérarchiques) et R4 (associative) avec une longueur de marche de 30 pas).

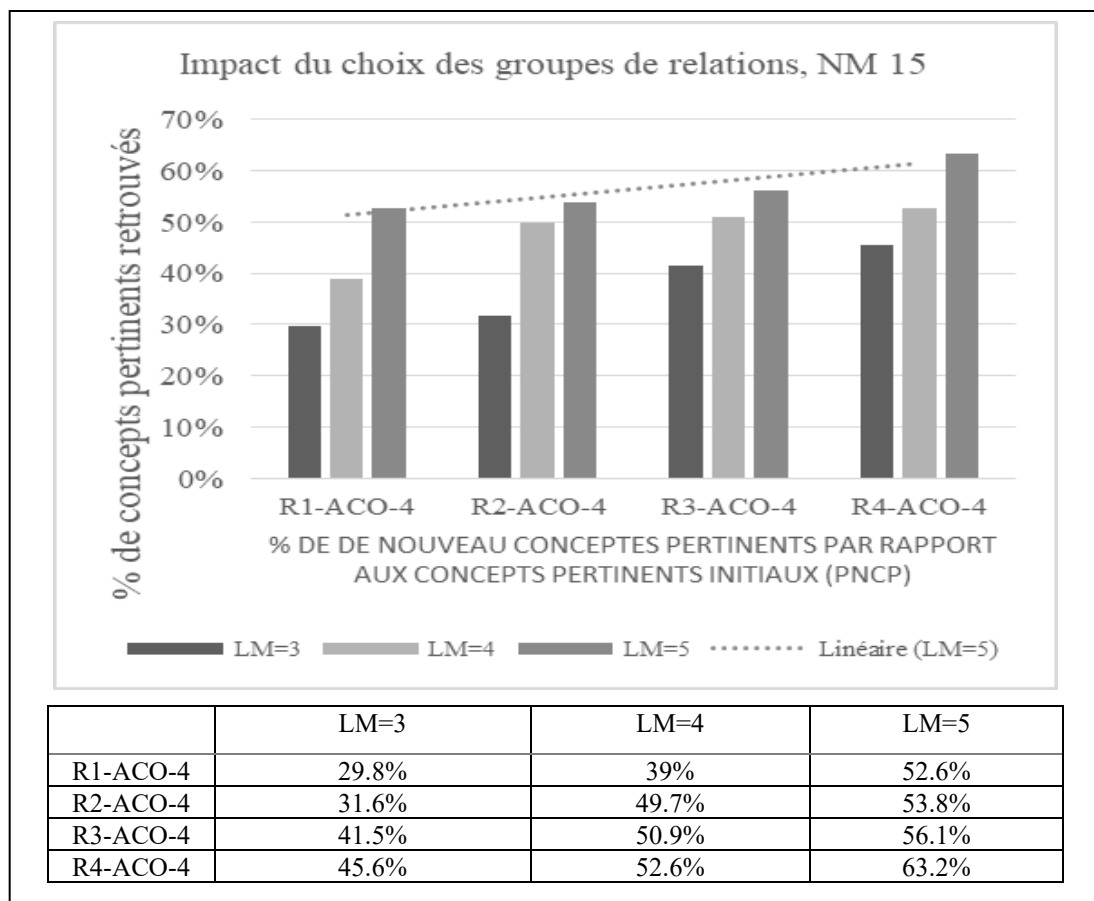


Figure 2.25 Impact du choix des groupes de relation, NM =15.

#### 2.10.6 Impact du seuil de cooccurrence

Dans cette phase d'expérimentation, nous avons cherché à connaître l'impact du choix du seuil de cooccurrence. Pour nous, l'augmentation ou la diminution du seuil pourrait

reflète le degré de relation entre les deux concepts « *relatedness* ». Favoriser cette relation pourrait aider à trouver plus de concepts pertinents avec moins d'effort. Les résultats ci-dessous, montrent exactement notre intuition, plus ce seuil est important pour élire les concepts qui pourront être visités, plus le nombre de concepts pertinents retrouvés est important (voir les graphes, les tableaux et l'explication ci-dessous).

- NM =10, LM=5 et SV =4

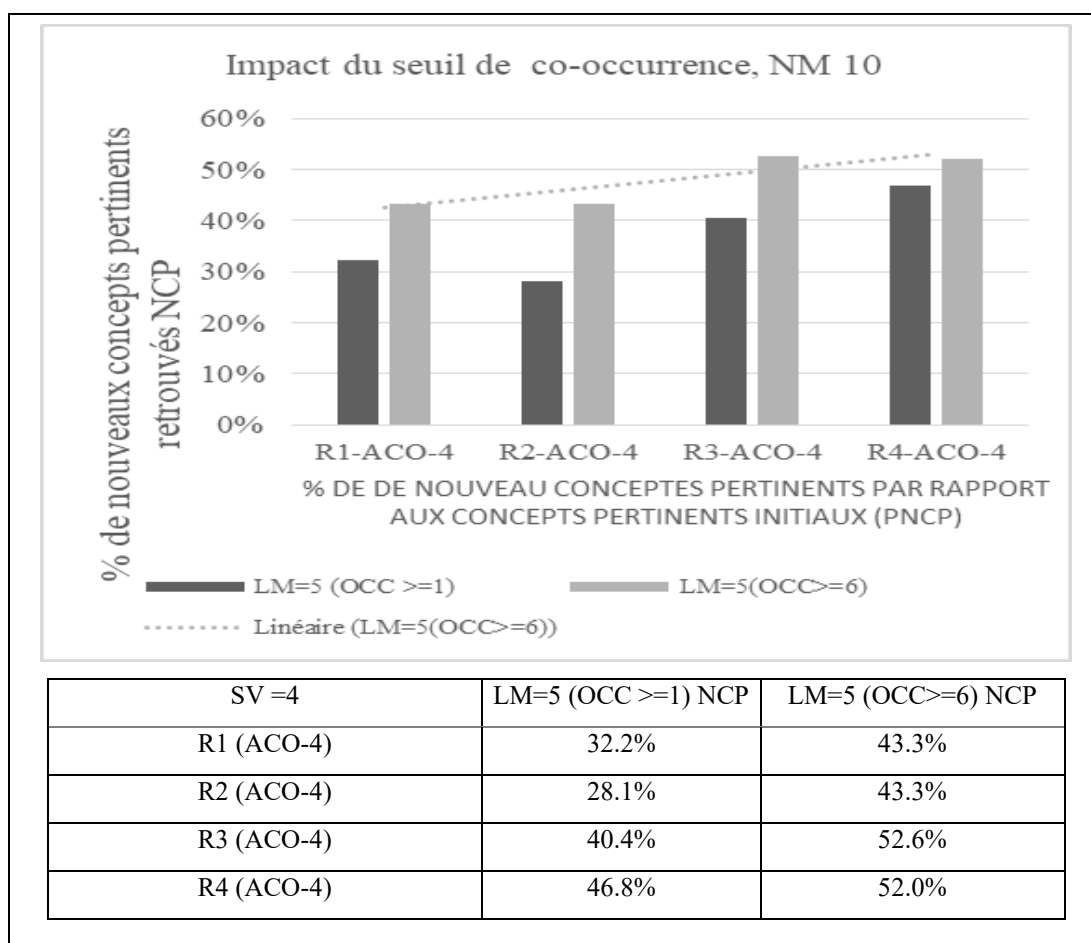


Figure 2.26 Impact du choix du seuil de cooccurrence, NM 10.

- NM =15, LM=5 et SV =4

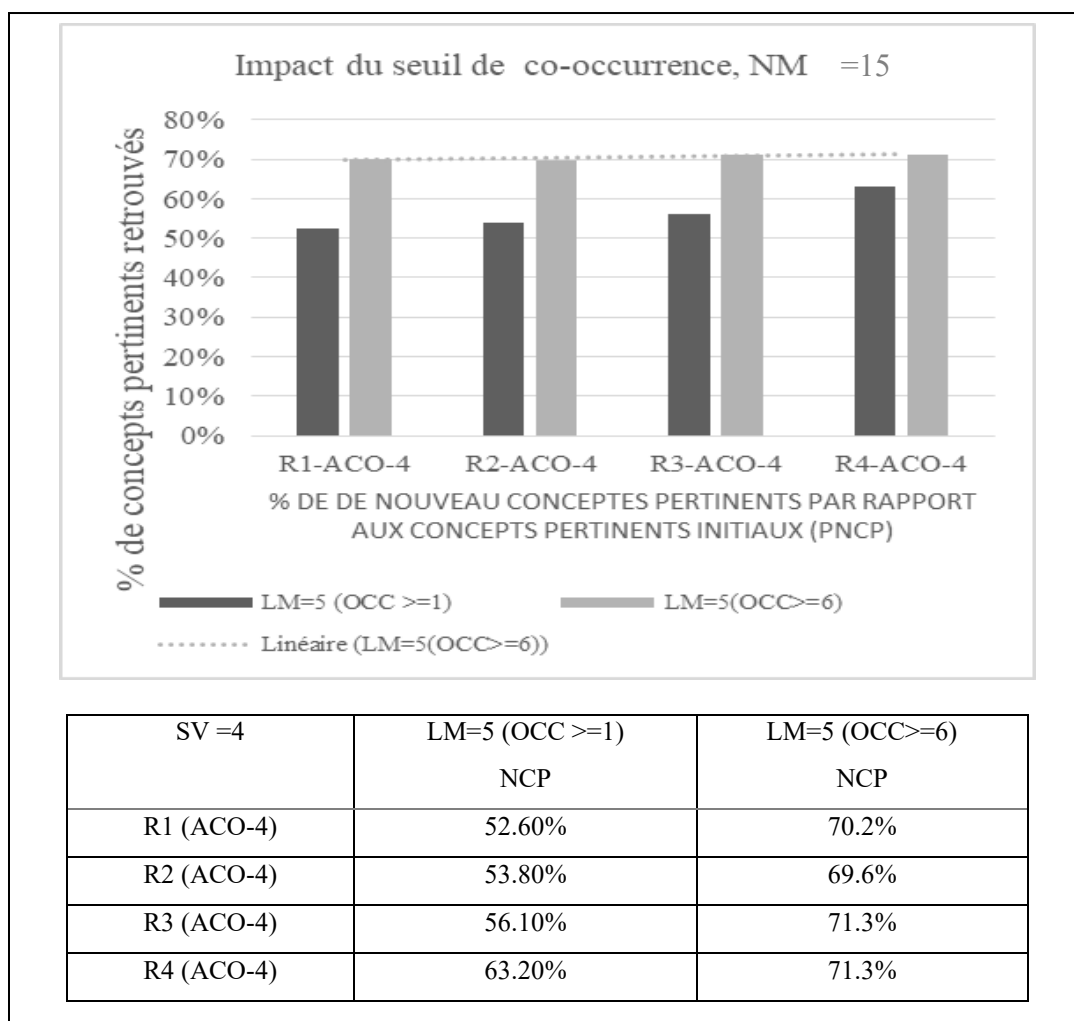


Figure 2.27 Impact du choix du seuil de cooccurrence, NM 15.

En effet, l'expérimentation montre que l'augmentation du seuil de cooccurrence (avec le concept de départ) permet une augmentation des concepts pertinents retrouvés (Figure 2.26 et figure 2.27), mais elle réduit aussi la distinction entre les différents groupes de relations utilisées lorsque le nombre de marches augmente (de 10 à 15 dans nos deux figures) aussi. Ceci pourrait être expliqué par le fait que la liste des concepts à visiter

est de plus en plus réduite, car seuls les concepts avec un nombre de cooccurrences supérieur à 06 (comme dans nos exemples) seraient éligibles au choix aléatoire. Ceci n'est pas un désavantage, car le processus stochastique de la marche aléatoire devient déterministe au fur et à mesure que le nombre d'itérations augmente. De plus, nous avons constaté que le nombre de nouveaux concepts pertinents a atteint 70 % (en excluant ceux considérés comme pertinents après le mappage MetaMapLite).

#### 2.10.7 Impact du nombre de marches et du seuil de visite : (2, 4, 6,8)

Pour avoir une idée sur l'impact du seuil de visite à partir duquel un concept est retenu ou pas. Nous avons mené un test avec les paramètres suivants :

- Groupe de relations RO (car nous avons constaté que c'est un groupe propice à l'exploration) ; une cooccurrence  $\geq 6$  ; longueur de la marche 5 ; sources vocabulaire comme les tests précédents (MSH, MTH et US\_SNOMEDCT).

a) Impact du nombre de marches sur le pourcentage des nouveaux concepts retrouvés

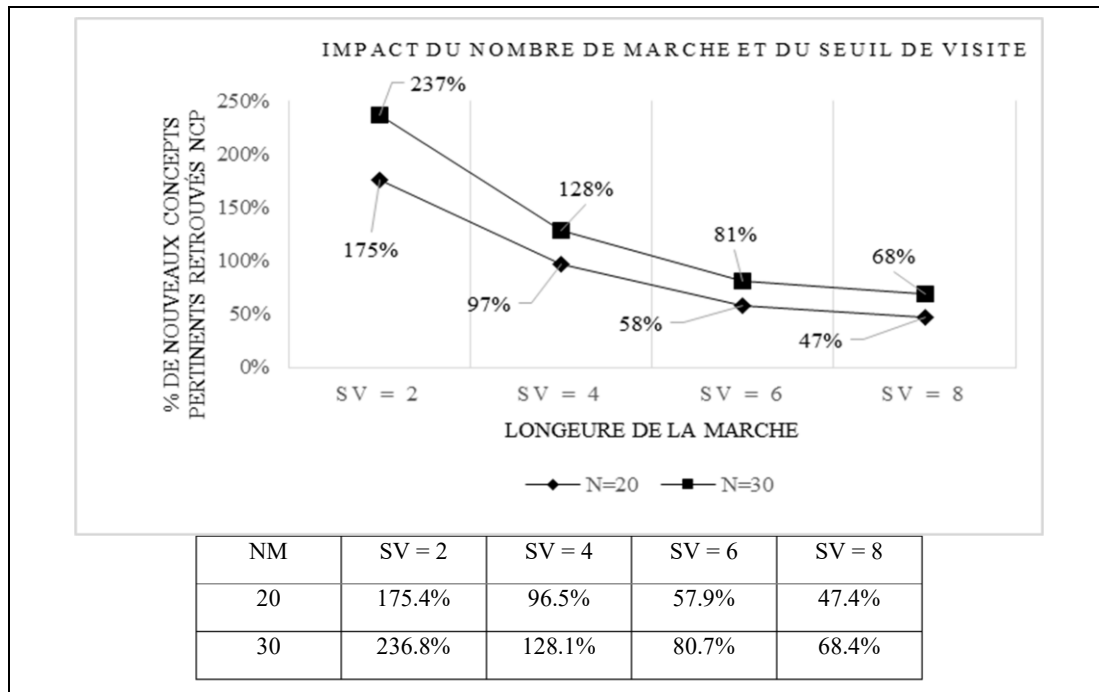


Figure 2.28 Impact sur le pourcentage des nouveaux concepts retrouvés (NCP)

Du graphique de la figure 2.28, nous pouvons déduire deux informations importantes :

1. Plus le nombre de marches est élevé, plus la découverte de concepts pertinents est importante. Pour une dizaine d'itérations supplémentaires, le nombre de concepts pertinents peut doubler.
2. Plus le seuil du nombre de visites est bas, plus le nombre de nouveaux concepts pertinents est important. Ceci s'explique par le nombre total de concepts retrouvés qui augmente la chance de trouver des concepts plus en lien avec le domaine.



### b) Impact du nombre de marches sur le « surplus » de concepts

Pour cette analyse, nous avons cherché à connaître quel serait le « surplus » de concepts que la marche aléatoire avec tous les paramètres que nous avons fixés peut générer. Le « surplus » dans notre contexte réfère au nombre de concepts rapportés par l'algorithme, mais qui ne figurent pas directement dans le dictionnaire de référence utilisé. Ceci n'implique pas que ces concepts sont inutiles, mais plutôt des concepts probablement éloignés du domaine spécifique de la tâche (ou activité) objet de nos travaux.

Le constat est qu'un seuil de visite plus bas permet de retrouver plus de concepts pertinents. En même temps suivant les résultats ci-dessous, on se retrouve avec plus de concepts en surplus qui peuvent aussi diluer l'efficacité des extractions.

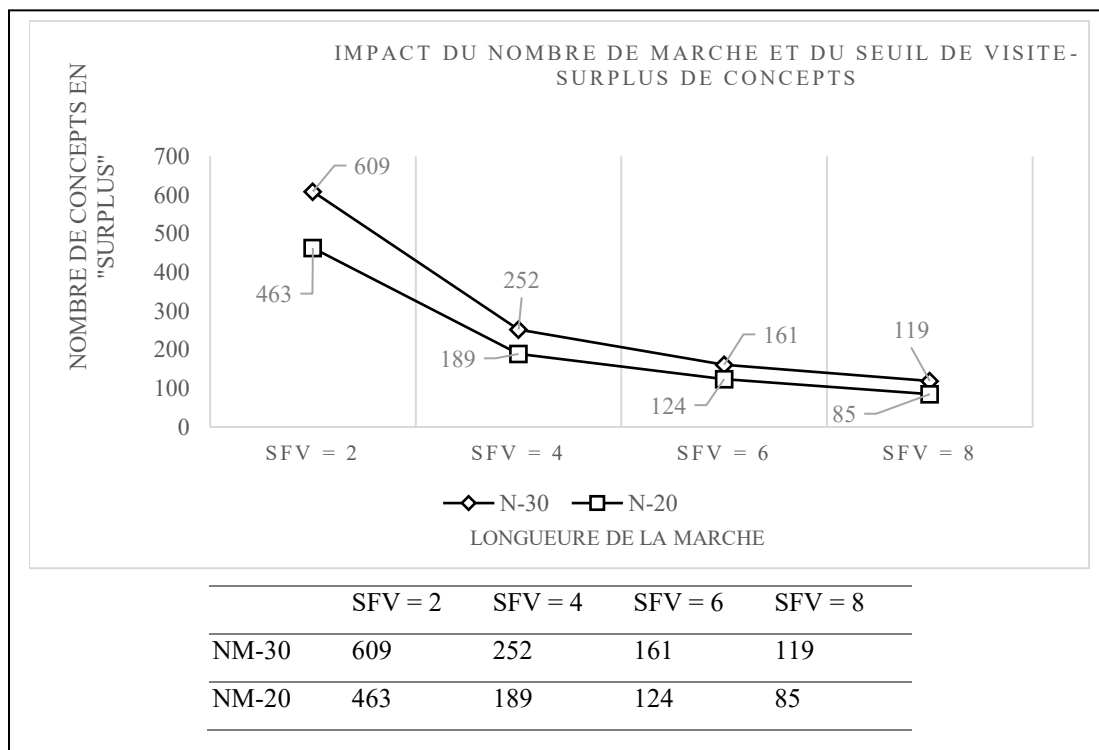


Figure 2.29 Impact du nombre de marches sur le « surplus » de concepts

On pourrait déduire qu'il faudrait probablement choisir un seuil de visite qui réduira le nombre de concepts en surplus même si le nombre de concepts pertinents est moins

important. Le but étant de garder un ratio équilibré entre le nombre de concepts pertinents et le surplus. Sinon, la pertinence des concepts accumulée pour le SADC sera diluée.

## 2.11 Discussion

Dans cette étude, le domaine d'activité médicale est représenté par les règles de décisions d'un SADC. Ces règles reflètent en quelque sorte les tâches majeures d'une activité particulière (des conditions à vérifier et des décisions à prendre). Les différentes expérimentations que nous avons menées, indépendamment du choix arbitraire des paramètres de départ, permettent d'apprécier l'apport d'un algorithme de marche aléatoire « modifié » pour l'identification de concepts pertinents/importants en parcourant les structures termino-ontologique comme celle d'UMLS. À partir des résultats de précisions que nous avons obtenus, il apparaît évident que l'utilisation de la cooccurrence pour limiter la liste des prochains nœuds (concepts) à visiter augmente l'efficacité de l'algorithme de 56 % en moyenne par rapport à une marche totalement aléatoire. Plus le seuil de la fréquence de cooccurrence entre le concept de départ et les futurs concepts à visiter est élevé (supérieur à 6 et plus, dans notre contexte) et plus le pourcentage des nouveaux concepts importants/pertinents au domaine est important (+ 70 % en moyenne pour un NM=15, voir la figure 2.27).

Nous remarquons aussi qu'un seuil de cooccurrence plus élevé permet une plus grande stabilité (moins de variabilité) de concepts entre les différentes exécutions de l'algorithme (ceci s'explique par la réduction du nombre de concepts « à visiter » à partir de chaque nœud ou sommet).

L'utilisation des termes qui existent dans les règles de décision s'est avérée un bon point de départ pour naviguer dans UMLS sans aucune connaissance préalable du domaine d'activité (en l'occurrence, la transfusion). Toutefois, cette exploration ne peut

pas couvrir l'ensemble des termes pertinents énumérés dans le dictionnaire de référence en utilisant seulement ces règles de décision comme point de départ pour les raisons suivantes :

- Les règles de décisions contiennent des abréviations que nous n'avons pas traitées et que le résultat dépend entièrement de l'efficacité de MetaMapLite à les identifier (ce qui n'est pas un point fort de MetaMapLite).
- Les règles ne couvrent pas la totalité des tâches d'un domaine donné, néanmoins elles couvrent une certaine portion. Plus d'expérimentation avec une longueur de marche plus importante, ainsi que d'autres paramètres pourraient nous aider à explorer au mieux UMLS afin d'atteindre cet objectif. Le résultat serait intéressant pour les travaux qui visent à utiliser notre approche pour extraire un sous-ensemble ontologique (pour bâtir une ontologie par exemple) sans un grand effort au départ (pour faciliter considérablement le travail des ontologistes).

Un autre aspect des résultats de ces expérimentations est le fait que certaines relations (ou plus exactement une combinaison de relation) peuvent avoir plus d'impact que d'autres (Ex. RO, RN, RB). Nous estimons que ces relations concentrent des connaissances plus que d'autres. Ces connaissances sont l'apport des efforts d'alignement et du traitement effectué par les spécialistes en terminologies, les ontologistes et les spécialistes du domaine médical pour les différentes sources vocabulaire et d'UMLS. De plus, le réseau sémantique ajoute une nouvelle couche de connaissances et des relations qui n'existaient presque pas comme telle dans les sources vocabulaire d'origines.

En parlant des relations sémantiques, il serait intéressant de mentionner que le paramétrage de notre algorithme ne faisait aucune restriction sur le choix des relations sémantiques lors de l'exploration. Cela dit, il se pourrait qu'un choix limitant les relations sémantiques à explorer (par exemple, favoriser certaines relations sémantiques ex :

« *Sign or symptom* », pourrait servir d'indice pour trouver les réponses à des questions telles que « quels sont les signes d'une hémorragie ». Il en va de même pour la plupart des types de relation sémantique (*Substance, Clinical Drug, Enzyme, Pathological Function, Gene or Genome, etc.*) qui peuvent aider à cerner une pathologie ou une procédure particulière au plus près. Dans le même sens, donner des poids (ou des probabilités) pour discriminer ces relations sémantiques lors de la navigation pourrait changer profondément le comportement de l'algorithme d'exploration ainsi que les résultats de ce dernier. Toutefois, au vu de l'objectif de notre travail de recherche, il s'avère difficile de paramétrer l'algorithme dans cette optique, car le choix de ces relations ou de leur attribuer des valeurs de pondération nécessite des bases de connaissances médicales très avancées dans le domaine.

Ce que les résultats obtenus dans ce chapitre nous apprennent, ce sont les différents paramètres qu'il faut optimiser pour que cet algorithme puisse atteindre l'objectif du chapitre III de cette thèse. Ce chapitre de recherche consiste à améliorer la recherche d'information par l'extension ontologique des requêtes en utilisant les nouveaux concepts résultant de cette exploration. Choisir le nombre de marches, la longueur de cette marche ainsi que les différents seuils pour avoir plus de concepts pertinents/importants appartenant au domaine d'activité et moins de concepts superflus.

Les concepts résultants de l'exploration de l'algorithme peuvent offrir de nouvelles connaissances aux systèmes d'aide à la décision pour améliorer la recherche d'information scientifique au service de la MBE (objet de nos travaux). Cette recherche d'information optimisée peut aussi améliorer le raisonnement des systèmes d'aide à la décision si un mécanisme d'identification de nouvelles règles de décision est envisagé à partir de cette documentation. Par ailleurs, ces ensembles de concepts résultants de l'exploration peuvent être utilisés par les ontologistes comme des ensembles de départ pour entamer l'ingénierie d'une ontologie de domaine ou de celle d'une tâche particulière.

## 1. Complément d'information sur les cas des patients.

Dans le cadre du projet de recherche sur les systèmes d'aide à la décision en transfusion, l'utilité de notre outil se traduit par la possibilité de collecter des informations supplémentaires afin de pouvoir garder un profil des « cas » qui se présentent au système. Ceci permettra de bâtir une base de « cas » qui comporterait plus d'information que celle qui a servi à déclencher les règles de décision du SADC.

Exemple : le terme « transfusion » peut ramener des concepts comme « le test de l'histocompatibilité » ou « transplantation ». Ceci permettra au système d'obtenir d'autres informations sur le patient à partir des systèmes informatiques administratifs ou de laboratoires qui traitent de ces actes médicaux. Ces informations n'étant pas nécessairement liées à la prise de décision, car ils n'existaient pas au départ dans les clauses des règles, mais ils peuvent rajouter une nouvelle couche d'information sur les cas traités. Ce nouvel instantané des « cas » servira comme une optimisation de l'aide à la décision en rajoutant une mémoire à base de cas.

## 2. Construction d'ontologie

L'autre exemple d'utilisation des résultats de cet algorithme est celui de l'aide à la construction d'ontologies spécifique des domaines particuliers en médecine. En effet, au début de nos travaux, nous n'avons pas trouvé une ontologie propre à la transfusion et encore moins, un outil capable d'extraire une partie ontologique d'UMLS qui concerne une activité spécifique. Alors nous avons entamé ce travail pour répondre à notre besoin et celui des chercheurs qui ont les mêmes attentes. Cet outil réduira le temps et l'effort des ontologistes, car, construire une ontologie reste une tâche très coûteuse et la réduction de ce coût par l'utilisation de cet algorithme peut faciliter cette tâche.

### 3. Recherche des publications scientifiques de qualité

Pour revenir à l'objet de notre étude, l'utilisation de cet algorithme a pour but d'utiliser des concepts ontologiques supplémentaires afin d'améliorer la recherche d'informations par l'extension ontologique de la requête de recherche. De ce fait et pour chaque règle de décision exécutée, le système sera capable d'aller chercher les plus pertinentes publications scientifiques en lien avec le cas qui a déclenché cette ou ces règles. Ceci peut se faire juste en utilisant les termes initialement présents dans les clauses conditionnelles ou décisionnelles de ces règles.

#### 2.11.1 Aspects cognitifs

Dans nos travaux de recherche et pour ce premier chapitre, l'essentiel de l'effort a été porté sur l'aspect informatique de l'identification des concepts potentiellement importants, car nous avons dû faire face à beaucoup de défis d'ordre algorithmique et techniques. Cependant, un fort aspect cognitif se trouve à la base du développement de l'ICPU avec comme perspective des recherches plus approfondies en science et processus cognitifs. Avant de décrire ces aspects cognitifs, il serait important de souligner qu'à ce stade de nos travaux, notre objectif ne tend pas à expliquer le mécanisme cognitif sous-jacent à l'identification des concepts pertinents/importants dans les dédales des connaissances structurées des termino-ontologies, que les concepteurs ont consciemment ou inconsciemment incrustées. Nous cherchons plutôt à commencer la construction d'outils qui permettront cette réflexion sur de possibles schémas en lien avec ces processus cognitifs qui pourraient se baser peut-être sur la centralité<sup>34</sup> des concepts (ou coefficient de regroupement local - local clustering coefficient<sup>35</sup>) ainsi que l'exploitation des différentes interconnexions entre thésaurus que ce soit au niveau hiérarchique et sémantique des termino-ontologies dans un métathésaurus comme celui d'UMLS.

Comme déjà mentionné, au centre de notre approche se retrouvent deux composantes défendues par les approches émanant de la science des réseaux pour étudier les processus cognitifs complexes, 1) la construction cognitive en tant que réseau et 2) les processus qui exploitent ce réseau cognitif.

Un réseau est une représentation abstraite des interactions dans un système (Newman 2018). Selon Siew *et al.* (2019), la science des réseaux donne la capacité de quantifier les aspects liés à la structure des systèmes de la cognition et de pouvoir modéliser les processus qui opèrent au sein de ces structures. Aussi, la science des réseaux facilite une meilleure compréhension de la connaissance humaine en permettant aux chercheurs de réfléchir à la structure en réseau de la cognition et aux processus opérant sur ces réseaux qui peuvent interagir pour produire des phénomènes comportementaux. Des exemples concrets ont été modélisés pour la mémoire sémantique — notre mémoire des faits ou des événements (Beckage et Colunga, 2016) et pour le lexique mental — la partie du mémoire à long terme où sont stockées les représentations lexicales (par exemple les travaux de (Aitchison, 2012; Wulff *et al.*, 2019)).

#### a) La modélisation cognitive par réseau

Les réseaux sont composés de deux éléments : des nœuds (sommets) qui représentent les entités conceptuelles ou théoriques d'intérêt et les arcs représentent les relations entre ces unités. Le premier défi pour étudier les systèmes cognitifs en tant que réseaux, c'est de pouvoir les représenter d'une manière significative en termes de nœuds et d'arcs. Aussi, les relations entre les nœuds (c'est-à-dire les arcs) peuvent être aussi importantes que les nœuds eux-mêmes, sinon plus importantes. Cette représentation est assez intuitive pour les cerveaux où les nœuds peuvent correspondre à des neurones, des circuits, ou des régions anatomiques micro-, méso- ou macroscopiques (par exemple, dans le dernier cas, le « *Large-Scale Brain Networks* ») et les arcs peuvent représenter des connexions structurelles ou fonctionnelles (Park, H.-J. et Friston, 2013) entre ces structures. Ceci explique le développement relativement avancé des réseaux

cérébraux (ou « connectomes »). Donc, un premier défi pour étudier les systèmes cognitifs en tant que réseaux, c'est de représenter ces systèmes de manière significative en termes de nœuds et d'arcs. Voici quelques exemples (Tableau 2.17) de constructions cognitives en réseau et leurs applications cognitives.

Tableau 2.17 Exemples de réseaux cognitifs et leurs applications cognitives (Siew et al., 2019).

Réseau cognitif	Nœuds	Arcs	Domaines de recherche pertinents
Réseau sémantique	Mots	Relations sémantiques, taxonomiques, cooccurrence, rôles sémantiques	Acquisition du langage ; vieillissement cognitif ; créativité/perspécité ; recherche cognitive et navigation ; sémantique ; Mémoire
Réseau conceptuel	Concepts, idée	Cooccurrence ; causalité ; fonctionnalité et similarité	Apprentissage ; mémoire ; formation de concept
Clinique, réseaux de personnalité	Symptômes ; traits de personnalité ; éléments sur un questionnaire	Relation statistique telle que corrélations partielles ; comorbidité	Psychopathologie clinique ; troubles de la personnalité
Réseaux sociaux	Utilisateurs	Relation amicale ; abonnés sur les réseaux sociaux médias ;	Résolution collective de problèmes ; la prise de décision ; polarisation

Dans nos travaux, l'outil ICPU est bâti sur une modélisation des termino-ontologies d'UMLS sous forme de graphe avec des nœuds qui représentent les concepts médicaux et les arcs qui représentent différentes relations associatives ou hiérarchiques qui relient ces concepts. Ces relations représentent un travail cumulatif de plusieurs années et des couches successives de connaissances provenant des thésaurus et des codeurs d'UMLS. Cependant, ce n'est qu'aujourd'hui que nous commençons à posséder les méthodes



analytiques nous permettant de comprendre la structure de ces réseaux ainsi que leur dynamique, si nous disposons d'informations concernant le déroulement de leur activité dans le temps ou les étapes de leur développement. Ces connaissances pourraient être mises à profit pour améliorer l'outil ICPU.

#### b) Processus sur les réseaux cognitifs

Une des théories mises en avant pour l'étude des processus cognitifs sur des réseaux est celle de la théorie de la propagation de l'activation (*Spreading activation*) (Collins et Loftus, 1975). Cette théorie a permis de généraliser l'idée que l'utilisation d'une structure en réseau en tandem avec un processus de récupération peut rendre compte du comportement humain. La principale contribution de cette théorie a été de rendre explicite le processus opérant sur un réseau sémantique lors de l'exécution d'une tâche cognitive et de pouvoir relier les résultats de ce processus aux performances des participants et leur temps de réponse.

En d'autres termes, le processus de la propagation de l'activation suppose implicitement la présence de certaines ressources cognitives (c.-à-d. activation) qui peuvent être attribuées à des nœuds spécifiques, et qui peut se propager vers des nœuds connectés dans un réseau prédéfini (Anderson, 1983; Dell, 1986). Cette propagation décline avec le temps et la distance dans un réseau sémantique (Balota et Lorch, 1986). Globalement, le succès de la propagation de l'activation démontre l'importance de formaliser un processus qui capture la recherche dans une représentation cognitive. Une donnée contemporaine dont ne disposaient pas les défenseurs de cette théorie dans les années 1980 est que cette propagation est grandement influencée par la structure du réseau. Les réseaux cérébraux ont tendance à s'organiser dans des structures où des régions modularisées (c'est-à-dire densément interconnectées) sont reliées par des projections de longue distance à d'autres régions modularisées par le biais de « hubs », voire de hubs dits « *rich club* » (un ensemble de hubs densément connectés entre eux)

(Griffa et Van den Heuvel, 2018). Les neurosciences cognitives par réseaux (Medaglia *et al.*, 2015) étudient présentement les propriétés des systèmes cognitifs émergeant de telles structures et ces données et théories pourront éventuellement contribuer au développement de notre outil.

L'une des implémentations de la théorie de la propagation de l'activation sur des réseaux est celle des modèles basés sur la notion et la puissance de la marche aléatoire (Siew *et al.*, 2019). Comme nous l'avons décrit dans la section 2.8.2, un modèle de marche aléatoire naïf est un processus de recherche qui se déplace de nœud en nœud en fonction d'un ensemble de probabilités de transition spécifiant la probabilité de passage aux nœuds voisins directement connectés. Le modèle de marche aléatoire est similaire à un modèle d'activation de la propagation, mais il diverge sur les objectifs de la propagation d'activation suivant la définition de (Collins et Loftus, 1975). Plus précisément, les modèles de marche aléatoire produisent des chemins individuels pris par le marcheur, alors qu'un processus de propagation de l'activation produit un schéma des niveaux d'activation entre les nœuds du réseau ainsi que les changements de ces schémas à travers le temps. En d'autres termes, la propagation l'activation comme schéma représente le comportement global à long terme à partir d'un processus de base sous-jacent qui pourrait être modélisé comme une marche aléatoire.

Dans nos travaux, l'exploration des concepts pertinents par l'outil ICPU utilise un modèle de marche aléatoire avec un comportement qui favorise la cooccurrence entre concepts comme biais de sélection pour le passage d'un nœud à un autre. De plus, un certain nombre de paramètres peuvent être modulés pour modifier le comportement du processus. Formellement, nous n'avons pas utilisé pour le moment cet outil pour analyser un comportement cognitif particulier. Toutefois, ce biais de cooccurrence dans la marche aléatoire pourrait être une base pour comprendre nos mécanismes cognitifs qui permettent de prioriser certains concepts plus que d'autres suivants leur centralité dans

nos mémoires (ici la mémoire ce sont les réseaux hiérarchique et sémantique des termino-ontologies d'UMLS) et la cooccurrence des termes dans l'historique des publications médicales spécialisées.

Plusieurs pistes de développement pouvant contribuer au thème de ce chapitre peuvent être envisagées en se basant sur nos connaissances des sciences cognitives. Une première piste pourrait être de structurer les concepts de manière à recréer la structure en « hubs » de l'organisation cérébrale. Ceci permettrait d'ajouter à l'exploitation de la structure conceptuelle locale associée à un concept, une exploration efficace d'autres champs sémantiques plus distants, mais sémantiquement liés, de manière à permettre des propositions d'articles moins ciblées, mais permettant peut-être des découvertes inattendues. Une autre piste qui pourrait contribuer directement aux objectifs de ce chapitre serait d'essayer de lancer la marche aléatoire sous la forme d'une « propagation d'activation » avec une certaine profondeur. Ceci reviendrait à faire la propagation au début à partir des concepts initiaux, lorsque la première cohorte de nouveaux concepts pertinents est sélectionnée une autre propagation est effectuée à partir de ce nouvel ensemble et ainsi de suite. La profondeur représente le nombre des différentes couches des nouveaux concepts pertinents successives. Dans ce cas de figure, la cooccurrence entre concepts n'est plus liée uniquement aux termes de départ, mais elle se déplacera elle aussi à chaque couche. L'utilisation des relations associatives pourrait permettre d'étendre efficacement cette propagation. L'objectif est de faire avancer l'exploration et la découverte d'autres zones termino-ontologiques plus éloignées.

Les avenues ouvertes par le lien avec les sciences cognitives peuvent être multiples et nous nous limiterons donc à une dernière piste intéressante. Celle-ci pourrait utiliser une évaluation par des médecins des concepts pertinents retrouvés par ICPU. Le résultat sera utilisé comme un indicateur pour changer les poids des connexions entre concepts, auxquels on avait uniquement assigné une valeur égale à 1. Les chemins parcourus lors de recherches amenant des concepts positivement évalués pourraient voir les

poids entre les connexions des concepts contenus dans ce chemin quelque peu augmentés, et inversement pour les chemins menant à des concepts évalués négativement. Ceci implémenterait une forme d'apprentissage qui permettrait, par exemple, d'ajuster le système aux préférences locales des médecins.

## 2.12 Conclusion et travaux futurs.

Nous présentons une nouvelle approche matérialisée par un logiciel qui pourrait être utilisé pour l'identification automatique des concepts pertinents/importants pour un domaine d'activité médicale. Ce processus utilise comme point de départ les termes existants dans les règles de décisions d'un SADC. L'algorithme se base sur une marche aléatoire d'exploration dans UMLS, cette marche est orientée par la cooccurrence qui existe entre les différents concepts dans les publications scientifiques. L'avantage de cette approche c'est l'identification des concepts pertinents pour un domaine d'activité médical sans un effort humain considérable ou une connaissance approfondie du domaine objet de l'exploration. Au fait, cette méthode peut être étendue à n'importe quel type de tâche ou activité médicale. Ceci avec la possibilité de construire à moindres coûts des sous-ensembles de structures ontologiques et des bases de connaissances sans pour autant mobiliser d'importantes ressources humaines.

Les résultats expérimentaux montrent que notre méthode permet d'obtenir une meilleure efficacité pour l'identification des concepts pertinents/importants qu'une marche totalement aléatoire ou uniquement avec MetaMapLite.

Notre algorithme pourrait faire l'objet de plusieurs travaux qui seront conduits suivant les pistes que nous considérons comme très prometteuses, parmi celles-ci :

- 1) Ajouter un autre niveau de profondeur

Les nœuds de niveau 2 et 3 peuvent être considérés comme de nouveaux points de départ et faire l'objet d'une nouvelle marche aléatoire. Ceci conduit à une plus grande profondeur de recherche et d'exploration.

- 2) Donner un poids pour les différentes relations selon une certaine importance.

Il serait probablement pertinent de favoriser certaines relations par l'affectation d'un poids aux arcs du graphe au lieu de les considérer comme identiques (équivalent à 1 dans notre cas)

- 3) Étudier l'impact de restreindre les relations sémantiques

Comme évoqué dans la section discussion, restreindre les types sémantiques pourrait conduire à un comportement plus ciblé de notre algorithme. Néanmoins, cet aspect nécessite l'identification des relations sémantiques sur la base de connaissances médicale profonde.

- 4) Mieux exploiter les abréviations et la réécriture des règles de décision

Nous avons utilisé les règles de décisions dans leur version brute sans aucun traitement préalable. Les abréviations peuvent être un frein pour le nombre et la qualité des termes de départs. MetaMapLite n'est pas un outil très performant lorsqu'il s'agit d'abréviations. Un prétraitement plus adéquat qui prend en charge ces aspects pourrait, à notre avis, augmenter les chances d'une exploration plus performante.

## CHAPITRE III

RECHERCHE DE RÉSUMÉS D'ARTICLES PAR L'EXTENSION DE RE-  
QUÊTE AVEC DES CONCEPTS ISSUE DE LA MARCHE ALÉATOIRE SUR  
UMLS ET UNE ORIENTATION PICO.

### 3.1 Contexte

L'une des difficultés fondamentales de la mise en pratique des résultats de la MBE réside dans la nature de cette dernière qui, par la publication de nouvelles publications scientifiques, crée un vaste bassin de connaissances et d'option de traitements (Frey *et al.*, 2016). Cette multitude d'options peut facilement accabler les cliniciens qui tentent de se tenir au courant des dernières conclusions et peuvent facilement inhiber l'effort de déterminer le meilleur traitement possible pour un patient particulier. Cependant, la capacité de rapidement localiser les preuves scientifiques pertinentes est une habilité de la recherche d'informations (RI) (Roberts *et al.*, 2017). Plusieurs études se sont intéressées à optimiser la recherche d'articles scientifiques (Boudin, Shi, *et al.*, 2010; Demner-Fushman et Lin, 2007; Huang, X. *et al.*, 2006), le but était de retrouver les articles ou les résumés qui adhèrent au mieux au besoin du praticien en termes de recherche d'information.

### 3.2 Motivation

Notre motivation est guidée par le souhait d'améliorer les capacités des SADC pour qu'ils puissent proposer aux utilisateurs des publications scientifiques par une recherche d'information en lien avec les cas qu'ils traitent. Cette amélioration des capacités de recherche pourrait permettre à un SADC d'aller chercher des publications scientifiques récentes à l'extérieur du système avec comme point des concepts résidents dans ces propres règles de décisions par exemple. Ces concepts seront aussi enrichis par d'autres connaissances qui proviennent de source extérieure comme les thésaurus et les ontologies.

Dans ce chapitre, nous présentons une approche qui utilise les concepts initiaux (retrouvés dans les règles de décision ou même d'une autre source comme du texte libre) comme points de départ pour améliorer la recherche d'information, et ce, en récupérant

d'autres concepts sémantiquement liés et potentiellement pertinents. Ces nouveaux concepts deviendront une nouvelle extension pour la requête initiale et reflètent, indirectement, les connaissances médicales enchâssées dans les termino-ontologies comme celles d'UMLS.

L'outil ICPU a été utilisé pour cette extension. Ce dernier, développé et expliqué dans le chapitre II précédent, permet l'identification par une marche aléatoire des concepts pertinents à partir d'UMLS en favorisant les concepts qui cooccurrent le plus souvent dans les publications scientifiques. L'outil permet aussi de donner une importance à chaque concept candidats à l'extension par le biais de leurs fréquences de visite. De plus, nous voulions explorer l'influence d'une orientation PICO (voir la section 3.7.4) dans le choix de ces concepts sur les résultats de la recherche d'information médicale dans les publications scientifiques.

### 3.3 Questions de recherche

Les deux questions de recherche pour ce chapitre sont les suivantes :

*Peut-on améliorer la pertinence des résultats de la recherche d'information avec une extension de requête qui exploite les concepts identifiés dans l'étape précédente et une pondération reflétant l'importance de ces concepts dans leur entourage termino-ontologique ?*

*Mener cette extension de requête en choisissant les nouveaux concepts candidats suivant leur appartenance à l'une ou l'autre des catégories du modèle PICO (Patient ou Population, Intervention, Comparateur, Outcome — voir section 3.7.4) pourrait-il améliorer cette recherche d'information ?*



### 3.4 Contribution

La principale contribution dans ce chapitre est une méthode pour l'extension de requêtes basée sur les résultats d'exploration termino-ontologique qui entourent un ensemble de termes de départ appartenant à ces requêtes initiales. Cette approche innovante utilise une pondération qui reflète la centralité de ces nouveaux concepts dans leur graphe termino-ontologique au sein d'UMLS. Ce graphe est parcouru par le processus de marche aléatoire introduit dans le chapitre précédent. De plus, une catégorisation de ces concepts par rapport au paradigme PICO a été ajoutée. Pour déterminer à quelle catégorie PICO ces nouveaux concepts s'apparentent, nous proposons une autre méthode qui exploite les informations semi-structurées dans un registre d'études cliniques. Ceci a permis de produire des ensembles de références pour cette catégorisation.

### 3.5 Structure du chapitre

Après la présentation du contexte, la motivation et les questions de recherche à la base du développement de ce chapitre, une introduction générale à la recherche d'informations et l'évaluation des systèmes de recherche a été nécessaire pour expliquer des notions importantes pour la compréhension du raisonnement globale, de la méthodologie de recherche et les expérimentations. Parmi celles-ci, les modèles de recherche d'information, l'évaluation de ces systèmes ainsi que les outils et ressources disponibles pour la réalisation et l'évaluation des nouveaux processus de recherche proposés. Après ces introductions, une revue de la littérature est présentée pour énumérer certains travaux effectués dans le domaine en concordance avec les prémisses des hypothèses. Par la suite, nous introduisons notre méthodologie d'extension de requêtes ainsi que les différentes variations que nous avons expérimentées. Sans oublier d'expliquer dans cette section les nombreux prétraitements et les outils utilisés pour préparer les données

avant l'évaluation des approches. Par après, les expérimentations (incluant les paramètres et les mesures utilisés) seront présentées et expliquées afin d'alimenter les sections suivantes notamment les résultats et les discussions. Finalement, le chapitre se terminera par des conclusions et les futures pistes de recherches possibles.

### 3.6 Introduction et notions générales en recherche d'information

#### 3.6.1 Recherche d'information (RI)

Les systèmes de recherche d'informations (RI) ont été d'abord développés pour faciliter la gestion des informations documentaires, pour certains professionnels, dans les bibliothèques et ce, on leur donnant rapidement un accès aux livres et aux revues (Baeza-Yates et Ribeiro-Neto, 1999; Chowdhury, 2010). Actuellement, la recherche d'informations devient la forme dominante de l'accès à l'information; chaque jour, des millions d'utilisateurs l'utilisent que ce soit pour leur recherche sur le web ou le contenu de leurs courriels (Mogotsi, 2010; Schütze *et al.*, 2008). Schütze *et al.* (2008) proposent la définition suivante : « la *recherche d'information (RI)* consiste à trouver du matériel (généralement des documents) d'une nature non structurée (généralement du texte) à partir de grandes collections (généralement stockées sur des ordinateurs) avec comme objectif de satisfaire à un besoin d'information ». De cette définition, trois notions importantes pour la recherche d'information s'érigent : 1) la notion de « document » qui englobe toute unité qui peut constituer une réponse à une requête d'un utilisateur. 2) La notion de « requête » qui exprime le besoin d'information d'un utilisateur et 3) la notion de « pertinence » qui consiste à retourner seulement les documents dont l'utilisateur a besoin.

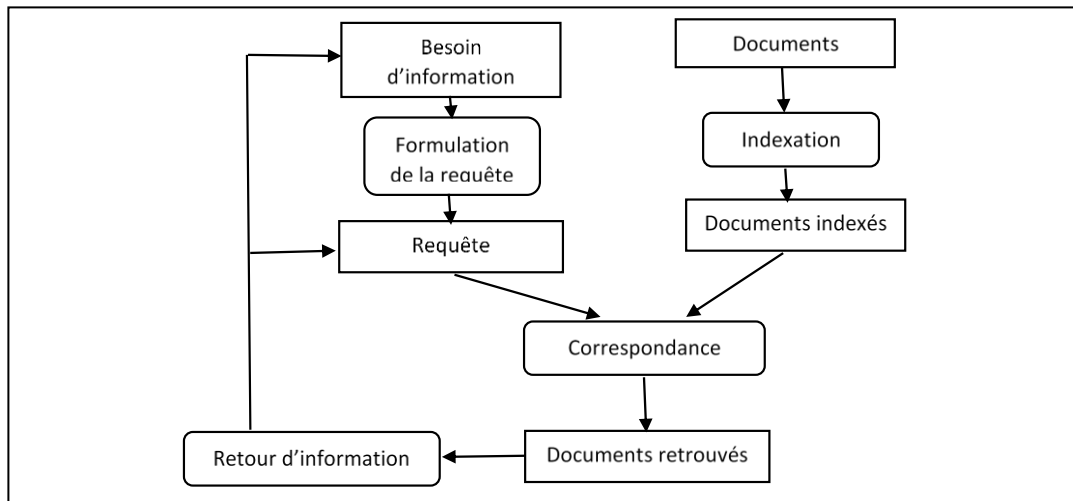


Figure 3.1 Processus de recherche d'informations (Hiemstra, 2009).

La plupart des systèmes de recherche d'information, y compris les moteurs de recherche, s'appuient, en tout ou en partie, sur le calcul de l'importance des termes qui apparaissent dans la requête et dans les documents pour déterminer leurs réponses (Carpineto et Romano, 2012). Pour y parvenir, un processus d'indexation des documents et des requêtes est opéré pour donner une représentation des significations (sémantiques) et la facilitation pour la manipuler computationnelle de la recherche (Figure 3.1). Cette indexation permet de trouver les mots qui représentent le mieux le contenu d'un document. Toutefois, décider de l'importance d'un terme pour résumer le contenu d'un document n'est pas une tâche triviale. Tous les termes indexés ne sont pas égaux pour décrire le contenu du document, certains termes sont plus informatifs et d'autres sont plutôt plus vagues. Pour établir cette pondération, différentes approches, comme celle de la fréquence des termes (*TF*) et la fréquence inverse des documents (*IDF*) ont été proposées. Ces dernières sont devenues les fondements des schémas de pondération les plus populaires en IR, comme le *TF-IDF*.

Pour rappel, la « *TF* » (*term frequency*) représente le nombre de fois (fréquence) où le terme (*t*) apparaît dans un document (*d*) avec  $TF(t, d) = f_{t,d}$ . Cette fréquence peut être ajustée à la longueur du document.

Quant à la IDF (*inverse document frequency*), celle-ci représente la fréquence inverse du document dans une collection et qui cherche à déterminer la quantité d'informations fournies par le mot (est-il commun ou rare dans l'ensemble des documents) (Baeza-Yates et Ribeiro-Neto, 1999). Il s'agit de la fraction inverse à l'échelle logarithmique des documents contenant le mot (obtenue en divisant le nombre total de documents par le nombre de documents contenant le terme, puis en prenant le logarithme de ce quotient).  $IDF = \log \left( \frac{N}{1 + |\{d \in D : t \in d\}|} \right)$  Avec N égale au nombre total des documents dans la collection et la valeur  $|\{d \in D : t \in d\}|$  comme étant le nombre de documents ou le terme (*t*) apparaît (plus cette valeur est proche de 0 et plus le terme est très commun). La multiplication des deux facteurs donne la pondération du terme (*t*). Plus ce facteur est grand et plus le terme est pertinent pour ce document.

En plus de ces mesures qui permettent d'évaluer l'importance des termes dans les documents et les collections, la recherche d'information se base aussi sur des modèles de recherches, qui permettent de *comparer les représentations de documents et la requête afin de parvenir à un score de correspondance*. Parmi les familles de modèles les plus utilisés, on retrouve les modèles statistiques et les modèles probabilistes.

Pour ce qui est des modèles statistiques, ces derniers se basent sur le degré de similitude entre la représentation de la requête et les représentations des documents dans la collection afin de classer les résultats de la recherche (Hiemstra, 2009a, 2009b; Luhn, 1957). Une multitude d'approches ont été proposés pour calculer cette similitude, du simple comptage du nombre d'éléments que la requête et la représentation de l'index du document en commun aux modèles d'espace vectoriel et ces variantes ultérieures (Dillon, 1983). Ces modèles considèrent les représentations d'index du document et de la requête comme des vecteurs intégrés dans un espace euclidien de grande dimension, où chaque terme se voit attribuer une dimension. La mesure de similitude est généralement le cosinus de l'angle qui sépare les deux vecteurs (Hiemstra, 2009b).

Dans les modèles probabilistes, la recherche d'information est considérée comme un processus incertain et imprécis, avec une imprécision dans l'expression des besoins (la requête) et une incertitude dans la représentation des informations (les documents recherchés), ce qui rend la théorie des probabilités adéquate pour prendre en compte cette incertitude et cette imprécision. L'idée de base de ces modèles probabilistes est de pouvoir déterminer les probabilités de la pertinence et de la non-pertinence d'un document par rapport à une requête donnée. En résumé de cette introduction, plusieurs paradigmes sont à la base d'innombrables approches qui visent essentiellement à améliorer la performance des systèmes de recherche d'information, et ce, en se rapprochant des besoins d'information des utilisateurs.

### 3.6.2 Évaluation des systèmes de RI

#### 3.6.2.1 La notion de pertinence

La pertinence est la notion centrale dans la recherche d'information, car toutes les évaluations s'articulent autour de cette notion. La pertinence est considérée comme la correspondance (un degré de relation, mesure d'informativité) entre un document et une requête (Xu et Chen, 2006). Un système de recherche d'information parfait ne récupérerait que les documents pertinents et aucun document non pertinent. Cependant, selon Hiemstra (2009b) les systèmes parfaits n'existent pas et n'existeront pas, car la recherche d'information est nécessairement incomplète et la pertinence dépend de l'avis subjectif de l'utilisateur. Malgré des années de travaux, il n'est pas certain que l'on sache mesurer l'efficacité d'un système de recherche d'informations du point de vue de l'utilisateur. En effet, il a été démontré que les différentes tentatives faites pour apprécier l'expérience de l'utilisateur avec des mesures mathématiques sont un échec (Turpin et Scholer, 2006). L'évaluation des utilisateurs est construite sur plusieurs couches d'interprétations qui sont dérivées à partir des expériences, des perceptions et

des connaissances privées des individus liés aux problèmes d'information qu'ils recherchent (Park, T. K., 1993). Toutefois et comme présenté dans les sections concernant la conférence TREC (section 3.9.4), une approche d'évaluation appelée « *pooling* » a été adoptée et qui permet d'instaurer un consensus pour évaluer la pertinence des documents retrouvés par rapport aux requêtes d'information dans le cadre de cette conférence.

### 3.6.2.2 La conférence « TREC » et les corpus de test pour la RI

L'expérimentation dans le domaine de la recherche d'information a été pour longtemps basée sur de petits corpus de test qui ne reflètent pas les problèmes posés par des environnements bibliographiques plus larges. Pour qu'un corpus de test soit significatif, il faut qu'il possède un nombre de documents assez élevé. Les premiers corpus de test développés dans les années 1960 renferment quelques milliers de documents (Pinel-Sauvagnat et Mothe, 2013). De plus, il était difficile de faire la comparaison de l'efficacité des systèmes de recherche d'information, car les différents groupes de chercheurs focalisés sur différents aspects de la recherche d'information. Pour faire face à ces divergences et donner un espace commun à tous les travaux de recherche, de nombreuses campagnes d'évaluation ont vu le jour, parmi eux on peut citer TREC (*Text REtrieval Conference*”, <http://trec.nist.gov>) ou encore CLEF (*Conference and Labs of the Evaluation Forum*, <http://clef2018.clef-initiative.eu/>) (Baeza-Yates et Ribeiro-Neto, 2011; Pinel-Sauvagnat et Mothe, 2013).

La TREC est une campagne annuelle dédiée à l'évaluation des recherches classiques avec de très larges corpus de test (millions de documents). Ces corpus ont été développés pour différents champs de recherche d'information (web, médecine de précision, aide à la décision...) avec trois principales composantes : 1) un ensemble de documents, 2) un ensemble de requêtes et 3) un ensemble de documents jugés pertinents pour chaque sujet de recherche (Baccini et al., 2010). Chaque année la TREC fournit

un ensemble de documents et des sujets de recherche que les participants doivent exploiter avec leurs propres systèmes de recherche d'information. Les résultats des expérimentations appelés « exécutions » seront soumis sous forme de listes des documents les mieux classés par leurs systèmes de recherche. Après cela, la TREC évalue l'ensemble des résultats et organise des ateliers pour la communauté et les participants (Pinel-Sauvagnat et Mothe, 2013).

### 3.6.2.3 Mesures utilisées pour l'évaluation des systèmes de RI

#### 3.6.2.3.1 Précision et rappel

La qualité d'un système doit être mesurée en comparant les réponses du système avec les réponses idéales que l'utilisateur espère recevoir. Plus les réponses du système correspondent à celles que l'utilisateur recherche, plus le système sera considéré comme étant performant (Saracevic, 1995). La comparaison des réponses d'un système pour une requête avec les réponses idéales peut se faire par les deux unités de mesure très connues dans le domaine de la recherche d'information.

- Précision : la précision mesure la proportion de documents pertinents retrouvés parmi tous les documents retrouvés par le système.

$$\text{Précision} = \frac{\text{\#documents pertinents retrouvés}}{\text{\#documents retrouvés}}$$

- Rappel : le rappel mesure la proportion de documents pertinents retrouvés parmi tous les documents pertinents de la collection.

$$\text{Rappel} = \frac{\text{\#documents pertinents retrouvés}}{\text{\#documents pertinents dans la base}}$$

Les valeurs de ces mesures sont comprises entre 0 et 1 et sont optimales pour 1. Ces deux mesures varient en sens inverse : les méthodes permettant d'augmenter la précision ont tendance à dégrader le rappel et vice versa (Koehrsen, 2018).

### 3.6.2.3.2 P@k

La précision peut être calculée également lorsque l'on considère les premiers documents retrouvés. Ce type de mesure est en particulier utilisé pour évaluer la « haute précision ». Par exemple, la P@10 correspond à la précision obtenue si seuls les 10 premiers documents de la liste des documents retrouvés étaient considérés (Pinel-Sauvagnat et Mothe, 2013). Exemple, supposons qu'il existe 20 documents pertinents dans une collection et que le résultat retourné par une requête (Q) avec la pertinence du document est indiqué dans la dernière colonne, la P@10 est alors égale à 0,70 (Tableau 3.1) :

Tableau 3.1 Exemple de calcul de la P@k pour une requête donnée.

k	P@k	Document retourné par la requête (Q)
1	$(1/1) = 1,0$	P
2	$(1/2) = 0,5$	NP
3	$(2/3) = 0,67$	P
4	$(3/4) = 0,75$	P
5	$(4/5) = 0,80$	P
6	$(5/6) = 0,83$	P
7	$(6/7) = 0,86$	P
8	$(6/8) = 0,75$	NP
9	$(7/9) = 0,78$	P
10	$(7/10) = 0,70$	NP

L'avantage de la mesure P@k est sa facilité de calcul et d'interprétation. Le k est généralement choisi dans l'ensemble suivant P@ [1, 5, 10, 20].



### 3.6.2.3.3 La précision moyenne (AP-*Average Precision*)

Afin de connaître la performance que l'on peut associer à une requête, la communauté scientifique a adopté comme mesure principale la précision moyenne (Abdou et Savoy, 2007). Son calcul s'opère selon le principe suivant : pour chaque requête, on détermine la précision après chaque document pertinent, puis on calcule une moyenne arithmétique sur l'ensemble de ces valeurs. Elle est définie par :

$$AP(Q) = \frac{\sum_{k=1}^n P@k \times rel(k)}{\# \text{ documents pertinents pour la requête } (Q)}$$

Avec  $n$  comme étant le nombre de documents retrouvés ( $k$ ) est le rang du document et  $P@k$  est la précision lorsque les  $k$  premiers documents retrouvés sont considérés.  $rel(k)$  est égale à 1 si le document au rang ( $k$ ) est pertinent et 0 sinon (Pinel-Sauvagnat et Mothe, 2013), un exemple de calcul de la précision moyenne est illustré dans le tableau 3.2.

Tableau 3.2 Précisions moyennes de deux requêtes ayant trois documents pertinents (notés P) et non pertinents (NP) présentés dans des rangs différents ((Abdou et Savoy, 2007).

Rang	Requête A	Requête B
1	NP	P 1/1
2	P 1/2	P 2/2
3	P 2/3	NP
...	NP	NP
35	P 3/35	NP
...	NP	NP
108	NP	P 3/108
AP	0.4175	0.6759

### 3.6.2.3.4 La moyenne de la précision moyenne (*MAP-Mean Average Precision*)

La valeur moyenne de la précision moyenne (MAP) est considérée comme la mesure offrant le meilleur jugement en présence de requêtes multiples (Schütze *et al.*, 2008). Dans nos expériences et pour évaluer les performances d'extraction, les valeurs de précision moyenne (MAP) ont été prises en compte. Cette mesure est fortement recommandée par la communauté TREC parce qu'il a été démontré que le MAP avait une discrimination et une stabilité particulièrement bonnes (Schütze *et al.*, 2008). La MAP peut être calculée en utilisant l'équation suivante avec  $n$  représentant le nombre de requêtes qu'on veut mesurer ( $n$  étant le nombre de requêtes) :

$$\text{MAP} = \frac{1}{n} \sum_n \text{AP}_n$$

### 3.6.2.3.5 R-Précision

R-Précision est défini comme étant le rapport entre le nombre de documents pertinents extraits jusqu'à un rang ( $r$ ) sur le nombre de documents pertinents qui existe dans la collection au total ( $R$ ) (Zhou et Yao, 2010). Ci-dessous la formulation de ce rapport :

$$\text{R - Précision} = \frac{\# \text{ documents pertinents extraits jusqu'au rang } (r)}{\# \text{ documents pertinents de la collection } R}$$

Par exemple : dans une collection de 100 documents au total et dont 30 sont pertinents ( $R=30$ ), les autres étant sans objet. Le système récupère les 30 premiers documents (au même nombre des documents pertinents dans la collection). Sur ces 30 documents, 10 sont pertinents ( $r = 10$ ) alors la R-précision est égal à  $\frac{10}{30} = 0,33$  (Baeza-Yates et Ribeiro-Neto, 1999).

La mesure *R-Précision* met moins l'accent sur le classement exact des documents pertinents récupérés, ce qui peut être particulièrement utile dans le TREC où il existe un grand nombre de documents pertinents. La R-précision et la précision moyenne sont fortement corrélés et ont une stabilité similaire en termes de comparaison des systèmes utilisant différentes requêtes (Aslam, J. A. *et al.*, 2005).

La moyenne de la R-précision est calculée en prenant la moyenne des R-précisions des différentes requêtes. Par exemple : Deux requêtes Q1 et Q2, un avec 50 documents pertinents pour Q1 et 10 documents pertinents Q2. Si le système renvoie 17 documents pertinents dans les 50 premiers documents pour Q1 et 7 documents pertinents dans les 10 premiers Q2, la R-précision pour les deux requêtes est alors égale à :

$$\text{R-précision} = \frac{\left(\frac{17}{50}\right) + \left(\frac{7}{10}\right)}{2} = 0.52$$

#### 3.6.2.3.6 Autres mesures

Il existe d'autres mesures utilisées dans l'évaluation des résultats de recherche d'information comme le gain cumulative (*Cumulative Gain*), gain cumulé actualisé (*Discounted Cumulative Gain-DCG*) ou infNDCG (*inferred Normalized Cumulative Gain (NDCG)*) et bien d'autres. Pour l'exemple et à titre non exhaustif, infNDCG est une mesure très populaire pour l'évaluation des systèmes de recherches, cette mesure se base sur deux hypothèses : 1) les documents très pertinents sont plus utiles que les documents les moins pertinents et 2) cette pertinence est liée à la position du document dans le classement retourné par le système qui dénote son utilité. Plus il est loin dans ce classement, moins il est utile à l'utilisateur, car il est moins susceptible d'être examiné. (Dupret, 2011).

### 3.6.2.4 Plate-forme d'évaluation et de développement de SRI

Plusieurs plates-formes logicielles sont utilisées pour le développement des systèmes de recherche d'information. Parmi les systèmes les plus connus, en retrouve : ATIRE (Trotman *et al.*, 2012), Galago (Cartright *et al.*, 2012), Indri (Metzler et Croft, 2004), JASS (Lin et Trotman, 2015), Lucerne (Białecki *et al.*, 2012), MG4J (Boldi et Vigna, 2005) et Terrier (Ounis *et al.*, 2006). L'ensemble de ces systèmes implémentent les deux aspects de la recherche d'information (l'indexation et la recherche). Toutefois chaque système se distingue par les techniques d'implémentation. Dans une étude comparative sur l'efficacité des systèmes, Lin *et al.* (2016) ont établi un classement suivant

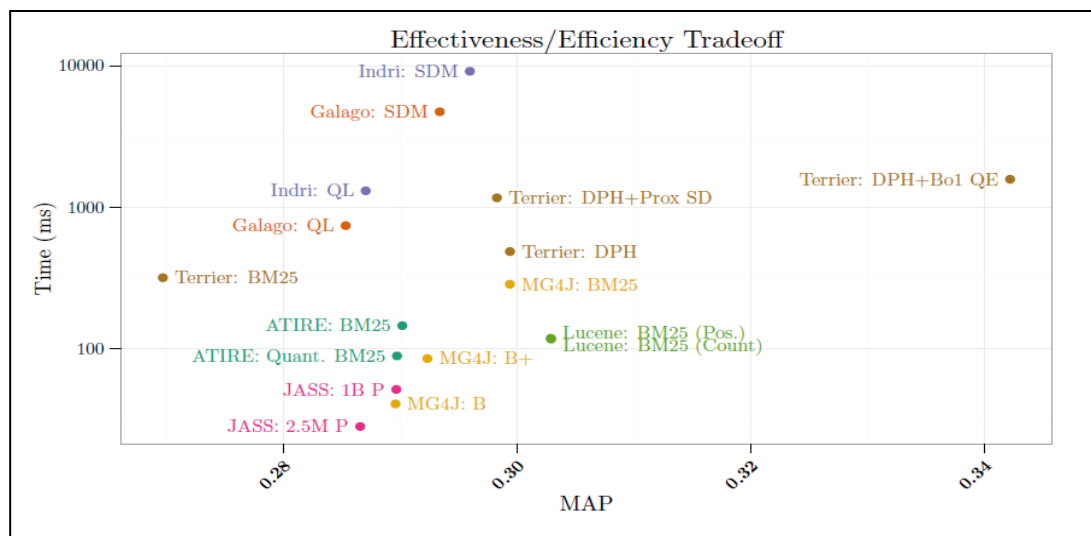


Figure 3.2 Compromis entre efficacité et efficacité (Lin et al. (2016))

l'efficacité (en termes de MAP) et efficacité en ce qui concerne le temps d'exécution pour 150 requêtes de base (TREC) sur des serveurs EC2 identiques. Il ressort de cette étude que la plate-forme terrier se distingue avec son modèle de pondération DPH (voir la section 3.6.2.5.1) et son extension Bo1 (la section 3.9.8) comme étant le système qui conjugue le mieux l'efficacité et l'efficacité dans le processus d'indexation et de recherche (Figure 3.2).

### 3.6.2.5 Plate-forme d'évaluation « Terrier »

Terrier pour « *Terabyte Retriever* », est un projet qui a été créé à l'Université de Glasgow en 2000 dans le but de fournir une plate-forme souple pour le développement rapide des applications de recherche d'information à grande échelle, ainsi qu'un banc d'essai de pointe pour la recherche et l'expérimentation dans ce domaine (Ounis *et al.*, 2006). Cette plate-forme implémente les différentes étapes de la recherche d'information classique (Figure 3.1) et offre, en plus, un cadre pour l'évaluation des résultats de recherche pour différentes applications (Ounis *et al.*, 2006). Terrier a été largement testé et éprouvé (Middleton et Baeza-Yates, 2007). Le choix de cette plate-forme pour notre évaluation est dû essentiellement à sa capacité à traiter de grandes collections de documents et aux résultats d'efficacité et d'efficience évoquées dans la section précédente. Voir (Annexe B) pour plus de détails sur l'architecture et l'utilisation de cette plate-forme.

#### 3.6.2.5.1 Modèles de pondération

Terrier est considéré comme une plate-forme mature et à la pointe de la technologie (Macdonald *et al.*, 2005), utilisant des modèles de recherche d'information et de pondération très efficaces basés sur le modèle DFR « *Divergence from Randomness* ». La performance de ces modèles est au moins comparable, sinon mieux, à ceux des plus récents modèles d'indexation et de recherche (Ounis *et al.*, 2006), y compris sur des collections de tests à très grande échelle. Le modèle de DFR, proposé par Amati, Gianni et Van Rijsbergen (2002), calcule la pondération des termes en mesurant la divergence entre une distribution des termes produits par un processus aléatoire (dans l'ensemble de la collection de documents) et la distribution réelle des termes (dans le document) (Wilkie et Azzopardi, 2014). Cela est basé sur deux hypothèses : 1) tous les mots n'ont pas la même importance pour décrire le contenu des documents et que 2) les mots qui contiennent peu d'informations sont supposés être distribués d'une façon aléatoire sur

l'ensemble des documents de la collection comme le décrit par Baeza-Yates et Ribeiro-Neto (1999).

La formulation ci-dessous définit le poids du terme comme étant inversement proportionnel à la probabilité de la fréquence du terme au sein du document (d) obtenu par un modèle M d'aléatoire :

$$\text{Poids}(t|d) \propto -\log \text{Prob}_M(t \in d | \text{Collection})$$

Le tableau 3.3 décrit quelques modèles et techniques de pondération et d'extraction de « terrier » utilisées pour nos expérimentations. En particulier, l'Okapi BM25 et trois modèles du cadre appelé DFR « *Divergence From Randomness* » les plus performants : le modèle de Poisson-Laplace avec leur deuxième normalisation de la fréquence des termes (PL2), le modèle (DPH) hypergéométrique utilisant la normalisation de Popper et le modèle InL2 (Wilkie et Azzopardi, 2014).

Terrier permet aussi, la mise en œuvre d'autres modèles de pondération issus du cadre Divergence from Randomness (DFR) parmi lesquels : BB2 (DFR); DFR\_BM25 (DFR); DLH (DFR); DLH13 (DFR) (voir le lien suivant pour plus d'informations [http://terrier.org/docs/v3.5/dfr\\_description.html](http://terrier.org/docs/v3.5/dfr_description.html)).

Tableau 3.3 Description des modèles de pondérations utilisés.

Modèle	Description	Formulation
Okapi BM25	Okapi BM25 est une application du modèle probabiliste de pertinence, proposé en 1976 par Robertson et Jones. Ce modèle ordonne les documents en fonction de la fréquence des termes qui apparaissent dans chaque document, indépendamment des relations pouvant exister entre ces termes ou de leurs proximités relatives au sein du document.	$score(d, Q) = \sum_{i=1}^n IDF(q_i) \frac{tf(q_i, d) \cdot (k_1 + 1)}{tf(q_i, d) + k_1 \cdot (1 - b + b \cdot (\frac{ d }{avgdl}))}$ <p>Où <math>tf(q_i, d)</math> est la fréquence du terme <math>q_i</math> dans le document (<math>d</math>). <math> d </math> est la longueur du document (<math>d</math>) en mots et <math>avgdl</math> est la longueur moyenne des documents dans la collection de textes à partir de laquelle les documents sont extraits. <math>k_1</math> et <math>b</math> sont des paramètres libres, généralement choisis, en l'absence d'optimisation avancée, comme <math>k_1 \in [1.2, 2.0]</math> et <math>b = 0,75</math>.</p> <p><math>IDF(q_i) = \log \frac{N - n(q_i) + 0.5}{n(q_i) + 0.5}</math> Où <math>N</math> est le nombre total de documents de la collection et <math>n(q_i)</math> est le nombre de documents contenant (<math>q_i</math>).</p>
DPH (DFR)	DPH pour ( <i>Different hyper-geometric DFR model using Popper's normalization</i> ) Amati, Giambattista, 2006) avec l'utilisation de la fonction de distribution hyper-géométrique réduite à une distribution binomiale avec des termes non uniformes ou la fonction de pondération est donnée par (Lu et al., 2013)	$score(d, Q) = \sum_{t \in Q} qtw(t) \cdot \frac{tf \cdot \log_2 \left( \frac{tf}{l(d)} \cdot \frac{TFC}{TF} \right) + 0.5 \cdot \log_2 \left( 2\pi \cdot tf \cdot \left( 1 - \frac{tf}{l(d)} \right) \right)}{tf + 1}$ <p>Où <math>qtw(t)</math> est le poids du terme de requête de (<math>t</math>), donné par le rapport <math>\frac{qt_f}{qt_f \max}</math>, avec <math>qt_f</math> le nombre d'occurrences de (<math>t</math>) dans la requête, et <math>qt_f \max</math> le maximum de tous les termes de la requête. (<math>tf</math>) est la fréquence du terme dans le document <math>d</math>, <math>l(d)</math> est la longueur du document. <math>TF</math> est le nombre des occurrences de (<math>t</math>) dans l'ensemble de la collection des documents et <math>TFC</math> est le nombre de jetons dans la collection.</p>
InL2 (DFR)	Inverse document frequency model for randomness, Laplace succession for first normalization, and Normalization 2 for term frequency normalization. (Benkoussas et Bellot, 2015)	$score(d, Q) = \sum_{t \in Q} qtw(t) \cdot \frac{1}{tf_n + 1} (tf_n \cdot \log \frac{N + 1}{n_t + 0.5})$ <p>Où. <math>N</math> est le nombre de documents dans toute la collection et <math>n_t</math> est le nombre de documents contenant (<math>t</math>). <math>tf_n</math> correspond à la somme pondérée de fréquences de terme normalisées <math>tf_f</math> pour chaque champ <math>f</math>, dite : Normalisation 2 et donnée par la formule : <math>tf_n = tf \cdot \log \left( 1 + c \cdot \frac{avgdl}{l(d)} \right)</math> (<math>c &gt; 0</math>). <math>avgdl</math> et <math>l(d)</math>, <math>qt_w</math> et <math>tf</math> (voir ci-dessus.)</p>
LP2 (DFR)	Estimation de Poisson pour le caractère aléatoire, succession de Laplace pour la première normalisation et Normalisation 2 pour la normalisation de la fréquence des termes.	$score(d, Q) = \sum_{t \in Q} qtw \cdot \frac{1}{tf_n + 1} (tf_n \cdot \log_2 \frac{tf_n}{\lambda} + (\lambda - tf_n) \cdot \log_2 e + 0.5 \cdot \log_2 (2\pi \cdot tf_n))$ <p><math>\lambda</math> est la variance et la moyenne d'une distribution de Poisson. Il est donné par <math>F/N</math> et <math>F</math> est beaucoup plus petit que <math>N</math>. <math>F</math> est le terme fréquence de <math>t</math> dans toute la collection.</p>

### 3.6.2.5.2 Trec\_eval

Le « trec\_eval » est un outil standard utilisé par la communauté TREC pour évaluer les classements retournés par les exécutions de requêtes (Buckley, 2004 ; Peddi et al., 2010). L'évaluation avec « Trec\_eval » repose sur l'utilisation de deux principaux fichiers : le premier, appelé « qrels » [pertinence des requêtes], c'est le fichier qui contient les jugements de pertinence pour chaque requête. Le second fichier contient le résultat de classement des documents renvoyés par le système de recherche d'information utilisé. L'exécution de « trec\_eval » avec les options par défaut, retourne les mesures et les informations notamment : la moyenne de la précision moyenne [MAP], précision des R premiers documents [R-précision], précision de cinq premiers documents [P@5] et la précision aux dix premiers documents [P@10] ainsi que de multiples autres mesures.

### 3.6.2.5.3 Langage de requête dans « Terrier »

Terrier offre un puissant langage de requête qui permet à l'utilisateur de spécifier des opérations supplémentaires en plus des requêtes probabilistes normales (Ounis *et al.*, 2007) (<http://terrier.org/docs/v5.1/querylanguage.html>). Ce dernier est basé sur le langage de requête « *Inquery* », a été conçu pour être robuste (Strohman *et al.*, 2005). Il peut gérer à la fois des requêtes avec des termes simples et des requêtes extrêmement complexes. Voici quelques exemples d'écriture de requêtes avec langage de Terrier (Figure 3.3).

- Q : term1 term2 → Récupère les documents contenant term1 ou term2 (ne dois pas nécessairement contenir les deux).
- Q : (term1 term2) → Récupère les documents qui contiennent term1 ou term2, où ils sont traités comme des synonymes l'un de l'autre (ils n'ont pas besoin de les contenir tous les deux).
- Q : term1^2,3 → Le poids de term1 est multiplié par (2,3).
- Q : +term1 +term2 → Récupère les documents contenant à la fois les termes terme 1 et t2.
- Q : +term1 — term2 → Récupère les documents contenant le terme 1 et ne contenant pas le

Figure 3.3 Exemples d'opérations en utilisant le langage de requêtes de terrier.



D'autres fonctions existent aussi, selon les besoins des utilisateurs, il serait possible de reformuler la requête de départ en manipulant le poids et la présence ou pas des termes. Ce langage a été utilisé dans le cadre de ce travail pour la pondération des termes candidats issus des différentes stratégies d'extension suivant notre méthodologie (section 3.8 ; section 3.10.4 et section 3.10.3).

### 3.7 Revue de littérature

#### 3.7.1 Extension de requêtes en RI

Selon Bhogal *et al.* (2007), il est difficile d'exprimer un besoin d'information en utilisant uniquement les termes d'origine d'une requête. C'est pour cette raison que l'extension de cette requête permettra d'effectuer des recherches sur les variations morphologiques des termes d'origines et/ou d'inclure d'autres termes résultant de la désambiguïsation du sens des mots provenant de leur contexte lexical ou sémantique (Bhogal *et al.*, 2007; Mitra, M. *et al.*, 1998). La notion de contexte ici, renvoie vers un domaine de connaissances particulier ou une tâche spécifique (Bhogal *et al.*, 2007). De ce fait, les nouveaux termes fournissent des informations contextuelles à la requête initiale en vue d'améliorer les résultats de la recherche. Les informations contextuelles peuvent être acquises à partir de la rétroaction sur la pertinence, de la cooccurrence des termes et, plus récemment, elle est dérivée de modèles de connaissances comme l'ontologie (Bhogal *et al.*, 2007; Cieslewicz *et al.*, 2018). Cependant, si les termes d'extension ne sont pas choisis avec soin, il existe un risque d'inclure des termes d'extension parasites, ce qui peut élargir les interprétations potentielles de la requête modifiée et ainsi réduire le nombre de documents retournés en lien avec la recherche souhaitée (Crimp et Trotman, 2018).

Comme mentionné, plusieurs approches ont été développées pour permettre cette extension. Ces approches peuvent être classées suivant le paradigme utilisé pour déterminer les caractéristiques d'extension (Carpineto et Romano, 2012). Un exemple de paradigme et d'extension est celui de l'extension par la « rétroaction sur la pertinence » ou en anglais « relevance feedback » qui est une technique (modèle de langage) efficace et couramment utilisée pour l'ajouter des mots utiles à une requête (Mitra, M. *et al.*,

1998). Comme son nom l'indique, cette technique se base sur l'appréciation des utilisateurs des documents retournés et ainsi chercher d'autres mots pertinents dans ces derniers. La pseudo-rétroaction (ou à l'aveugle) est une variante qui à défaut des rétroactions des utilisateurs, considère comme pertinents les documents d'un petit ensemble de documents récupérés à l'aide de la requête initiale. À partir de ce point et sans aucune intervention des utilisateurs, ces documents seront utilisés dans un processus de rétroaction de pertinence pour construire une requête étendue, qui est ensuite exécutée pour récupérer l'ensemble de documents qui seront effectivement présentés aux utilisateurs (Mitra, M. *et al.*, 1998).

D'autres techniques comme celles de l'analyse linguistique exploitent les propriétés globales du langage telles que les variations morphologiques, lexicales, les relations syntaxiques et sémantiques de mots pour développer ou reformuler les termes de la requête. Ils sont généralement basés sur des dictionnaires, des thésaurus ou d'autres sources de connaissances similaires comme *WordNet* (Carpineto et Romano, 2012).

La « navigation » dans les ontologies est une autre technique bien connue et spécifique au langage. Cette approche exploite des modèles de connaissances comme les ontologies ou les thésaurus pour paraphraser la requête de l'utilisateur dans un contexte ou un domaine spécifique. Plusieurs ontologies (spécifiques au domaine ou indépendantes du domaine) ont été utilisées pour la recherche d'information, y compris la combinaison de plusieurs thésaurus (Bhogal *et al.*, 2007; Carpineto et Romano, 2012; Mitra, M. *et al.*, 1998).

Les termes identifiés pour l'extension des requêtes peuvent être lexicalement liés, comme dans un thésaurus ou un système vocabulaire, ou liés statistiquement, comme par une table de co-occurrence dans une collection document. L'extension lexicale peut lier des termes entre différents systèmes de vocabulaire, par exemple un terme très

général comme le terme « cancer » pourrait être étendu par un terme plus clinique comme « carcinome » (Wright *et al.*, 2017).

### 3.7.2 Extension de requêtes en utilisant les ontologies

Utiliser une ontologie dans la recherche d'information consiste à trouver des concepts au sein de cette dernière ayant un lien avec les termes qui existent dans la requête. Ces nouveaux concepts peuvent être utilisés pour étendre cette requête d'origine (Paz-Trillo *et al.*, 2005). Ceci découle selon Dridi (2008) de la vision cognitive du monde, qui présume que la signification d'un texte (mot) dépend des relations conceptuelles avec des objets dans le monde plutôt que des relations linguistiques ou contextuelles trouvées dans des textes ou des dictionnaires. Ce processus d'ajout de termes peut être manuel, automatique ou assisté par l'utilisateur (Bhogal *et al.*, 2007).

De ce fait, l'utilisation des ontologies est considérée comme une approche alternative (ou complémentaire) qui vise à améliorer la recherche en se basant sur la notion de concepts. En effet, les documents et les requêtes peuvent être représentés par des concepts du domaine, au lieu de mots-clés et des termes (ou en plus des termes originaux) permettant ainsi d'effectuer la recherche dans cet espace conceptuel. Selon Egozi *et al.* (2011), cette approche détient la promesse que la représentation des documents et des requêtes par des concepts aboutiront à un modèle de récupération documentaire qui dépend moins des termes utilisés dans la requête originale.

### 3.7.3 Extension des requêtes en utilisant les ontologies dans le domaine médical

Wright *et al.* (2017) dans leur étude a évalué l'apport d'une extension de requêtes en utilisant les synonymes du vocabulaire (MeSH) pour les ajouter à la requête d'origine et ainsi améliorer la récupération documentaire. Les résultats de cette étude montrent que l'utilisation requête limitée à cinq MeSH termes pour chaque terme dans la requête avec une pondération relativement légère, a produit une amélioration globale de

l'infNDCG de 11 % par rapport le système de référence (les requêtes de base sans extension) (voir section 3.6.2.3.6 pour l'unité de mesure infNDCG).

Cieslewicz *et al.* (2018), pour étendre les requêtes, les auteurs ont utilisé des termes similaires obtenus à partir d'une collection de citations de revues PubMed et de la collection du défi de données bioCADDIE<sup>33</sup>. La similarité a été calculée pour chaque ensemble de données à l'aide de word2vec (Mikolov *et al.*, 2013), un modèle permettant d'apprendre des représentations vectorielles de mots à partir de données textuelles non structurées. Un filtrage par la collection de termes médicaux MeSH a été effectué de sorte que les termes d'expansion des requêtes restent dans le domaine médical. Les résultats montrent un gain en pourcentage un peu plus de 4 % (infNDCG) ; obtenu par l'extension de la requête par rapport aux requêtes de base. Les auteurs précisent que les données fournies aux défis de la bioCADDIE ont une structure assez irrégulière (certains types de données manquent dans de nombreux documents), ce qui peut faire une différence.

Ces deux études utilisent la similarité entre les termes des requêtes initiales avec des termes dans les thésaurus. Ceci n'inclut pas l'exploration du contexte sémantique ou lexical de ces termes qui peuvent être déduits en utilisant de plus larges spectres de relations qui existent dans les termino-ontologies.

Pour ce qui est de l'utilisation d'UMLS proprement dite dans l'extension de requête, on peut citer l'étude de Aronson et Rindfleisch (1997). L'approche proposée utilise le programme MetaMap (décrit dans le chapitre II précédent, section 2.9.2) pour associer les concepts UMLS Métathésaurus à la requête d'origine. L'étude conclut que cette

---

<sup>33</sup> BioCADDIE : pour « *Biomedical and Healthcare Data Discovery Index Ecosystem* » est une fondation qui développe des outils d'indexation comme « *DataMed* » qui indexe les métadonnées de base disponibles pour la plupart des ensembles de données. <https://datamed.org/about.php>.

extension est efficace et se compare favorablement aux méthodes basées sur la pseudo-rétroaction. Ils préconisent que la stratégie optimale serait une combinaison des deux techniques (MetaMap-UMLS et la pseudo-rétroaction).

Une autre étude utilisant UMLS et celle de (Zhu *et al.*, 2006), qui compare trois différentes approches pour l'extension de requête : l'analyse locale, l'analyse globale<sup>34</sup> et la repondération des termes basée sur l'ontologie. Les expérimentations ont été menées avec différents moteurs de recherche. La repondération des termes basée sur l'ontologie semble fonctionner mieux que les deux autres approches. Leur approche consiste à reformuler la requête initiale en repondérant les termes clés et en leur associant des synonymes d'UMLS. Les résultats des expériences montrent que la précision moyenne est améliorée jusqu'à 20,3 % et 12,1 % (utilisant les outils LUCENE<sup>35</sup> et LEMUR<sup>36</sup> respectivement) par rapport à la requête de base. De leur côté, Dong *et al.* (2011) ont proposé une extension de requête basée sur un graphe ontologique. Ce graphe est construit en utilisant quatre vocabulaires contrôlés qui existent dans UMLS (MSH, SNOMEDCT, CSP, AOD)<sup>37</sup>. En utilisant un algorithme de type PageRank personnalisé sur ce graphe et avec une pondération soigneusement choisie afin de sélectionner

---

<sup>34</sup> L'analyse globale examine les occurrences de mots et les relations dans le corpus et qui forment un tout. Puis utilise ces informations pour étendre une requête particulière. L'analyse locale, en revanche, n'implique que les documents les mieux classés récupérés par la requête d'origine.

<sup>35</sup> Lucene est un logiciel libre offrant de puissantes fonctionnalités d'indexation et de recherche, ainsi que des fonctionnalités de vérification orthographique et d'analyse/tokénisation avancée. <https://lucene.apache.org/>

<sup>36</sup> L'outil Lemur est conçu pour faciliter la recherche dans la modélisation du langage et la recherche d'informations (IR). Cet outil prend en charge la construction de systèmes de recherche de texte de base en utilisant des méthodes de modélisation du langage, ainsi que des méthodes traditionnelles telles que celles basées sur le modèle d'espace vectoriel et Okapi. <https://www.lemurproject.org/>

<sup>37</sup> MSH: Medical Subject Headings; SNOMEDCT: SNOMED Clinical Term; CSP: CRISP Thesaurus; AOD: Alcohol and other Drug.

les « meilleurs » concepts biomédicaux. Leurs expérimentations ont démontré une amélioration de 8,2 % par rapport à une extension de requête basée sur la rétroaction de pseudo-pertinence et de 17,7 % par rapport à une requête sans extension (Mesure utilisée : précision moyenne interpolée en 11 points). Aussi, l'étude a permis d'établir que les concepts biomédicaux généralisés sont l'une des raisons de la dégradation des performances de l'extension de requête. Cette dégradation est de l'ordre de 13 % (précision moyenne interpolée en 11 points) par rapport à une requête de base.

Les études citées et indépendamment des approches utilisées, des collections de tests, des ontologies ainsi que les mesures d'évaluation adoptées, démontrent que l'extension de requête en utilisant des thésaurus ou UMLS est une méthode efficace et moins onéreuse. En utilisant uniquement des ressources ontologiques, ces méthodes permettent d'améliorer la recherche d'information. Cette extension est comparable à celle de la rétroaction de pseudo-pertinence avec des améliorations qui oscillent entre 10 et 20 % par rapport aux requêtes de base.

#### 3.7.4 Utilisation des éléments PICO dans la recherche d'information

Les études cliniques et les questions de recherche s'articulent sur quatre aspects qui sont : la population, patient ou problème (P), l'intervention (I), la comparaison (C) et le résultat (O), qui sont connus sous le nom d'éléments PICO (voir le tableau 3.4 pour les exemples) (Boudin, Shi, *et al.*, 2010; Demner-Fushman et Lin, 2007; Huang, X. *et al.*, 2006). Il serait donc intuitivement plus avantageux d'utiliser ces éléments dans la recherche des articles scientifiques pertinents. Voici un exemple d'une question en format « PICO » : *chez les patients avec le diabète de type II (P) qui présentent un risque*

élevé d'événements macro-vasculaire (Pr), est-ce que la pioglitazone <sup>38</sup>(I) réduit la mortalité (O) ?

Tableau 3.4 Catégorie PICO et exemple de concepts y afférents.

PICO	Exemple
P : Population, Patient ou problème	Problèmes: <i>[Disease] e.g., "panic disorder"</i> <i>[Behaviour] e.g., "oppositional behaviours"</i> <i>[Symptom] e.g., "leg cramps"</i> Population: <i>[Age] e.g., "40-years-old"</i> <i>[Gender] e.g., "male"</i> <i>[Treatment Status] e.g., "delayed treatment"</i> <i>[Physical Condition] e.g., "healthy"</i> <i>[Medical History] e.g., "with prior attacks"</i> <i>[Treatment &amp; Drug] e.g., "taking hormone replacement therapy."</i> <i>[Disease] e.g., "nonvalvular atrial fibrillation"</i>
I : Intervention (exposition à considérer qui peut être un traitement, un test)	Intervention et Comparaison :
C : contrôle (comparaison par rapport à d'autres traitements, placebo ou standards de traitement)	<i>[Treatment &amp; Drug] e.g., "warfarin"</i> <i>[Procedure] e.g., "transvaginal ultrasound"</i> <i>[Diagnostic test] e.g., "Pap smear"</i> <i>[exposure] e.g., "maternal smoking"</i> <i>[disease] e.g., "a flare-up of Crohn's"</i>
O : outcome (résultats recherchés)	Outcome: <i>[Treatment outcome] e.g., "Fibroid volume reduction"</i> <i>[Patient outcome] e.g., "decreased mortality"</i>

<sup>38</sup> La pioglitazone est un médicament utilisé pour diminuer le taux de glucose sanguin élevé (le sucre dans le sang) associé au diabète de type 2.



Huang, X. *et al.* (2006) et à la suite de leur étude d'évaluation des éléments PICO pour la recherche d'articles scientifique, dénotent que les questions formulées en langage naturel par les cliniciens se sont avérées dépourvues des éléments qui constituent une « requête bien formulée » dans la plupart des cas. Ils confirment aussi comme les études de (Aslam, S. et Emmanuel, 2010; Ebell, 1999; Richardson *et al.*, 1995), que l'utilisation d'une question bien formulée selon la structure PICO est plus efficace pour faciliter la recherche d'une réponse précise au sein d'une grande base de données de citation médicale (Schardt *et al.*, 2007).

L'une des premières études pour l'utilisation des éléments PICO dans la recherche d'information et sans doute celle de Demner-Fushman et Lin (2007). Cette étude propose un modèle de réponse aux questions (*question-answering*) avec en premier temps, une approche pour encoder (manuellement) la question sous forme de la structure PICO (Patient/Problème, Intervention, Comparaison et Outcome) pour rechercher des articles scientifiques à partir de sources externes (citation dans MEDLINE et PubMed). Par la suite, un reclassement de la liste initiale de citations retrouvées est effectué sur la base de la pertinence d'un document est calculée par un score linéaire ( $S_{PICO} = S_{\text{problème}} + S_{\text{population}} + S_{\text{intervention}} + S_{\text{résultat}}$ ) avec les éléments PICO détectés dans ce document à partir de la question PICO. À partir des articles les mieux notés, une nouvelle extraction des éléments PICO est effectuée pour permettre de générer une réponse à la question initialement posée. Les expériences effectuées sur 24 questions cliniques réelles montrent que le résultat de l'approche est supérieur à une recherche PubMed traditionnelle. Toutefois, ces travaux s'appuient fortement sur une annotation automatique et manuelle (en prétraitement) des éléments PICO dans les requêtes et dans les documents.

Dans une autre étude intéressante sur l'utilisation des éléments PICO, Boudin, Nie, *et al.* (2010) présentent une approche de modélisation du langage qui intègre le document et la structure PICO à des fins de recherche d'information clinique. Leur idée consiste

à détecter les éléments PICO dans les documents et d'utiliser ces résultats dans le processus de recherche. Plus explicitement, si le modèle du document est susceptible de générer la requête alors le document correspond à cette requête. Les auteurs ont utilisé la classification automatique des éléments PICO dans la requête (rédigé manuellement) et dans les résumés d'articles scientifiques (PubMed). Cependant, cette approche ne fonctionne pas bien en raison de la difficulté à parvenir à un balisage cohérent de ces éléments PICO dans le document. C'est pour cette raison qu'ils proposent une approche moins exigeante qui attribue différents poids à différentes parties d'un document avec une détection manuelle de ces éléments dans les documents. L'étude présente trois différentes approches qui étendent l'approche du modèle de langage classique<sup>39</sup>. La première approche utilise les informations structurelles PICO des documents, la seconde tire parti de la structure PICO dans la requête tandis que le troisième combine simplement les deux premiers modèles. Pour résumé, les résultats varient suivant les approches entre 22,5 % à 61,5 % MAP par rapport aux résultats de recherche des questions de base limitée à d'une dizaine de questions sur le diabète. Malgré cela, l'étude dénote la difficulté de définir les éléments PICO que ce soit dans le document ou la requête. Beaucoup d'études (BoudinNie, *et al.*, 2010; Huang, K.-C. *et al.*, 2013; Jin et Szolovits, 2018, 2020) se sont intéressés par la suite à la complexité de l'identification des éléments PICO dans des résumés d'article et dans des corps complets d'articles, mais pas nécessairement pour améliorer la recherche d'information. Ces recherches seront abordées dans la section 3.9.6.

En ce qui concerne nos travaux, nous avons expérimenté l'utilisation de l'orientation PICO dans le choix et la pondération des concepts retrouvés par ICPU pour l'extension

---

<sup>39</sup> Suivant les approches du modèle de langage ; un document correspond bien à une requête si son modèle de langage est susceptible de générer la requête, le modèle uni gramme - également appelés modèles de sacs de mots – qui suppose qu'il n'y a pas de structure dans les requêtes ou les documents.

de requête, et ce, afin de comprendre l'effet de cette discrimination sur les résultats de recherche d'articles scientifiques (voir section 3.8 et section 3.9.6).

### 3.8 Méthodologie pour extension de requêtes à partir d'extraction ICPU avec ou sans orientation PICO.

Notre méthodologie de recherche s'articule principalement sur l'utilisation du système ICPU (Identification des concepts pertinents à partir d'UMLS) développé au chapitre précédent. Cet outil permet l'extension de requêtes à partir du contexte sémantique et lexical de chacun des concepts constituant ces dernières. Ces requêtes peuvent être des clauses d'une règle de décision d'un SADC ou simplement du texte libre décrivant la condition médicale problématique d'un patient. Le processus de recherche d'information utilise la nouvelle requête étendue par les nouveaux concepts avec un effort de pondération suivant leur importance dans le graphe UMLS exploré. Pour rappel, ICPU parcourt le voisinage des concepts de départ par une marche aléatoire paramétrée sur certaines termino-ontologies d'UMLS. Cette méthode exploite les relations hiérarchiques ou sémantiques ainsi que la co-occurrence entre ces concepts dans l'historique des publications scientifiques. Les requêtes proviennent du corpus de test de la campagne TREC (voir section 3.9.2) pour la médecine de précision et constituent un ensemble de 30 requêtes de base.

Les extensions ont été générées suivant trois approches : 1) une extension avec uniquement les nouveaux concepts issus de l'ICPU 2) une extension avec ICPU, mais en favorisant les concepts appartenant à des catégories PICO en alternance et finalement 3) appliquer une autre extension de requête bo1 supplémentaire sur les ensembles précédents (section 3.8 et 3.10.5).

Ces différentes approches subiront deux variations : 1) le type de relations utilisées pour l'exploration (associatives ou hiérarchiques) et l'utilisation (ou pas) d'une pondération que nous préconisons dans le cadre de ce travail de recherche. Ceci donnera un ensemble de variantes détaillées dans la section 3.10.5.

### 1) Extension avec ICPU

Cette approche consiste à extraire un ensemble de concepts potentiellement pertinents à partir d'UMLS, et ce, en utilisant les termes initiaux du sujet recherché et l'outil ICPU avec un paramétrage que nous considérons adéquats pour cet objectif (longueur de la marche, nombre de marches, seuil de sélection des concepts, type de relations hiérarchiques ou sémantiques ainsi que les vocabulaires liés aux domaines médicaux). Une pondération sera appliquée aux nouveaux termes (voir section 3.10.3). Pour finir, une évaluation de la qualité des résultats de recherche d'information de ce type d'extension sera réalisée (Figure 3.4).

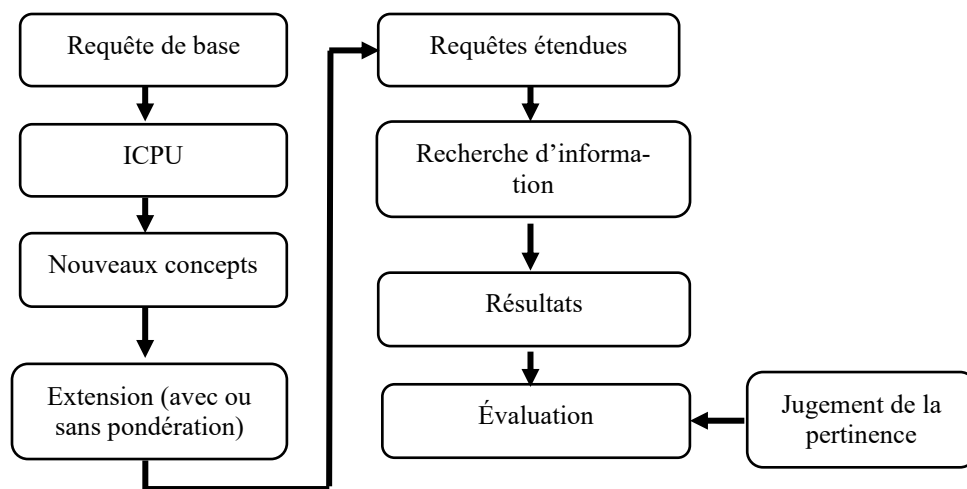


Figure 3.4 Méthodologie d'extension des « sujets » en utilisant ICPU.

### 2) Extension ICPU avec orientation PICO

Dans ce volet, l'approche proposée consiste à ajouter aux concepts de la requête de base et/ou concepts issus d'ICPU, de nouveaux concepts identifiés comme appartenant à l'une ou l'autre des catégories PICO (Problème, intervention, et outcome voir section 3.9.6 et 3.9.7) (Figure 3.5). Ces nouveaux concepts figurent dans des études cliniques enregistrées auprès d'un organisme gouvernemental et considérées comme

proches de la problématique médicale traités par le sujet recherché. Ce rapprochement est obtenu par la sélection des études pour lesquels les concepts issus d'ICPU figurent dans leur problématique médicale. Il est à noter que la catégorie « comparaisons » n'est pas utilisée dans nos expérimentations. Du fait que cette catégorie n'est pas explicitement définie dans la structure XML des fichiers d'enregistrements des études cliniques (voir section 3.9.6).

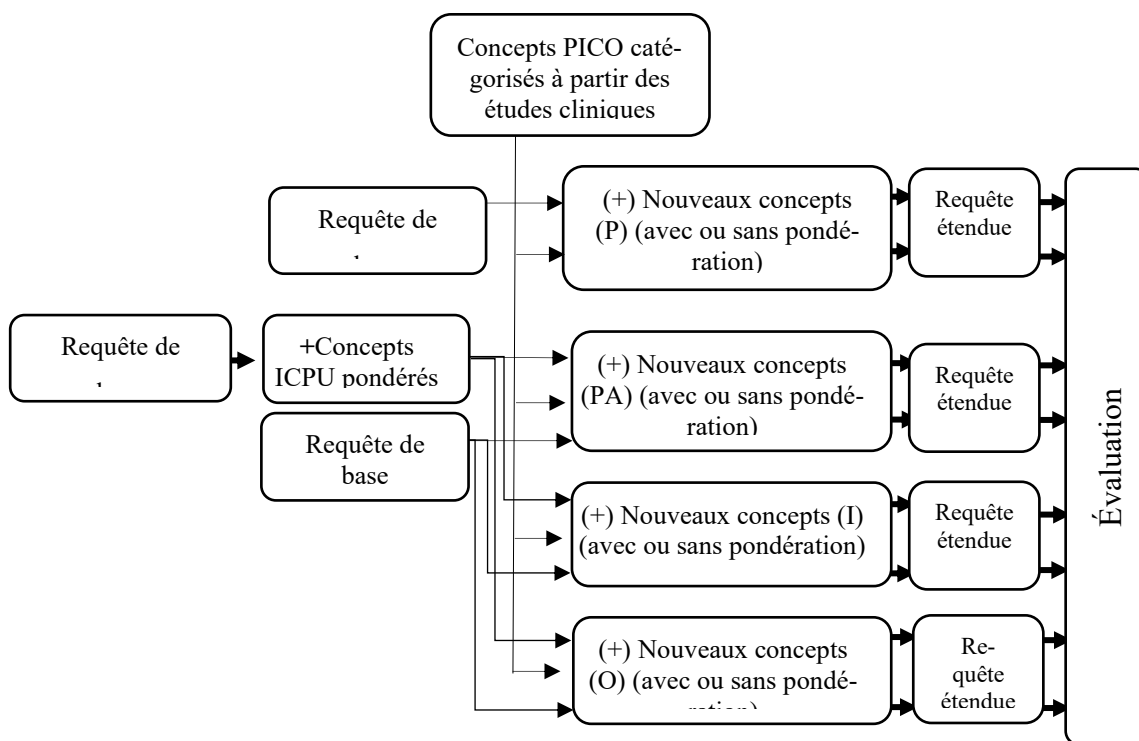


Figure 3.5 Variantes d'extension suivant les différents groupes de concepts PICO.

Voici les différentes variantes pour l'application de l'extension orientée PICO (Figure 3.5) :

- Concepts (P) : Exceptionnellement, cette variante n'ajoute pas de nouveaux concepts, mais plutôt elle permet d'utiliser uniquement les concepts qui figurent

dans les concepts retrouvés par ICPU et en même temps dans les concepts catégorisés comme (P) dans les études cliniques. L'objectif de cette sélection est de connaître si un ensemble « minimale » de concepts (P) permet d'améliorer la recherche (sans les autres concepts ICPU supplémentaires). Cette extension est testée uniquement avec la requête initiale.

- Concepts (PA) : PA pour dire « problèmes autres ». Ceci désigne les concepts de la catégorie « problème\population » qui n'existaient pas dans la requête initiale ou parmi les concepts issus d'ICPU. Ces concepts sont fréquemment utilisés dans la description des études en recherche clinique. La figure 3.6 décrit les étapes et des exemples de concepts ainsi retrouvés et utilisés.

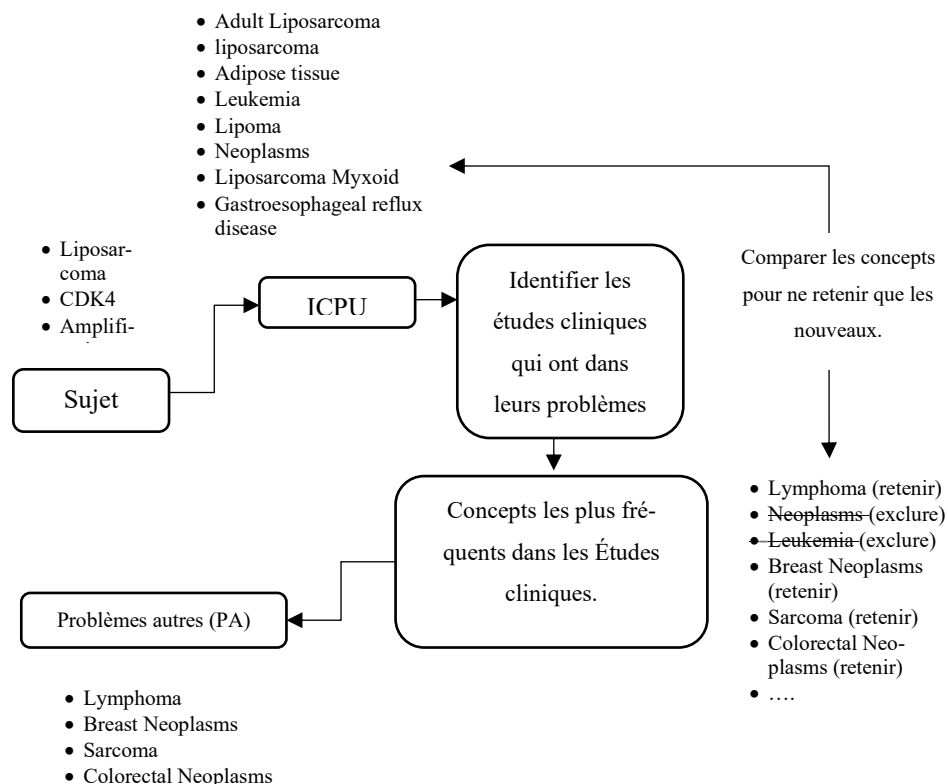


Figure 3.6 Processus d'identification des problèmes "autres" fréquemment utilisés des études cliniques.

- Concepts (I) : Ces concepts sont identifiés comme appartenant à la catégorie « intervention ». Ces derniers n'existent pas dans la requête initiale et le principe pour les identifier est identique à la catégorie (PA) précédente. Cela revient à utiliser les concepts ICPU pour identifier les études cliniques qui traitent de la problématique médicale. Ensuite, les concepts (I) seront sélectionnés par ordre de fréquence dans les études cliniques.
- Concept (O) : Idem que pour la section précédente. Développer.

### 3) Appliquer une extension supplémentaire

Pour les deux volets précédents, une extension supplémentaire (Bose-Einstein 1 ; voir section 3.9.8) sera appliquée sur les termes des requêtes déjà étendues. Cette extension ajoute 10 nouveaux termes qui seront extraits à partir de trois premiers documents les mieux classés retournés par les extensions. L'objectif de cette démarche permettra de déterminer si une extension de type « rétroaction » sur les résultats des requêtes déjà étendus suivant les scénarios précédents donne de meilleurs résultats.



### 3.9 Ressources, prétraitements et préparation

Pour l'évaluation de notre méthodologie d'extension des requêtes initiales, nous avons utilisé les ressources suivantes :

1. L'outil ICPU ;
2. Les « sujets » de recherches pour le TREC en Médecine de précision Ad-hoc à la conférence TREC ;
3. Collection des résumés scientifiques de Pub Med (un instantané) ;
4. Fichier d'appréciation de la pertinence pour les résumés scientifiques ;
5. Plate-forme de recherche d'information « Terrier » (section 3.6.2.5).

Aussi, d'autres ressources ont été nécessaires pour la conduite des expérimentations, parmi celles-ci :

6. Indexation de la collection documentaire PubMed à l'aide de « terrier » ;
7. Extraction des concepts usuellement utilisés dans les éléments PICO des études cliniques enregistrés sur le site [Clinicaltrial.gov](http://Clinicaltrial.gov) ;
8. Codage des concepts PICO.

#### 3.9.1 L'outil ICPU (Identification des concepts pertinents à partir des UMLS)

Cet outil a été développé et détaillé dans le chapitre précédent (voir 2.9.3 du Chapitre II). Il permet d'extraire des concepts potentiellement pertinents à partir d'UMLS en utilisant le contexte sémantique des termes initiaux de la requête. Le résultat sera utilisé pour améliorer les résultats de recherche d'information par extension de requête dans le cadre de ce chapitre.

### 3.9.2 Les « Sujets » de la conférence TREC ad hoc pour Médecine de précision et l'aide à la décision.

Le volet de la conférence TREC ad hoc pour la médecine de précision et d'aide à la décision (<http://www.trec-cds.org/2017.html>) met l'accent sur un cas d'utilisation important dans l'aide à la décision clinique : fournir des informations utiles sur la médecine de précision aux cliniciens traitant des patients atteints de cancer. Deux importants points nous ont incités à utiliser les « sujets » (ou « *topics* » *en anglais*) de cette campagne :

- 1) Cette campagne propose des cas synthétiques (sujets) élaborés par des oncologues du « *M.D. Anderson Cancer Center* » de l'Université du Texas. Chaque cas décrira la maladie du patient (type de cancer), les variantes génétiques (les gènes impliqués), les informations démographiques de base (âge, sexe) et d'autres facteurs potentiels pouvant être pertinents (Figure 3.7).
- 2) La disponibilité du fichier de « Jugement » (voir section 3.9.4) qui permet d'évaluer les résultats du système de RI au vu des documents considérés comme pertinents pour chaque sujet de cette campagne ad hoc du TREC.

```
<topics task="2017 TREC Precision Medicine">
<topic number="1">
<disease> Acute lymphoblastic leukemia </disease>
<gene> ABL1, PTPN11 </gene>
<demographic>12-year-old male </demographic>
<other> No relevant factors </other>
</topic>
```

Figure 3.7 Exemple de « sujet » du TREC 2017.

### 3.9.3 La collection de résumés scientifiques PubMed « un instantané »

Il existe deux collections de documents utilisés par les candidats à la campagne TREC pour le volet « Médecine de précision » : les résumés scientifiques et les essais cli-

niques. (<http://www.trec-cds.org/2017.html>). La collection utilisée est celle des résumés scientifiques des articles publiés entre 1996 et jusqu'à l'année de la campagne TREC de 2017. Cet ensemble est fourni sous forme de fichier XML compressé de 20 giga-octets. Pour information, le fichier des essais cliniques qui est un instantané des études cliniques enregistrées auprès du site « ClinicalTrials.gov ».

### 3.9.4 Fichier d'appréciation de la pertinence pour les résumés scientifiques.

Le jugement de pertinence des documents retournée par les SRI constitue la tâche la plus difficile lors de la construction d'un corpus de test (Samimi et Ravana, 2014). Le résultat de ce jugement, dans le cadre de la TREC, est un fichier indiquant pour chaque requête et pour chaque document le degré de pertinence. (Figure 3.8).

```

1 0 10220412 0
1 0 10323080 1 →pour la requête n° 1, le document n° 10323080 est pertinent.
1 0 10466061 0 →pour la requête n° 1, le document n° 10466061 n'est pas pertinent.
1 0 10513887 0
1 0 10587703 0
1 0 10634644 0
1 0 10727978 0
1 0 10755400 2 →pour la requête n° 1, le document n° 10755400 très pertinent.
1 0 10780569 0

```

Figure 3.8 Exemple de fichier de jugement avec explication (<https://trec.nist.gov/data/precmed/qrels-final-abstracts.txt>).

Il serait intéressant d'expliquer que ce fichier de résultats de pertinence a été produit par une approche dite de « *pooling* ». En effet, il est irréalisable de juger 800 000 documents pour chaque sujet, ceci prendrait 6 500 heures, en supposant que chaque document puisse être jugé en seulement 30 secondes (Voorhees, E. et Harman, D., 2000). Cette méthode permet d'éviter cette évaluation exhaustive, et ce, en utilisant qu'un sous-ensemble restreint de documents appelés : le « pool » (Sparck Jones et Van Rijsbergen, 1975; Voorhees, E. et Harman, D., 2000). Chaque document de ce pool est jugé par l'auteur du sujet. Les documents qui ne sont pas dans le pool sont censés être « sans rapport » avec ce sujet. Le « pool » est alimenté par les X premiers documents

(généralement,  $X = 100$ ) d'un certain nombre d'exécutions (runs) des participants au concours TREC. Ces documents sont considérés comme étant les plus susceptibles d'être pertinents pour le sujet en question. L'utilisation de cette méthode a été remise en question parce que les documents non évalués sont censés ne pas être pertinents (le pool est considéré comme incomplet). Les critiques soutiennent aussi que les scores d'évaluation pour les méthodes de recherche qui n'ont pas contribué au pool seront pénalisés par rapport aux méthodes qui ont contribué, et ce, parce que les non-contributeurs auront des documents hautement classés, mais non jugés. Selon des études comme celles de (Voorhees, 2000; Voorhees, E. M. et Harman, D. K., 2000; Zobel, 1998), rapporté par (Voorhees, E. et Harman, D., 2000), démontrent qu'il y'a une preuve très forte sur la fiabilité de cette méthode dite de « pooling » pour les évaluations comparatives des cycles d'extraction et que la différence entre les scores de pertinences résultants du pool « incomplets » sont très minime par rapport à l'utilisation d'autres méthodes d'évaluation de pertinence.

### 3.9.5 Indexation de la collection documentaire PubMed à l'aide de « terrier »

Pour les besoins de nos expérimentations et telle que proposé par la TREC (ad hoc pour la médecine de précision et d'aide à la décision), un instantané des résumés de PubMed<sup>40</sup> à la date de janvier 2017 a été utilisé pour les résumés scientifiques. Cet ensemble est au format XML et un prétraitement est requis pour le mettre au format TREC afin que le logiciel Terrier puisse l'indexer. Après décompression des données, l'ensemble des fichiers XML a été analysé pour extraire uniquement le numéro PubMed de l'article et le résumé (Abstract). Le résultat sera au format TREC conventionnel pour l'élaboration d'une collection de tests (Figure 3.9).

---

<sup>40</sup> PubMed comprend plus de 29 millions de citations dans la littérature biomédicale de MEDLINE, des revues de sciences de la vie et des livres en ligne. (<https://www.ncbi.nlm.nih.gov/pubmed/>)

```

<DOC>
<DOCNO>9041915</DOCNO>
Excessive blood transfusion during orthotopic liver transplantation (OLT) is correlated with a lower graft survival rate. Experimentally, excessive hemorrhage is associated with endotoxemia and release of pro-inflammatory cytokines. To measure the concentrations of plasma endotoxin and pro-inflammatory cytokines during OLT and to investigate their relation to intraoperative blood loss and graft viability. A prospective case series in a liver transplantation centre. Thirty patients who underwent OLT. Group 1 comprised 6 patients whose operative blood transfusion requirement was 10 U or more; group 2 comprised 24 patients whose operative blood transfusion requirement was less than 10 U. The following factors were measured in the plasma before and after OLT: interleukin (IL)-1 beta, IL-6, tumour necrosis factor alpha, hepatocyte growth factor, endotoxin, hyaluronic acid, and lactate. Graft viability. Two patients in group 1 died. All 24 patients in group 2 survived after they underwent OLT. The responses of IL-6 and IL-1 beta in group 1 were striking compared with those in group 2, and they were accompanied by an elevation of the endotoxin concentration and a subsequent elevation of the concentrations of hepatocyte growth factor, hyaluronic acid, lactate, and other factors that reflected graft viability. The changes in IL-6 seemed to respond to the excessive intraoperative hemorrhage, to provoke the elevation of the endotoxin concentration, and to be associated with the graft viability. The prevention of excessive intraoperative bleeding and the subsequent response of proinflammatory cytokines may be contributing factors to the success of liver transplant surgery. </DOC>

```

Figure 3.9 Format du document TREC conventionnel pour la collection

### 3.9.6 Extraction des concepts usuellement utilisés dans les éléments PICO des études cliniques enregistrés sur le site Clinicaltrial.gov

Dans l'objectif de sélectionner les termes d'extension en utilisant une orientation PICO, il devient nécessaire de trouver un moyen pour identifier à quels éléments PICO ces concepts s'apparentent. Cette tâche est généralement considérée comme une tâche de classification des phrases ou des segments de texte (Wallace *et al.*, 2016). Selon Boudin, Nie, *et al.* (2010), l'extraction des éléments PICO dans des documents est une tâche particulièrement difficile. En effet, l'annotations manuelle pour délimiter les sections PICO du texte nécessite des experts humains ; elle a été grandement utilisée comme référence pour l'apprentissage et les tests des classifieurs dans la détection des éléments PICO (Boudin, Nie, *et al.*, 2010; Demner-Fushman et Lin, 2007; Huang, K.-C. *et al.*, 2013; Huang, X. *et al.*, 2006). L'inconvénient de ces approches reste le travail

manuel fastidieux d'annotation qui se traduit par de petits ensembles de données étiquetés. De plus, il peut y avoir des divergences entre les annotateurs sur la classification des textes par rapport au cadre PICO. Néanmoins, Boudin et al. (2010 b, a) ont montré que l'annotation PICO générée automatiquement pour des résumés peut améliorer la recherche d'information biomédicale, même si le résultat de cette annotation est entaché de bruit (des erreurs dans l'annotation). D'autres approches ont été proposées pour améliorer l'identification automatique des éléments PICO dans des textes ou dans les résumés d'articles comme celle de (Wallace *et al.*, 2016) ou celles de (Jin et Szolovits, 2018, 2020). Wallace *et al.* (2016) proposent un modèle d'apprentissage automatique pour extraire des phrases relatives aux éléments PICO à partir d'articles en texte intégral décrivant les études cliniques randomisé.

L'annotation qui servira à entraîner le modèle sera établie sur la base d'une ressource (semi-structurée) existante - la base de données Cochrane<sup>41</sup> - et d'une méthode appelée « supervision à distance (DS) ». Cette supervision à distance est générée à l'aide d'heuristiques pour mapper à partir des structures de données existantes des pseudo-annotations qui se rapprochent des étiquettes cibles (P/I/O). Une autre approche est celle de Jin et Szolovits (2020), ces derniers proposent un modèle d'apprentissage en profondeur pour la reconnaissance des éléments PICO dans les résumés basée sur la mémoire bidirectionnelle à long terme bi-LSTM<sup>42</sup>). Ce modèle a été évalué par une validation

---

<sup>41</sup> La Cochrane Library est une bibliothèque qui comprend plusieurs bases de données, spécialisées en médecine et santé, produites par la Cochrane. Parmi ses bases figure celle des revues systématiques Cochrane évaluant les effets des interventions en santé (en prévention, diagnostic, thérapeutique et rééducation) et les axes d'amélioration de la méthodologie des études. (<https://www.cochrane.org/>). La Cochrane est une organisation caritative internationale britannique formée pour disséminer les résultats de la recherche médicale .

<sup>42</sup> Bi-LSTM : (*Bidirectional recurrent neural networks*) le modèle des réseaux neuronaux récurrents bidirectionnels connectent deux couches cachées de directions opposées à la même sortie. Avec cette forme d'apprentissage profond génératif, la couche de sortie peut obtenir simultanément des informations sur les états passés et futurs du réseau. (Wikipédia).

croisée par 10<sup>43</sup> sur des corpus annotés par auteurs (PubMed-PICO). Les résultats montrent des valeurs entre 0,92 de précision et 0,89 pour le rappel pour le classement des éléments de texte dans la catégorie Patient/Population par exemple.

L'intégration et l'adaptation des nouvelles méthodes évoquées précédemment nécessitent un effort considérable d'intégration et l'utilisation d'annotateurs pour l'apprentissage et l'évaluation de ces implémentations. Ceci peut dépasser nos objectifs de recherche dans le cadre de ce chapitre où l'amélioration de l'identification PICO dans des textes n'est pas l'objet de ces travaux. Néanmoins, ceci reste une voie d'amélioration possible pour accentuer la qualité de la classification des concepts suivant les catégories PICO de notre méthodologie dans une étape future.

Pour pouvoir réaliser les objectifs de notre travail de recherche et ainsi explorer la piste de l'extension en priorisant les concepts avec une orientation PICO, nous nous sommes inspirés avec modification de l'approche de (Kim, H. *et al.*, 2016) qui a proposé d'utiliser une source structurée (ou plutôt semi-structurée.) Pour permettre de rattacher des concepts à certains éléments PICO. La ressource semi-structurée provient du registre des essais cliniques du site « ClinicalTrials.gov ». Ce service d'inscription en ligne des essais de recherche clinique mis en place par l'US National Library of Medicine (NLM) en 1999 dans le but d'accroître la transparence de la recherche clinique. En 2019, « ClinicalTrials.gov » archivait plus de 327 011 essais réalisés dans 192 pays. Depuis 2007,

---

<sup>43</sup> La validation croisée (« cross-validation ») est, en apprentissage automatique, une méthode d'estimation de fiabilité d'un modèle fondé sur une technique d'échantillonnage. La validation croisée à k blocs, « k-fold cross-validation » : on divise l'échantillon original en k échantillons (ou « blocs »), puis on sélectionne un des k échantillons comme ensemble de validation pendant que les k-1 autres échantillons constituent l'ensemble d'apprentissage. Après apprentissage, on peut calculer une performance de validation. Puis on répète l'opération en sélectionnant un autre échantillon de validation parmi les blocs prédéfinis. À l'issue de la procédure nous obtenons ainsi k scores de performances, un par bloc. La moyenne et l'écart type des k scores de performances peuvent être calculés pour estimer le biais et la variance de la performance de validation.

les essais couverts par la loi portant modification de la « Food and Drug Administration Act (FDAA) » doivent être mis à jour dans un délai de 20 ans avec des données sommaires sur les résultats (<https://clinicaltrials.gov/ct2/about-site/for-media>).

Le registre « ClinicalTrials.gov » contient des informations détaillées sur chaque essai clinique au format semi-structuré en XML. Chaque enregistrement d'un essai ou d'une étude contient des informations telles que : l'objectif de l'étude, les conditions médicales d'inclusion et d'exclusion des participants, la taille de l'échantillon (nombre de participants à l'étude), la méthodologie de l'étude, les bras d'étude et les résultats de l'étude. Dans la présente étude. En partant des conclusions de l'étude de Kim, H. *et al.* (2016), nous avons repris le concept en modifiant l'algorithme pour extraire d'une façon plus ciblée est sur un plus grand nombre d'études.

- 4) Exemple d'une partie du fichier XML pour la description/enregistrement d'une étude (Figure 3.10) :

```

▼<clinical_study>
  ▼<!--
    This xml conforms to an XML Schema at:
    https://clinicaltrials.gov/ct2/html/images/info/public.xsd
  -->
  ▼<required_header>
    ▼<download_date>
      ClinicalTrials.gov processed this data on January 15, 2020
    </download_date>
    <link_text>Link to the current ClinicalTrials.gov record.</link_text>
    <url>https://clinicaltrials.gov/show/NCT03478891</url>
  </required_header>
  ▼<id_info>
    <org_study_id>180069</org_study_id>
    <secondary_id>18-I-0069</secondary_id>
    <nct_id>NCT03478891</nct_id>
  </id_info>
  ▼<brief_title>
    Safety and Pharmacokinetics of a Human Monoclonal Antibody, VRC-EB0MAB092-00-AB (Mab114), Administered Intravenously to Healthy Adults
  </brief_title>

```

Figure 3.10 Partie d'un fichier XML d'enregistrement d'un essai clinique (<https://clinicaltrials.gov/ct2/show/NCT03478891?displayxml=true>)



Les règles d'extraction XPath<sup>44</sup> que nous avons établies :

Tableau 3.5 Règles d'extraction XPath des concepts pour chaque élément PICO

Élément Pico	XPath	Exemple
Problème/Population	<pre>//clinical_study/condition_browse/mesh_term //clinical_study/condition</pre>	<pre>&lt;Condition&gt; HIV Infections &lt;/condition&gt; &lt;Mesh_term&gt; Immunologic Deficiency Syndromes &lt;/mesh_term&gt; &lt;Mesh_term&gt; Retinitis &lt;/mesh_term&gt; &lt;Mesh_term&gt; Cytomegalovirus Retinitis &lt;/mesh_term&gt; &lt;Mesh_term&gt; Acquired Immunodeficiency Syndrome &lt;/mesh_term&gt;</pre>
Intervention	<pre>//clinical_study/intervention/intervention_name //clinical_study/intervention/intervention_type //clinical_study/intervention_browse/mesh_term</pre>	<pre>&lt;Intervention_type&gt; Drug &lt;/intervention_type&gt; &lt;Intervention_name&gt; Ganciclovir &lt;/intervention_name&gt; &lt;Intervention_name&gt; Foscarnet &lt;/intervention_name&gt; &lt;Mesh_term&gt; Phosphonoacetic Acid &lt;/mesh_term&gt;</pre>
Outcome	<pre>//clinical_study/primary_outcome/measure //clinical_study/secondary_outcome/measure</pre>	<pre>&lt;Measure&gt; Morbidity &lt;/measure&gt;</pre>

D'autres types d'information ont été extraits durant le traitement des fichiers XML, notamment l'intervalle d'âge (minimum, maximum) des patients inclus dans ces études, les critères d'inclusions et d'exclusions pour le recrutement. Toutes ces informations peuvent être importantes, car elles peuvent contenir des informations très pertinentes. Toutefois, un prétraitement est nécessaire pour pouvoir les exploiter. Pour nos

---

<sup>44</sup> XPath est un élément majeur de la norme XSLT. XPath peut être utilisé pour naviguer à travers des éléments et des attributs dans un document XML.

travaux, nous nous limiterons uniquement aux éléments repris dans le (Tableau 3.5). Ces rubriques peuvent faire l'objet d'extension de notre approche dans le futur.

### 3.9.7 Coder les concepts PICO en utilisant UMLS (Table TrialPico)

Les résultats d'extraction de la section 3.9.6 précédente sont enregistrés sous la forme d'une table que nous avons nommée « TrialPico » et qui comporte les éléments suivants : 1) le numéro de l'étude clinique donnée par le site (<https://clinicaltrials.gov/>) ; 2) le code UMLS concordant avec le concept retrouvé. Un champ qui indique à quelle section « PICO » appartient le concept retrouvé et finalement le texte tel que trouvé dans la section PICO du fichier XML.

Le code (*cui*) d'UMLS est récupéré par comparaison entre la colonne « description » de TrialPico et le champ STR de la table « *umls. mrconso* » (voir la section 2.9.1 du chapitre II). Nous avons dû optimiser cette requête vu la grandeur de la jointure entre les deux tables « *umls. mrconso* » (8 582 934 lignes) et TrialPico (4 392 844 lignes). Au moment de rédiger ce paragraphe, un ensemble de 184 700 lignes (juillet 2019) qui représente 21 780 (études cliniques).

Note importante : Quelques lignes de TrialPico restent sans code (*cui*) équivalent, car l'information récupérée du fichier XML n'a pas permis de retrouver les *cui umls* par une comparaison directe. Pour y parvenir, un second traitement du champ « description » serait nécessaire. Ce traitement nécessite un traitement de texte qui pourrait être fait dans la suite des travaux de recherche de cette thèse. Dans nos travaux, nous nous sommes limités au premier passage qui déjà a fourni beaucoup d'information.

### 3.9.8 Expansion de requête préexistante dans Terrier

La plate-forme d'indexation et de recherche « Terrier » offre des méthodes d'expansion basées sur la généralisation de la méthode de Rocchio (Rocchio, 1971) et de la pondération DFR (voir la section 3.6.2.5.1), comme le modèle Bo1 (Bose-Einstein 1)

(Plachouras *et al.*, 2004). Pour rappel, Rocchio propose une réinjection de la pertinence en utilisant les termes les plus informatifs dans les documents retournés et d'utiliser cette information de rétroaction pour étendre la requête. Pour le modèle de Bose-Einstein 1 (Bo1), le mécanisme d'extension consiste à extraire les termes les plus informatifs par le calcul de la divergence de leur distribution dans les documents les mieux classés et la distribution aléatoire de ces termes dans tous les documents retournés. Ces termes seront sélectionnés et ajoutés à la requête initiale suivant la formule (Pal *et al.*, 2013) :

$$w(t) = t f x . \log_2 \frac{1 + Pn(t)}{Pn(t)} + \log_2(1 + Pn[t]) \text{ avec } Pn(t) = \frac{t f c}{N}$$

Où  $w(t)$  est le poids ou la valeur informative du terme (t).  $Pn(t)$  est la probabilité de t dans toute la collection,  $t f x$  est la fréquence du terme dans les documents les mieux classés,  $\log_2(1 + Pn[t])$  est la fréquence de terme ( ) dans la collection, N est le nombre de documents de la collection. Dans ce travail de recherche et en plus de l'expansion de requête à partir des concepts retrouvés par l'algorithme de l'outil ICPU ou les concepts issus des catégories PICO, nous avons utilisé un des modèles d'expansion qui sont offerts par la plate-forme « Terrier » pour un besoin comparatif et de renforcement.

### 3.10 Expérimentation et évaluation

#### 3.10.1 Objectif

L'intérêt principal de ces expérimentations est de pouvoir répondre aux questions de recherche concernant l'extension de requêtes en utilisant un processus stochastique sur une termino-ontologie médicale et l'apport des concepts catégorisés comme appartenant au modèle PICO de la recherche clinique. Pour atteindre ces objectifs, nous avons choisi d'utiliser des corpus de test connus pour confronter nos hypothèses à des sources de jugements fiables et reconnues. Ceci explique l'utilisation du corpus test de la TREC pour la médecine de précision. Ce concours est une compétition qui regroupe différentes équipes de recherche et des groupes universitaires qui travaillent à l'amélioration de la recherche d'information dans la médecine de précision en oncologie. Le but étant de permettre aux médecins de retrouver des articles qui concernent un sujet donné d'une façon optimale (Roberts *et al.*, 2017).

Nous utilisons ces ressources pour deux raisons principales : 1) l'outil ICPU peut être utilisé d'une façon universelle et il n'est pas limité au seul domaine de la transfusion ou uniquement aux règles d'un SADC particulier ; 2) la conférence TREC fournit un corpus de test incluant des sujets (topics), ainsi que les fichiers de jugement « à posteriori » pour pouvoir faire la validation de la pertinence des résultats de recherche d'information de chaque sujet étendu par l'ICPU et/ou en ajoutant des concepts PICO.

Notre travail s'inscrit dans une démarche non compétitive dans la mesure où nous ne disposant pas des mêmes ressources que les autres équipes et que ce domaine précis reste très délicat et demande une expertise en oncologie (Tableau 3.6). Par exemple, l'équipe de UD\_GU\_BioTM (Université de Georgetown et Delaware — États-Unis (Mahmood *et al.*, 2017)) utilise pour l'extension de la requête en utilisant une base de données spécifique pour l'extension des variations génétiques (nommé : Entrez) et une ontologie spécifique pour extension des différentes maladies (*Disease Ontology*). En

plus, des oncologistes ont établi des règles d'associations entre les anomalies et traitement du cancer. Après l'extension de la requête, un traitement post-recherche d'information a été effectué avec d'autres outils pour marquer les maladies, les médicaments et les gènes dans les articles retrouvés et de détecter les possibles mutations concernées, et ce, en utilisant un outil développé en interne (BioNex). Cet exemple illustre la somme de ressources et d'interventions de spécialistes nécessaires pour relever le défi pour l'amélioration de la recherche d'articles en médecine de précision et en particulier en oncologie.

En effet et selon Roberts *et al.* (2017), l'un des facteurs qui fait de la médecine de précision un domaine difficile pour la RI est que les différents cas de patients nécessitent différents types de flexibilité. Pour certains patients, le type exact de cancer n'est pas pertinent alors que les données démographiques du patient ou d'autres facteurs peuvent peser plus lourdement. Le concept même de médecine de précision reconnaît le caractère unique du patient, et il faut donc s'attendre à ce qu'aucune correspondance parfaite ne soit trouvée. Non seulement les sujets fournis aux participants ne contiennent pas les informations nécessaires pour décider quels facteurs sont plus pertinents (par exemple, les traitements antérieurs du patient), mais à bien des égards, il n'est pas réaliste d'attribuer cette responsabilité au système IR. Néanmoins nous avons utilisé nos résultats pour nous positionner et nous comparer aux autres équipes (voir section 3.12). Ceci nous permettra de démontrer le caractère universel de notre approche d'extension indépendamment du domaine médical d'application. De plus, ceci pourrait bien initier une participation à ces compétitions en utilisant possiblement un calibrage dédié spécifiquement à l'oncologie et éventuellement une collaboration avec des experts humains. Notre approche pourrait permettre un avancement dans l'amélioration de la recherche d'information.

Tableau 3.6 Ressources et procédés utilisés par les équipes participantes à la TREC 2017.

Équipes	Extension de requête	Autres ressources utilisées	Spécialistes	Reclassement des articles
UKNLP (University of Kentucky Natural Language Processing) (Noh et Kavuluru, 2017)	Extension avec 5 concepts UMLS les plus proches du National Cancer Institute (NCI) ; extension avec 10-15 concepts proches de MeSH avec tous « atoms » (variations de termes utilisés à la différence du terme préféré) d'UMLS (MeSH and NCI) ; augmenter la requête suivant l'importance des sections (gène, âge,) dans le cas de patients ;	<i>Catalogue of Somatic Mutations in Cancer</i> (COSMIC) base de données pour le traitement des variations génétiques.	Intervention des experts en oncologie pour l'optimisation des facteurs pour pondérer les différentes parties de la requête.	Aucun traitement pour le reclassement
UD_GU_BioTM (Georgetown University, University of Delaware) (Mahmood et al., 2017)	« Gene Expansion » : méthode utilisant une base de données spécifique pour l'extension des variations génétiques. « Disease Expansion » : utilisation d'une ontologie spécifique pour extension des différentes maladies.	Disease Ontology : pour les variations de maladies oncologiques. « Entrez database » pour les variations génétiques.	Intervention des oncologistes pour l'établissement de règles d'extraction des relations entre anomalie et traitement du cancer.	<i>Pubtator</i> : un outil pour marquer les maladies, les médicaments et les gènes ; un détecteur de mutation développé en interne ; <i>BioNex</i> : Détecteur d'acronymes ; Approche d'associations entre les anomalies génomiques et la réponse à certains traitements contre le cancer
imi_mug Medical University of Graz (López-García et al., 2017)	« <i>query template</i> » : Utilisation d'un « <i>framework</i> » pour la modification des requêtes ; « <i>Query decorators</i> » : un autre procédé pour décortiquer les requêtes selon le gène et les maladies. Disease Expansion : extension de à partir de synonymes d'UMLS. Gène Expansion : utilisation d'une ontologie spécifique.	Homo Sapiens : Lits de gènes fournis par le « National Center for Biotechnology Information ». Standard de référence à partir de termes suggérés par un expert médical de l'équipe identifiés lors de l'inspection manuelle des résultats.	La pertinence des documents a été jugée par au moins deux médecins et un informaticien.	Aucun

### 3.10.2 Paramétrage pour l'identification des termes d'extension

Choisir un paramétrage pour l'identification des termes d'extension des requêtes n'est pas une tâche facile, car plusieurs paramètres entrent en jeu et le nombre de possibilités peut vite devenir exponentiel. Voici l'ensemble des paramètres à définir pour l'identification :

- 1- Choix de la relation :
- 2- Choix du seuil de co-occurrence lors de la sélection des voisins a visité.
- 3- Choix de la longueur de la marche
- 4- Choix du nombre de marches.
- 5- Choix du seuil de visite à partir duquel le concept est sélectionné.

Dès le début de nos expérimentations, nous avons constaté que le nombre de possibilités de paramétrage est important. De plus, le fait de modifier un seul des paramètres cités en haut engendre énormément de temps de calcul. Pour ne pas nous perdre dans l'immensité des paramétrages possibles, nous avons combiné les connaissances obtenues dans le chapitre précédent et l'évaluation unitaire d'un sujet à faible restitution d'information. En effet, Roberts *et al.* (2017) ont constaté que certains sujets dans la liste des « sujets » du TREC 2017 présentent un R-précision ou une MAP très basse comme le sujet numéro (13), voir le tableau 3.7. Ces « sujets » ne permettaient pas de retrouver les articles pertinents justes en utilisant les termes initiaux. De ce fait, la MAP sur la trentaine de sujets est lourdement pénalisée. Ce constat a été déjà fait par Amati, Giambattista *et al.* (2004) sur d'autres concours de la TREC, ils indiquent que l'extension de requête dégrade toujours les performances du système de RI sur les requêtes les moins performantes. Les raisons de ce cette recherche infructueuse n'a pas été évoquée. Néanmoins, cela pourrait être due à la rareté des articles qui traite de ce sujet dans les collections sur les 1000 premiers retournés ou que les termes utilisés ne suffisent pas de retrouver les bons articles.

Tableau 3.7 Résultats de la RI avec les termes initiaux des sujets (modèle DPH, Pub-Med)

Q	(Sujets de base)			
	MAP	Rprec	P@5	P@10
1	0,213 2	0,322 6	0,600 0	0,400 0
2	0,139 3	0,243 8	0,800 0	0,700 0
3	0,105 8	0,156 9	0,200 0	0,200 0
4	0,050 1	0,172 3	0,200 0	0,300 0
5	0,016 6	0,044 4	0,000 0	0,000 0
6	0,202 6	0,320 3	0,800 0	0,500 0
7	0,049 4	0,190 8	0,200 0	0,200 0
8	0,105 3	0,225 0	0,400 0	0,300 0
9	0,181 9	0,295 9	1,000 0	0,700 0
10	0,063 4	0,164 9	0,200 0	0,100 0
11	0,044 7	0,000 0	0,000 0	0,000 0
12	0,129 4	0,190 9	0,600 0	0,600 0
13	<b>0,000 6</b>	<b>0,000 0</b>	<b>0,000 0</b>	<b>0,000 0</b>
14	0,012 6	0,000 0	0,000 0	0,000 0
15	0,038 9	0,100 0	0,200 0	0,100 0
16	0,136 2	0,232 4	0,800 0	0,800 0
17	0,101 7	0,206 9	0,000 0	0,100 0
18	0,078 2	0,170 1	0,400 0	0,500 0
19	0,023 6	0,085 7	0,000 0	0,100 0
20	0,072 4	0,142 9	0,200 0	0,200 0
21	0,196 1	0,333 3	0,600 0	0,600 0
22	0,032 4	0,105 6	0,400 0	0,300 0
23	0,096 6	0,205 1	0,400 0	0,500 0
24	0,025 5	0,078 1	0,400 0	0,300 0
25	0,085 5	0,148 1	0,400 0	0,500 0
26	0,014 5	0,050 0	0,200 0	0,100 0
27	0,011 9	0,078 9	0,000 0	0,000 0
28	0,002 5	0,018 2	0,000 0	0,000 0
29	0,078 9	0,142 9	0,400 0	0,300 0
30	0,048 1	0,115 6	0,400 0	0,400 0
Moyenne	0,078 6			

À partir de ce constat, nous avons pris le sujet numéro « 13 » pour ajuster à titre indicatif et prospectif le choix des différents paramètres d'ICPU, car tester l'ensemble des configurations possibles serait prohibitif. Ceci nous a permis de statuer sur un paramétrage qui favorise l'apport positif du système sans dégradations majeures des résultats



du tableau 3.8. Les résultats du tableau ci-dessous expriment une valeur MAP pour extension des deux sujets sans pondération.

Tableau 3.8 Évaluation de la MAP des paramètres pour le sujet 13 et sujet 1

Extension ICPU avec relation RO, LM = 6 et co-occurrence =6			Sujet 13	Sujet 1
Vocabulaire	NM	SV	MAP	MAP
	30	10	0,000 7	0,001 1
	<b>30</b>	<b>12</b>	<b>0,001 2</b>	<b>0,144 7</b>
	30	14	0,003 7	0,137 2
Uniquement des termes initiaux			0,000 6	0,213 2

Pour la suite des expérimentations, nous avons choisi les paramètres suivants :

- 1- Relation : RO (associative) et CHD, PAR, RB, RN (hiérarchique)
- 2- Seuil de co-occurrence lors de la sélection des voisins a visité (=6)
- 3- Longueur de la marche (= 6)
- 4- Nombre de marche (= 30)
- 5- Seuil de visite à partir duquel le concept est sélectionné (=12)

### 3.10.3 Augmenter le poids des concepts d'extension ICPU

En réponse à la question de savoir si tous les termes de la requête étendus doivent avoir une pondération identique ou si les nouveaux termes devraient avoir une pondération plus élevée ou plus faible que les originaux, Voorhees (1994) a constaté que l'attribution de poids moindre aux concepts permet d'améliorer la précision de la récupération documentaire. Voorhees (1994) a utilisé un facteur compris entre 0 et 1 pour pondérer les termes ajoutés.

Pour la pondération du poids concepts qui vont servir à nos expérimentations, nous suivrons l'approche inspirée de l'étude de Cieslewicz *et al.* (2018) mais avec un changement majeur. Dans leur étude, Cieslewicz *et al.* (2018) proposent d'attribuer des poids différents aux mots dans une requête, selon que les mots appartiennent à la requête d'origine ou ils ont été ajoutés par le processus d'extension, ceci permet d'améliorer considérablement le résultat de leur extension. En effet, ils attribuent un poids de 100 aux termes de la requête d'origine, les termes obtenus de leur méthode d'extension à partir de PubMed seront augmentés par un facteur de 20 et par 1 si les termes proviennent de leur collection BioCADDIE (car selon eux très peu de termes). Cette pondération a augmenté positivement les résultats de recherche d'information des requêtes malgré l'aspect arbitraire du choix des valeurs, cette approche reprend les mêmes orientations proposées par Voorhees un peu plus en haut.

Nous avons repris le même principe, mais avec une attribution de poids sur l'information relative à la centralité des concepts définie par la fréquence de visite de chaque concept sélectionné pour l'extension par l'outil ICPU. Plus le concept est visité plus ce dernier est important. Les termes originaux seront distingués des termes d'expansion par la multiplication par un facteur de 100.

Notre pondération permet d'augmenter la puissance de chaque terme contenu dans les concepts retrouvés et candidats à l'extension de la requête, par une normalisation de leurs fréquences de visite par rapport à la somme globale des fréquences du groupe de concepts. Ce poids sera arrondi à la borne supérieure et multipliée par 100 suivant la formule ci-dessous :

*Poids du terme d'extention*

$$= \text{arrondi} \left( \frac{\text{fréquence de visite du concept}}{\sum \text{fréquences des concepts retrouvés}} \times 100 \right)$$

Exemple :

1) Termes originaux

*Breast ; cancer ; NRAS ; Amplification*

2) Concepts retrouvés avec leurs fréquences de visite respectives :

*Breast Carcinoma (40); Malignant neoplasm of breast (49); Woman (14); Gene Amplification Technique (31).*

Avec un total des fréquences de tous les concepts égalant 134

Cela donne après pondération et normalisation :

*Breast<sup>30</sup> Carcinoma<sup>30</sup> Malignant<sup>37</sup> neoplasm<sup>37</sup> breast<sup>37</sup>; Woman<sup>11</sup>; Gene<sup>24</sup> Amplifications<sup>24</sup> techniques<sup>24</sup>.*

Nous avons pondéré les termes originaux par un facteur de 100 et la requête finale sera de la sorte :

*Breast<sup>100</sup> cancer<sup>100</sup> NRAS<sup>100</sup> Amplification<sup>100</sup> Breast<sup>30</sup> Carcinoma<sup>30</sup> Malignant<sup>37</sup> neoplasm<sup>37</sup> breast<sup>37</sup>; Woman<sup>11</sup>; Gene<sup>24</sup> Amplification<sup>24</sup> Technique<sup>24</sup>*

*Note important: les termes peuvent apparaitre dans de multiples concepts et les termes originaux de la requête. Ceci implique que la multiplication par un facteur de pondé-*

*ration sur un terme qui se répète se traduirait finalement par un cumulatif qui augmenterait la puissance représentative de ce terme. Ceci est l'approche adoptée dans le cadre de nos expérimentations. Toutefois, il serait possible de décider autrement et d'exclure la répression d'un terme et ne garder par exemple que le plus grand facteur de pondération ou la moyenne des pondérations pour ce terme. Aussi, pour l'extension nous avons pris l'ensemble des concepts issu d'ICPU.*

#### 3.10.4 Augmenter le poids des concepts d'extension orientés PICO.

Comme déjà mentionnée dans la méthodologie, l'idée serait de pouvoir examiner séparément l'impact de l'ajout des concepts appartenant uniquement aux éléments PICO à l'extension de la requête au lieu de rajouter l'ensemble des concepts retrouvés à partir d'UMLS. À noter que les concepts appartenant aux différentes catégories PICO, mais ne proviennent pas du processus d'extraction ICPU (voir la section 3.8 : méthodologie). Cela signifie que la pondération envisagée ne peut se baser sur la fréquence de passage sur les concepts retournés par ICPU. Par conséquent, nous avons utilisé le rapport entre la fréquence des concepts dans les études par rapports la somme des fréquences autres concepts dans la même classe.

*Poids du terme d'extention*

$$= \text{arrondi} \left( \frac{\text{fréquence du concept dans les études cliniques}}{\sum \text{fréquences des concepts retrouvés dans les études cliniques}} \times 100 \right)$$

Note : nous avons limité le nombre maximal des termes d'extension aux 50 premiers concepts, car le nombre de concepts retournés pourrait être beaucoup plus important. Définir le nombre optimal de termes à ajouter pour une extension de requêtes n'est pas unique et peut être très variable d'un système à l'autre et pour les sujets eux-mêmes dans un même système, selon Ogilvie *et al.* (2009). Certains systèmes peuvent être efficaces a avec dix termes ou moins et d'autres seront plus efficaces avec un nombre plus important. Pour notre étude, nous avons fixé ce nombre a un maximum de

50 termes pour les orientations PICO. Toutefois, ce nombre est exploratoire et pourrait bien faire l'objet d'une optimisation future.

#### 3.10.5 Différentes variantes pour l'extension des requêtes

En ce qui concerne la formulation des requêtes (15), quinze types de requêtes sont soumises au système expérimental. La première des requêtes (Qinit) sert de référence et les autres reflètent les différentes stratégies d'extension (voir les tableau 3.9, tableau 3.10 et tableau 3.11).

Tableau 3.9 Les différentes variantes d'extension de requêtes utilisées pour l'expérimentation.

	Ensemble d'extension	Description	Pondération	Termes d'extension	Objectif
1	Qinit	La dénomination « Qinit » est utilisée pour désigner la requête sous sa forme initiale. Les requêtes de cet ensemble (qui sont au nombre de 30) sont composées uniquement des termes initiaux du « sujet ».	n. a	n. a	Les mesures sur cet ensemble permettront d'établir une référence pour vérifier les améliorations/dégradation de la RI pour tous les ensembles subséquents.
2	Qinit + Meta-MapLite	Dans cet ensemble, de nouveaux termes issus du traitement « Mapping » avec MetaMapLite (voir section 2.9.2 du chapitre II) seront ajoutés aux termes initiaux de chaque requête comme dans le texte ci-dessous :	Sans pondération pour les termes ajoutés	MetaMapLite	Cet ensemble nous permettra de distinguer l'effet MetaMapLite des autres extensions subséquentes.
3	Qinit + ICPU (p)	Cet ensemble désigne les requêtes initiales augmentées par l'ajout des résultats d'extraction du système ICPU	Pondération des termes ajoutés selon la section 3.10.3	ICPU	Cet ensemble permet de vérifier l'effet des nouveaux termes pondérés provenant de l'extraction par la marche aléatoire paramétrée.
4	Qinit + ICPU (sp)	Cet ensemble est identique au précédent.	Sans pondération pour les nouveaux termes.	ICPU	Cet ensemble permet de vérifier l'effet des nouveaux termes d'ICPU sans pondération.
5	Qinit + ICPU (p) + PA (sp)	Ensemble d'extension avec ICPU et autres problèmes/conditions médicales fréquemment utilisés dans les études cliniques qui contiennent les concepts ICPU dans leur partie « problèmes ».	Sans pondération pour les nouveaux termes	ICPU + concepts PA retrouvés dans TrialPico	Évaluer l'effet d'une combinaison ICPU et autre problème/population et conditions de santé. (Plus fréquents dans les études cliniques pour cette catégorie de sujets).

Tableau 3.10 (Suite) Les différentes variantes d'extension de requêtes utilisées pour l'expérimentation.

	Ensemble d'extension	Description	Pondération	Termes d'extension	Objectif
6	Qinit + ICPU (p) + I (sp)	Ensemble d'extension avec ICPU et autres interventions fréquemment utilisés dans les études cliniques qui contiennent les concepts ICPU dans leur partie « problèmes ».	Sans pondération pour les nouveaux termes.	ICPU + concepts I retrouvés dans TrialPico	
7	Qinit + ICPU (p) + O (sp)	Ensemble d'extension avec ICPU et autres outcome fréquemment utilisés dans les études cliniques qui contiennent les concepts ICPU dans leur partie « problèmes ».	Sans pondération pour les nouveaux termes.	ICPU + concepts O retrouvés dans TrialPico	
8	Qinit + P (p)	Dans cette configuration, les concepts utilisés pour cette extension sont uniquement des concepts retrouvés par ICPU et qui se retrouvent aussi dans les champs « problèmes/condition et population » des études de recherche clinique et médicale (voir la section 3.7.2.2 et la section 3.7.2.3). La pondération est celle utilisée avec les concepts ICPU.	Pondération des termes ajoutés selon la section 3.10.3	P	Le but étant de voir si un nombre minimal de concepts liés uniquement aux conditions médicales améliorerait la recherche d'information.
9	Qinit + P (sp)	Ensemble identique à Qinit + P	Sans pondération pour les nouveaux termes.	P	Le but étant de voir si un nombre minimal de concepts liés uniquement aux problèmes et conditions médicales améliorerait la recherche d'information. (Sans pondération dans ce cas).
10	Qinit + PA (p)	Cet ensemble reflète une autre possibilité qui explore l'ajouter de concepts qui appartiennent aux classes PICO « P », mais qui ne figurent pas dans le lot de concepts extraits par l'ICPU à partir des termes de la requête initiale (Section 3.8)	Pondération des nouveaux termes suivant la section 3.10.4	Autres problèmes	Permet d'évaluer l'effet d'autres conditions médicales qui peuvent exister dans des recherches cliniques semblables.

Tableau 3.11 (Suite) Les différentes variantes d'extension de requêtes utilisées pour l'expérimentation.

	Ensemble d'extension	Description	Pondération	Termes d'extension	Objectif
11	Qinit + PA (sp)	Identique à l'ensemble précédent	Sans pondération pour les nouveaux termes.	Autres problèmes	Permet d'évaluer l'effet d'autres conditions médicales qui peuvent exister dans des recherches cliniques semblables. (Sans pondération)
12	Qinit + I (p)	Cet ensemble se fait par l'ajout des concepts considéré comme appartenant à la catégorie « Intervention ».	Pondération suivant la section 3.10.4	Intervention	Permet d'évaluer l'apport des concepts « intervention » aux termes initiaux.
13	Qinit + I (sp)	Identique à l'ensemble précédent.	Sans pondération pour les nouveaux termes.	Intervention	Objectif identique au précédent, mais sans pondération.
14	Qinit + O (p)	Cet ensemble se fait par l'ajout des concepts considéré comme appartenant à la catégorie « Outcome ».	Pondération suivant la section 3.10.4	Outcome	Permet d'évaluer l'apport des concepts « outcome » aux termes initiaux.
15	Qinit + O (sp)	Identique à l'ensemble précédent.	Sans pondération pour les nouveaux termes.	Outcome	Objectif identique au précédent, mais sans pondération.

Note : l'ensemble des termes initiaux dans les différentes variantes seront augmentés par un facteur 100 pour les distinguer des termes ajoutés.

Remarque importante : Pour les P, PA, I et O, le nombre maximal de concepts ajoutés ne dépasse pas (50) cinquante concepts.



### 3.11 Résultats et discussion

Les expérimentations ont porté sur deux configurations du logiciel ICPU avec des paramètres identiques sauf pour le type de relations à exploiter :

- 1) Longueur de marche =6, nombre de marche = 30, seuil de visite = 12, seuil de cooccurrence=6 ; Rela : RO ; Vocabulaires : MTH, MeSH, SNOMED\_CT, NCI.
- 2) Longueur de marche =6, nombre de marche = 30, seuil de visite = 12, seuil de cooccurrence=6 ; Rela : CHD, PAR, RB, RN, Vocabulaires : MTH, MeSH, SNOMED\_CT, NCI.

Les concepts candidats issus de chaque configuration de base peuvent être différents et ainsi permettre d'étudier la différence entre le choix de prioriser les relations hiérarchiques versus les relations associatives d'UMLS.

Nous avons établi (15) différentes combinaisons d'extension pour les 30 sujets fournis par la TREC (section 3.10.5). Pour chaque combinaison, les algorithmes de recherche DPH, BM25, InL2 et PL2 (voir section 3.6.2.4) ont été utilisés pour établir les différentes mesures d'évaluation. Les mesures calculées pour chaque algorithme sont les suivantes : MAP, R-précision, P@5 et P@10 (voir section 3.6.2.3 pour plus d'information). Une extension supplémentaire avec le modèle de bo1 (voir section 3.9.8) a été effectuée sur les mêmes ensembles précédents ( Tableau 3.12, tableau 3.13, tableau 3.15 et tableau 3.16). Pour certains ensembles d'extension qui présentent des valeurs améliorées, des mesures plus détaillées ont été fournies pour chaque sujet - au nombre de 30 (Tableau 3.14 et tableau 3.17).

Remarque importante : Les multiples possibilités de paramétrage de base du système ICPU et les différentes combinaisons d'extensions possibles peuvent rapidement devenir prohibitives en termes de temps de traitement et d'évaluation. Dans ce contexte, des

choix ont été pour définir les différents ensembles d'expérimentation (voir section 3.10.5).

Dans le tableau 3.12, nous pouvons constater que l'effet d'une extension avec l'utilisation des concepts retrouvés par l'ICPU (Qinit + ICPU (p)) et une pondération selon notre approche réalise la plus grande MAP avec 0,1170 en utilisant l'algorithme InL2 (10,6 % par rapport à la requête de bases et des scores de 0,1841, 0,3867, 0,3533 pour R-précision, P@5, P@10 respectivement). Cette MAP représente le double en pourcentage du résultat MAP d'une requête qui utilise uniquement les concepts retrouvés par MetaMapLite (4,8 %) (combinaison n° 3 du tableau 3.12). Le même ensemble de requêtes étendu par la combinaison (Qinit + ICPU (p)) présente aussi les plus fortes augmentations de MAP par rapport à la requête de base avec les différents algorithmes de recherche (DHP, BM25 et PL2 avec 14,1 %, 10,4 %, 11,5 % respectivement). Il est à noter aussi, que l'ajout de ces nouveaux concepts, même sans pondération, à la requête de base (Qinit + ICPU (sp)) permet d'avoir des valeurs supérieures à ce que MetaMapLite ou la requête de base peuvent réaliser. Cette augmentation de la valeur MAP (constatées pour les différents modèles de recherche) permet d'affirmer que les concepts ICPU sont des concepts qui peuvent apporter de la pertinence et que l'approche de pondération basée sur « la centralité » des concepts apporte une stabilité, quel que soit le modèle utilisé dans cette expérimentation. Pour rappel, l'approche permet d'augmenter le poids des concepts les plus fréquemment visités et aussi d'éliminer les concepts les plus faiblement visités si la normalisation du rapport de fréquence est proche du nul.

Par ailleurs, l'effet d'augmentation de la MAP avec l'ajout des concepts issus des différentes catégories PICO à la requête initiale (sans ICPU) reste modéré et se situe aux alentours de 9,2 % (sans pondération) et l'impact est plus faible (entre 0,6 et 7 %) avec la pondération proposée (requêtes 7, 8, 10, 11, 13 et 14 du tableau 3.12). Ceci suggère

que ces concepts, ainsi ajoutés, rapportent de l'information utile, mais que la pondération peut ne pas être efficace, car les concepts sans pondération produisent des scores plus élevés.

Pour rappel, la pondération proposée pour ces ensembles repose sur la fréquence des concepts appartenant aux catégories « intervention » ou « outcome » dans les études cliniques. Ces études sont sélectionnées à partir du moment où elles contiennent les concepts initiaux retrouvés par ICPU dans leurs sections problèmes et condition médicales (voir la section 3.8).

Ce faible rendement des extensions orienté PICO (avec ou sans pondération) signifie que les concepts les plus fréquents utilisés dans les catégories « intervention » et « outcome » des recherches cliniques ne suffisent pas à discriminer positivement les articles scientifiques pertinents pour les sujets recherchés.

Les expérimentations (5 et 6 du tableau 3.12), avec un ensemble minimal de concepts pertinents appartenant à la catégorie « problème et population », donnent des résultats modérés (9 %) sans pondération et faible avec la pondération. Cela signifie que les concepts ainsi filtrés ne peuvent pas à eux seuls améliorer les performances de recherche et que d'autres concepts qui ne se retrouvent pas forcément dans la catégorie des problèmes et populations sont nécessaires (comme des concepts décrivant l'âge des patients, Adult, enfant, le sexe...).

Les résultats pour les ensembles de combinaisons 9,12 et 15, qui représentent l'ajout aux requêtes déjà étendues par ICPU des concepts appartenant aux différentes catégories PICO (sans pondération), sont mitigés dans le sens où ils dégradent les résultats d'ICPU. Les concepts « intervention » et de « outcome » ne permettent pas d'augmenter les performances de recherche et ceci est due probablement aux mêmes phénomènes qu'avec les « termes » très fréquents décrits plus haut pour les ensembles 7, 8, 10, 11, 13 et 14 du tableau 3.12.

Tableau 3.12 Résultats d'évaluation des différents jeux d'extension de requêtes avec un paramétrage d'ICPU suivant : Longueur de marche =6, nombre de marche = 30, seuil de visite = 12, seuil de cooccurrence=6 ; Rela : RO Vocabulaires : MTH, MeSH, SNOMED\_CT, NCI.

Modèles		DPH					BM25					InL2					PL2				
		MAP	Δ	Rprec	P@5	P@10	MAP	Δ	Rprec	P@5	P@10	MAP	Δ	Rprec	P@5	P@10	MAP	Δ	Rprec	P@5	P@10
1	Qinit	0,0835		0,1530	0,3133	0,2933	0,1054		0,1741	0,3533	0,3167	0,1058		0,1755	0,3467	0,3200	0,0930		0,1696	0,3267	0,3200
2	Qinit + Meta-MapLite	0,0908	8,7 %	0,1578	0,3467	0,3233	0,1099	4,3 %	0,1769	0,3667	0,3467	0,1109	4,8 %	0,1776	0,3533	0,3500	0,0972	4,5 %	0,1705	0,3533	0,3300
3	Qinit + ICPU (p)	0,0953	14,1 %	0,1632	0,3600	0,3267	0,1164	10,4 %	0,1841	0,3867	0,3533	<b>0,1170</b>	<b>10,6 %</b>	<b>0,1805</b>	<b>0,3733</b>	<b>0,3633</b>	0,1037	11,5 %	0,1775	0,3533	0,3500
4	Qinit + ICPU (sp)	0,0928	11,1 %	0,1604	0,3667	0,3233	0,1116	5,9 %	0,1796	0,3867	0,3600	0,1131	6,9 %	0,1783	0,3733	0,3567	0,0992	6,7 %	0,1723	0,3467	0,3400
5	Qinit + P (p)	0,0891	6,7 %	0,1571	0,3467	0,3067	0,108	2,5 %	0,1779	0,3533	0,3433	0,1083	2,4 %	0,1761	0,3400	0,3433	0,0952	2,4 %	0,169	0,3333	0,3200
6	Qinit + P (sp)	0,0910	9,0 %	0,1599	0,3467	0,3200	0,1100	4,4 %	0,1783	0,3733	0,3500	0,1111	5,0 %	0,1774	0,3600	0,3533	0,0969	4,2 %	0,17	0,3467	0,3333
7	Qinit + PA (p)	0,084	0,6 %	0,1475	0,3333	0,3	0,1031	-2,2 %	0,1685	0,3467	0,3233	0,1045	-1,2 %	0,1657	0,3400	0,3300	0,0909	-2,3 %	0,16	0,3400	0,3100
8	Qinit + PA (sp)	0,0912	9,2 %	0,1613	0,3533	0,3233	0,1102	4,6 %	0,178	0,3733	0,3467	0,1114	5,3 %	0,1807	0,3533	0,3500	0,0974	4,7 %	0,1701	0,3400	0,3333
9	Qinit + ICPU (p) + PA (sp)	0,0943	12,9 %	0,1649	0,3533	0,3333	0,1154	9,5 %	0,1845	0,3733	0,3600	0,1164	10,0 %	0,1824	0,3733	0,3567	0,1024	10,1 %	0,1791	0,3533	0,3433
10	Qinit + I (p)	0,0910	9,0 %	0,1628	0,3333	0,3133	0,1101	4,5 %	0,1802	0,3533	0,3400	0,1113	5,2 %	0,1806	0,3467	0,3433	0,0977	5,1 %	0,1705	0,3400	0,3367
11	Qinit + I (sp)	0,0912	9,2 %	0,1613	0,3533	0,3233	0,1102	4,6 %	0,1780	0,3733	0,3467	0,1114	5,3 %	0,1807	0,3533	0,3500	0,0974	4,7 %	0,1701	0,3400	0,3333
12	Qinit + ICPU (p) + I (sp)	0,0953	14,1 %	0,1633	0,3600	0,3267	0,1164	10,4 %	0,1838	0,3867	0,3567	0,1170	10,6 %	0,1806	0,3733	0,3600	0,1037	11,5 %	0,1763	0,3533	0,3500
13	Qinit + O (p)	0,0902	8,0 %	0,1528	0,3733	0,3167	0,1089	3,3 %	0,1773	0,3800	0,3600	0,1100	4,0 %	0,1750	0,3867	0,3467	0,0965	3,8 %	0,1708	0,3800	0,3167
14	Qinit + O (sp)	0,0912	9,2 %	0,1587	0,3667	0,3200	0,1106	4,9 %	0,1783	0,3733	0,3567	0,1114	5,3 %	0,1778	0,3733	0,3500	0,0972	4,5 %	0,1710	0,3533	0,3300
15	Qinit + ICPU (p) + O (sp)	0,094	12,6 %	0,1623	0,3600	0,3233	0,1152	9,3 %	0,1815	0,4000	0,3567	0,1162	9,8 %	0,1799	0,3667	0,3600	0,1005	8,1 %	0,1738	0,3533	0,3367

Toutefois, dans le tableau 3.13, qui concerne l'expérimentation supplémentaire avec l'extension de requête « bo1 » de Terrier sur les ensembles déjà étendus par nos différentes méthodes d'extension, les résultats sont encore plus intéressants et importants sachant que l'amélioration par bo1 pour une requête de base est faible (3,8 % MAP pour le DHP). Les valeurs intéressantes sont celles de (Qinit + ICPU (p) +PA (sp) + bo1) et (Qinit +ICPU (p) +O (sp) + bo1) qui donnent les scores de MAP de 0,1181 et de 0,1182 en InL2 avec respectivement 11.6 % et 11.7 % d'augmentation par rapport à la requête de base. Ces deux valeurs peuvent être des indicateurs concernant le potentiel de l'ajout des concepts issus des catégories PICO, notamment les PA et les O. Ces concepts qui n'existaient ni dans la requête initiale, ni dans les concepts ICPU, permettent de ramener parmi les trois premières positions de classement des articles à partir desquels, dix termes informatifs seront extraits et utilisés pour d'améliorer la recherche (voir la section 3.9.8).

En constat général, les valeurs obtenues avec tous les ensembles utilisant les concepts ICPU avec pondération se situent entre 10,8 et 11,7 % d'amélioration de la MAP (InL2). Ceci confirme le résultat obtenu dans le tableau 3.12. Toutefois, les scores des ensembles (5, 6, 7, 8, 10, 11, 13 et 14) incluant uniquement des concepts « autres problèmes », « intervention » et « outcome » avec la requête initiale présentent des valeurs de MAP moins significatives, mais qui restent positives et notamment pour ceux ajoutés sans pondération.

Aussi, il est à noter que MetaMapLite réalise un score intéressant pour ce qui est PL2 avec 18,2 %, toutefois, il reste en deçà du 20,0 % de la requête 9 et les scores pour les autres algorithmes restent modérés.

Tableau 3.13 Résultats d'évaluation des différents ensembles d'extension, Rela : RO, avec extension supplémentaire de « bo1 ».

	Modèle	DPH					BM25					InL2					PL2								
		Requêtes	MAP	Δ	Rprec	P@5	P@10	MAP	Δ	Rprec	P@5	P@10	MAP	Δ	Rprec	P@5	P@10	MAP	Δ	Rprec	P@5	P@10			
	Qinit	0,0835			0,1530	0,3133	0,2933	0,1054			0,1741	0,3533	0,3167	0,1058			0,1755	0,3467	0,3200	0,0930			0,1696	0,3267	0,3200
1	Qinit + bo1	0,0867	3,8 %		0,1512	0,2867	0,28	0,1135	7,7 %		0,183 5	0,360 0	0,370 0	0,1014	-4,2 %		0,166 5	0,306 7	0,316 7	0,0971	4,4 %		0,1667	0,3067	0,3000
2	Qinit + Meta-MapLite + Bo1	0,0966	15,7 %		0,1611	0,3467	0,3067	0,1149	9,0 %		0,1813	0,36	0,3700	0,1174	11,0 %		0,1845	0,3867	0,3733	0,1099	18,2 %		0,178	0,3467	0,3467
3	Qinit + ICPU (p) + bo1	0,0988	18,3 %		0,1688	0,3533	0,3233	0,1164	10,4 %		0,1853	0,3733	0,3700	0,1172	10,8 %		0,1870	0,3667	0,3700	0,1075	15,6 %		0,1769	0,3467	0,3500
4	Qinit + ICPU (sp) + bo1	0,098	17,4 %		0,1659	0,3467	0,32	0,1141	8,3 %		0,182	0,3533	0,3633	0,1167	10,3 %		0,1845	0,3800	0,3733	0,1093	17,5 %		0,1797	0,3533	0,3500
5	Qinit + P (p) + Bo1	0,0909	8,9 %		0,1580	0,3533	0,31	0,1165	10,5 %		0,1803	0,3933	0,3633	0,1042	-1,5%		0,1681	0,3600	0,3433	0,1098	18,1 %		0,1803	0,3333	0,3467
6	Qinit + P (sp) + Bo1	0,0971	16,3 %		0,1618	0,3467	0,3133	0,1149	9,0 %		0,1813	0,38	0,3700	0,1167	10,3%		0,1827	0,3867	0,3700	0,1065	14,5 %		0,1735	0,3267	0,3333
7	Qinit + PA (p) + Bo1	0,0877	5,0 %		0,1446	0,3133	0,2833	0,1085	2,9%		0,1678	0,3533	0,3400	0,1097	3,7%		0,1691	0,3800	0,3400	0,1030	10,8 %		0,1594	0,3333	0,3033
8	Qinit + PA (sp) + Bo1	0,0969	16,0%		0,165	0,34	0,32	0,1154	9,5 %		0,1814	0,3733	0,3733	0,1172	10,8%		0,1838	0,3933	0,37	0,1075	15,6 %		0,1722	0,34	0,0969
9	Qinit + ICPU (p) + PA (sp) + bo1	0,0974	16,6 %		0,1657	0,36	0,3133	<b>0,1179</b>	<b>11,9 %</b>		<b>0,1882</b>	<b>0,3933</b>	<b>0,3733</b>	<b>0,1181</b>	<b>11,6 %</b>		<b>0,1879</b>	<b>0,3800</b>	<b>0,3700</b>	0,1116	20,0 %		0,180 0	0,353 3	0,350 0
10	Qinit + I (p) + bo1	0,0866	3,7 %		0,1520	0,3200	0,2800	0,1169	10,9 %		0,1830	0,3600	0,3700	0,116	9,6 %		0,1836	0,3933	0,3567	0,1076	15,7 %		0,1741	0,3400	0,3367
11	Qinit + I (sp) + bo1	0,0969	16,0%		0,1650	0,3400	0,3200	0,1154	9,5 %		0,1814	0,3733	0,3733	0,1172	10,8 %		0,1838	0,3933	0,3700	0,1075	15,6 %		0,1722	0,3400	0,3367
12	Qinit + ICPU (p) + I (sp) + bo1	0,0985	18,0 %		0,168 8	0,36	0,3167	0,1169	10,9 %		0,186	0,3867	0,37	0,1172	10,8 %		0,1874	0,36	0,3567	0,1077	15,8 %		0,1769	0,3467	0,3433
13	Qinit + O (p) + bo1	0,0962	15,2 %		0,159 8	0,360 0	0,316 7	0,1162	10,2 %		0,186 0	0,380 0	0,356 7	0,1146	8,3 %		0,184 4	0,380 0	0,363 3	0,1066	14,6 %		0,172 4	0,340 0	0,333 3
14	Qinit + O (sp) + bo1	0,0957	14,6 %		0,161 5	0,340 0	0,316 7	0,1161	10,2 %		0,182 4	0,353 3	0,366 7	0,1158	9,5 %		0,184 0	0,380 0	0,370 0	0,1069	14,9 %		0,171 9	0,326 7	0,336 7
15	Qinit + ICPU (p) + O (sp) + bo1	0,0989	18,4 %		0,169 2	0,353 3	0,320 0	0,1155	9,6 %		0,180 5	0,373 3	0,370 0	<b>0,1182</b>	<b>11,7 %</b>		<b>0,186 3</b>	<b>0,373 3</b>	<b>0,366 7</b>	0,1087	16,9 %		0,177 2	0,346 7	0,350 0

Le tableau 3.14 (ci-dessous) présente les résultats pour chaque sujet dans la combinaison « Qinit + ICPU (p) ». Il en ressort un effet significatif de l'extension pour certains sujets qui présentaient de très faibles scores MAP au départ dans la requête initiale. Exemple de la requête 13 avec une MAP de départ de 0,0008 et qui a fait un bond de 113 % avec l'extension ICPU pondérée.

Tableau 3.14 Mesures détaillées et impact sur les « sujets » pour « Qinit + ICPU (p) » avec rela = RO et InL2

N Req.	Qinit (termes de base)				Qinit+ICPU (p)				
	MAP	R précision	P@5	P@10	MAP	$\Delta$	R précision	P@5	P@10
1	0,2259	0,3387	0,6000	0,4000	0,2123	-6%	0,3387	0,6000	0,4000
2	0,1874	0,2576	0,8000	0,9000	0,1888	1%	0,2659	0,8000	0,9000
3	0,1299	0,2157	0,2000	0,1000	0,1835	<b>41%</b>	0,1765	0,4000	0,5000
4	0,0941	0,2360	0,2000	0,1000	0,1003	7%	0,2434	0,0000	0,1000
5	0,0246	0,0667	0,0000	0,0000	0,0273	11%	0,0667	0,0000	0,0000
6	0,2610	0,3750	0,6000	0,6000	0,2674	2%	0,3906	0,6000	0,7000
7	0,0699	0,1763	0,2000	0,1000	0,0767	10%	0,1763	0,4000	0,4000
8	0,1224	0,1833	0,2000	0,3000	0,1228	0%	0,1833	0,4000	0,3000
9	0,2395	0,3735	1,0000	0,9000	0,2404	0%	0,3673	1,0000	0,9000
10	0,0996	0,1959	0,2000	0,2000	0,1105	11%	0,2062	0,2000	0,2000
11	0,0661	0,0238	0,0000	0,0000	0,0644	-3%	0,0238	0,0000	0,0000
12	0,1523	0,2091	1,0000	0,7000	0,1512	-1%	0,2091	0,8000	0,7000
13	0,0008	0,0000	0,0000	0,0000	0,0017	<b>113%</b>	0,0000	0,0000	0,0000
14	0,0116	0,0000	0,0000	0,0000	0,0131	13%	0,0000	0,0000	0,0000
15	0,0570	0,1000	0,2000	0,1000	0,1059	<b>86%</b>	0,1000	0,2000	0,1000
16	0,1922	0,2606	0,8000	0,9000	0,1721	-10%	0,2606	0,8000	0,8000
17	0,1319	0,2328	0,0000	0,2000	0,1403	6%	0,2500	0,0000	0,2000
18	0,1019	0,1959	0,4000	0,6000	0,1046	3%	0,2216	0,4000	0,5000
19	0,0567	0,0857	0,4000	0,2000	0,0606	7%	0,0857	0,4000	0,2000
20	0,1194	0,1429	0,2000	0,3000	0,1210	1%	0,1429	0,6000	0,3000
21	0,2167	0,3333	0,4000	0,4000	0,2383	10%	0,3480	0,4000	0,5000
22	0,0503	0,1127	0,6000	0,4000	0,0560	11%	0,1127	0,6000	0,4000
23	0,1915	0,3179	0,4000	0,5000	0,3486	<b>82%</b>	0,4205	0,6000	0,8000
24	0,0784	0,1875	0,4000	0,3000	0,0849	8%	0,2031	0,4000	0,4000
25	0,0841	0,2037	0,2000	0,3000	0,0904	7%	0,1852	0,2000	0,4000
26	0,0159	0,0500	0,2000	0,1000	0,0227	43%	0,0500	0,2000	0,1000
27	0,0233	0,0658	0,0000	0,1000	0,0235	1%	0,0789	0,0000	0,1000
28	0,0065	0,0182	0,2000	0,1000	0,0064	-2%	0,0182	0,2000	0,1000
29	0,0928	0,1429	0,6000	0,4000	0,1005	8%	0,1190	0,6000	0,5000
30	0,0708	0,1633	0,4000	0,4000	0,0746	5%	0,1701	0,4000	0,4000
Moyenne	0,1058	0,1755	0,3467	0,3200	0,1170	10,6%	0,1805	0,3733	0,3633

En général, 22 sujets ont vu leurs scores s'améliorer avec une moyenne de 19 %. Toutefois, une dégradation est constatée pour les requêtes restantes avec une moyenne de -4 %.

Ceci indique qu'en moyenne, l'ajout des nouveaux concepts avec un facteur de pondération adéquat a permis de garder et d'augmenter la qualité des résultats de recherche sur l'ensemble des différents sujets.

Pour les expérimentations avec une extraction ICPU guidée par les relations hiérarchiques (Rela : CHD, PAR, RB, RN) (Tableau 3.15), le constat est le même, l'effet d'une extension avec l'utilisation des concepts retrouvés par l'ICPU et pondérés est significatif et présente la plus haute valeur MAP avec 0,1166 (Tableau 3.15- ensemble n° 3). Ceci confirme que la combinaison ICPU et pondération suivant la centralité des concepts permet de maintenir une amélioration stable des résultats de recherche quels que soient les algorithmes de recherche utilisés.

Dans le même tableau 3.15, il est à noter que l'utilisation directe des concepts issue de l'ICPU sans pondération provoque une dégradation des résultats MAP de recherche documentaire importante avec une moyenne de -34.6 % de MAP tous algorithmes confondus. Ceci n'enlève rien à leur potentiel de pertinence, il est important de rappeler que les concepts ainsi ajoutés ont le même niveau d'importance et cela se traduit par un classement différent des documents retrouvés par les algorithmes de recherche.

Par ailleurs et comparativement aux scores ICPU sans pondération avec les relations associatives (RO), ce score négatif des relations hiérarchiques (CHD, PAR, RB, RN) peut s'expliquer par les possibilités suivantes : nous avons constaté que le nombre de concepts retournés par les relations hiérarchiques est toujours plus important que celui des relations associatives (ces dernières peuvent avoir moins de relations avec leur voisinage et du coup la marche aléatoire d'ICPU ne parcourt plus que de courts chemins) et ainsi produire un effet inverse en dégradant la requête. L'autre possibilité est liée à



la pertinence de certains concepts issus de relation hiérarchique et qui peuvent tout simplement pousser la recherche à s'éloigner des articles recherchés en produisant plus de synonymes que les relations associatives. Ceci rejoint le même constat de (Dong *et al.*, 2011) qui a observé une dégradation des résultats de recherche des requêtes étendue et qui incombe principalement à la généralisation des concepts de départs. La généralisation et la spécialisation sont le propre des relations hiérarchiques et un déséquilibre vers une plus grande généralisation peut être la cause de la dégradation.

Quant à l'effet des différentes catégories PICO sur l'extension (scénarios 5, 6, 7, 8, 10, 11, 13 et 14 du tableau 3.15), cela reste modéré pour les catégories de concepts (P) (PA) et (O), mais une dégradation est constatée pour le (I). Ces concepts ainsi filtrés et ajoutés par catégories restent incapables d'améliorer à eux seuls la requête initiale. Comme nous l'avons souligné précédemment, probablement l'accumulation des facteurs suivants, en a été la cause : la fréquence et l'ordre à la base du rapport de pondération choisis ainsi que le nombre de concepts ajoutés qui ne permettrait pas une discrimination positive des articles potentiellement intéressants. Néanmoins la piste PICO reste pertinente au moins pour les concepts issus de (PA) et (O).

En effet, l'utilisation supplémentaire de l'algorithme d'extension de requête *bo1* sur ces ensembles d'extension (ou les sujets ont déjà été étendus) donne des résultats très encourageants pour l'orientation PICO.

Dans le tableau 3.16, le résultat d'évaluation de l'ensemble « Qinit +ICPU (p) + PA (sp) +*bo1* » donne le meilleur score de toutes les expérimentations avec une MAP égale à 0,1211 (InL2). Tandis que l'ensemble « Qinit +ICPU (p) + O (sp) + *bo1* » donne le troisième meilleur score avec de 0,1181 (InL2) après celui de Qinit + ICPU (p) + *bo1* avec 0,1186 de MAP (BM25). Cela permet de dire que l'ajout d'autres concepts PICO comme ceux de (PA) ou (O) peuvent contribuer à améliorer les résultats de recherche et ouvrir des pistes pour amélioration dans ce sens.

Tableau 3.15 Résultats d'évaluation des 15 ensembles d'extension de requêtes avec un paramétrage d'ICPU suivant : Longueur de marche =6, nombre de marche = 30, seuil de visite = 12, seuil de cooccurrence=6 ; Rela : CHD, PAR, RB, RN, Vocabulaires : MTH, MeSH, SNOMED

Modèles		DPH					BM25					InL2					PL2				
Requêtes		MAP	Δ	Rprec	P@5	P@10	MAP	Δ	Rprec	P@5	P@10	MAP	Δ	Rprec	P@5	P@10	MAP	Δ	Rprec	P@5	P@10
1	Qinit	0,0835		0,1530	0,3133	0,2933	0,1054		0,1741	0,3533	0,3167	0,1058		0,1755	0,3467	0,3200	0,0930		0,1696	0,3267	0,3200
2	Qinit + Meta-MapLite	0,0908	8,7%	0,1578	0,3467	0,3233	0,1099	4,3%	0,1769	0,3667	0,3467	0,1109	4,8%	0,1776	0,3533	0,3500	0,0972	4,5%	0,1705	0,3533	0,3300
3	Qinit + ICPU (p)	<b>0,0948</b>	<b>13,5%</b>	<b>0,165 2</b>	<b>0,346 7</b>	<b>0,326 7</b>	<b>0,1154</b>	<b>9,5%</b>	<b>0,187 1</b>	<b>0,366 7</b>	<b>0,356 7</b>	<b>0,1166</b>	<b>10,2%</b>	<b>0,183 8</b>	<b>0,373 3</b>	<b>0,370 0</b>	<b>0,1018</b>	<b>9,5%</b>	<b>0,175 0</b>	<b>0,366 7</b>	<b>0,350 0</b>
4	Qinit + ICPU (sp)	0,0542	-35,1%	0,105 6	0,214 3	0,182 1	0,0695	-34,1%	0,124 1	0,242 9	0,235 7	0,0699	-33,9%	0,124 2	0,250 0	0,225 0	0,0602	-35,3%	0,117 5	0,207 1	0,228 6
5	Qinit + P (p)	0,0896	7,3%	0,155 6	0,373 3	0,323 3	0,1092	3,6%	0,179 6	0,373 3	0,343 3	0,1101	4,1%	0,177 4	0,366 7	0,353 3	0,0963	3,5%	0,166 4	0,353 3	0,333 3
6	Qinit + P (sp)	0,0905	8,4%	0,159 8	0,346 7	0,323 3	0,1095	3,9%	0,178 0	0,380 0	0,350 0	0,1108	4,7%	0,177 0	0,346 7	0,353 3	0,0966	3,9%	0,170 8	0,340 0	0,330 0
7	Qinit + PA (p)	0,0896	7,3%	0,155 6	0,373 3	0,323 3	0,1092	3,6%	0,179 6	0,373 3	0,343 3	0,1101	4,1%	0,177 4	0,366 7	0,353 3	0,0963	3,5%	0,166 4	0,353 3	0,333 3
8	Qinit + PA (sp)	0,0902	8,0%	0,160 1	0,353 3	0,313 3	0,1093	3,7%	0,180 0	0,366 7	0,346 7	0,1110	4,9%	0,179 3	0,373 3	0,336 7	0,0966	3,9%	0,170 5	0,366 7	0,330 0
9	Qinit +ICPU (p) + PA (sp)	0,0944	13,1%	0,164 3	0,360 0	0,323 3	0,1147	8,8%	0,187 9	0,386 7	0,353 3	0,1165	10,1%	0,187 2	0,380 0	0,363 3	0,1018	9,5%	0,176 4	0,366 7	0,346 7
10	Qinit + I (p)	0,0780	-6,6%	0,146 9	0,292 9	0,275 0	0,0960	-8,9%	0,162 7	0,335 7	0,300 0	0,0970	-8,3%	0,165 1	0,314 3	0,289 3	0,0839	-9,8%	0,152 7	0,307 1	0,285 7
11	Qinit + I (sp)	0,0664	-20,5%	0,123 0	0,278 6	0,239 3	0,0825	-21,7%	0,141 6	0,328 6	0,267 9	0,0838	-20,8%	0,140 2	0,314 3	0,264 3	0,0712	-23,4%	0,133 5	0,300 0	0,260 7
12	Qinit + ICPU (p) +I (sp)	0,0721	-13,7%	0,136 9	0,335 7	0,282 1	0,0866	-17,8%	0,154 2	0,300 0	0,278 6	0,0858	-18,9%	0,152 4	0,321 4	0,285 7	0,0726	-21,9%	0,146 3	0,285 7	0,271 4
13	Qinit + O (p)	0,0864	3,5%	0,151 5	0,372 4	0,313 8	0,1022	-3,0%	0,174 1	0,372 4	0,334 5	0,1032	-2,5%	0,171 7	0,365 5	0,337 9	0,0899	-3,3%	0,164 6	0,365 5	0,306 9
14	Qinit +O (sp)	0,0910	9,0%	0,158 6	0,366 7	0,313 3	0,1104	4,7%	0,179 5	0,380 0	0,356 7	0,1112	5,1%	0,177 4	0,373 3	0,353 3	0,0971	4,4%	0,171 4	0,346 7	0,326 7
15	Qinit +ICPU (p) + O (sp)	0,0935	12,0%	0,164 5	0,360 0	0,330 0	0,113	7,2%	0,183 0	0,373 3	0,363 3	0,1146	8,3%	0,184 5	0,380 0	0,363 3	0,1001	7,6%	0,173 6	0,346 7	0,343 3

Tableau 3.16 Résultats d'évaluation des 15 ensembles d'extension avec extension supplémentaire de « bo1 ». Rela : CHD, PAR, RB, RN,

Modèle		DPH					BM25					InL2					PL2				
Requêtes		MAP	Δ	Rprec	P@5	P@10	MAP	Δ	Rprec	P@5	P@10	MAP	Δ	Rprec	P@5	P@10	MAP	Δ	Rprec	P@5	P@10
	Qinit	0,0835		0,1530	0,3133	0,2933	0,1054		0,1741	0,3533	0,3167	0,1058		0,1755	0,3467	0,3200	0,0930		0,1696	0,3267	0,3200
1	Qinit + bo1	0,0867	3,8%	0,1512	0,2867	0,28	0,1135	7,7%	0,1835	0,3600	0,3700	0,1014	-4,2%	0,1665	0,3067	0,3167	0,0971	4,4%	0,1667	0,3067	0,3000
2	Qinit + Meta-MapLite + Bo1	0,0966	15,7%	0,1611	0,3467	0,3067	0,1149	9,0%	0,1813	0,36	0,3700	0,1174	11,0%	0,1845	0,3867	0,3733	0,1099	18,2%	0,178	0,3467	0,3467
3	Qinit + ICPU (p) + bo1	0,0995	19,2%	0,1689	0,3600	0,3233	<b>0,1186</b>	<b>12,5%</b>	<b>0,1851</b>	<b>0,3867</b>	<b>0,3900</b>	0,1178	11,3%	0,1855	0,3600	0,3767	0,1128	21,3%	0,1797	0,3667	0,3567
4	Qinit + ICPU (sp) + bo1	0,0584	-30,1%	0,1088	0,2214	0,1929	0,0716	-32,1%	0,1224	0,2429	0,2286	0,0735	-30,5%	0,1219	0,2714	0,2321	0,0682	-26,7%	0,1186	0,2214	0,2214
5	Qinit + P (p) + Bo1	0,0977	17,0%	0,1629	0,3533	0,3067	0,1170	11,0%	0,1821	0,3733	0,3667	0,1054	-0,4%	0,1693	0,3533	0,3400	0,1112	19,6%	0,1812	0,3400	0,3467
6	Qinit + P(sp) + Bo1	0,0972	16,4%	0,1629	0,3467	0,3067	0,1154	9,5%	0,1804	0,3800	0,3667	0,1174	11,0%	0,1840	0,3867	0,3733	0,1098	18,1%	0,1806	0,3400	0,3467
7	Qinit + PA (p) + Bo1	0,0982	17,6%	0,1611	0,3600	0,3267	0,1165	10,5%	0,1849	0,3533	0,3700	0,1147	8,4%	0,1825	0,4000	0,3667	0,1112	19,6%	0,1746	0,3600	0,3433
8	Qinit + PA (sp) + Bo1	0,0964	15,4%	0,1626	0,3333	0,3167	0,1176	11,6%	0,1828	0,3667	0,3767	0,1178	11,3%	0,1857	0,3867	0,3700	0,1077	15,8%	0,1734	0,3533	0,3333
9	Qinit + ICPU (p) + PA (sp) + bo1	0,0974	16,6%	0,1655	0,3467	0,31	0,1187	12,6%	0,1856	0,3867	0,3867	<b>0,1211</b>	<b>14,5%</b>	<b>0,1904</b>	<b>0,3867</b>	<b>0,3833</b>	0,1110	19,4%	0,1755	0,3667	0,3467
10	Qinit + I (p) + bo1	0,0738	-11,6%	0,1335	0,2857	0,2464	0,1032	-2,1%	0,1651	0,3286	0,3214	0,1026	-3,0%	0,1653	0,3643	0,3143	0,097	4,3%	0,1591	0,3071	0,2929
11	Qinit + I (sp) + bo1	0,0699	-16,3%	0,1204	0,2643	0,2500	0,0848	-19,5%	0,1362	0,3214	0,2750	0,0864	-18,3%	0,1378	0,3357	0,2679	0,0804	-13,5%	0,1368	0,3000	0,2607
12	Qinit + ICPU (p) + I (sp) + bo1	0,0784	-6,1%	0,1348	0,3286	0,2607	0,0885	-16,0%	0,1448	0,3143	0,2893	0,0889	-16,0%	0,1489	0,2929	0,2857	0,0836	-10,1%	0,1414	0,3000	0,2571
13	Qinit + O (p) + bo1	0,0937	12,2%	0,1575	0,3724	0,3207	0,1089	3,3%	0,1765	0,3655	0,3483	0,1097	3,7%	0,1778	0,3793	0,3379	0,1023	10,0%	0,1664	0,3379	0,3207
14	Qinit + O (sp) + bo1	0,0959	14,9%	0,1634	0,3400	0,3167	0,1148	8,9%	0,1811	0,3733	0,3667	0,1163	9,9%	0,1855	0,3867	0,3700	0,1068	14,8%	0,1727	0,3400	0,3367
15	Qinit + ICPU (p) + O (sp) + bo1	0,0972	16,4%	0,1658	0,3467	0,3133	0,1158	9,9%	0,1813	0,3933	0,3733	<b>0,1181</b>	<b>11,6%</b>	<b>0,1853</b>	<b>0,3733</b>	<b>0,3833</b>	0,1116	20,0%	0,1791	0,3600	0,3500

Le tableau 3.17 nous permet de voir l'effet de l'extension bo1 supplémentaire de l'ensemble Qinit +ICPU (p) + PA (sp) sur les différents sujets. Pour certains sujets, les scores ont fait des bonds jusqu'à 125 %. Cette amélioration est accompagnée par une dégradation de certains sujets, mais qui reste globalement tolérable (-19.6 en moyenne).

Tableau 3.17 Mesures détaillées et impact sur les « sujets » pour « Qinit +ICPU (p) + PA (sp) +bo1 » avec rela = Rela : CHD, PAR, RB, RN et InL2

N	Initiale (sujet de base)				Initiales+ICPU+PA+bo1 (H) InL2				
	MAP	R prec	P@5	P@10	MAP	$\Delta$	R prec	P@5	P@10
1	0,225 9	0,338 7	0,600 0	0,400 0	0,218 5	-3%	0,338 7	0,600 0	0,300 0
2	0,187 4	0,257 6	0,800 0	0,900 0	0,193 9	3%	0,293 6	0,800 0	0,900 0
3	0,129 9	0,215 7	0,200 0	0,100 0	0,215 9	<b>66%</b>	0,254 9	0,600 0	0,600 0
4	0,094 1	0,236 0	0,200 0	0,100 0	0,110 8	18%	0,258 4	0,000 0	0,300 0
5	0,024 6	0,066 7	0,000 0	0,000 0	0,017 2	-30%	0,077 8	0,000 0	0,000 0
6	0,261 0	0,375 0	0,600 0	0,600 0	0,255 1	-2%	0,375 0	0,600 0	0,700 0
7	0,069 9	0,176 3	0,200 0	0,100 0	0,084 6	21%	0,173 4	0,400 0	0,500 0
8	0,122 4	0,183 3	0,200 0	0,300 0	0,127 9	4%	0,208 3	0,200 0	0,400 0
9	0,239 5	0,373 5	1,000 0	0,900 0	0,268 1	12%	0,381 6	1,000 0	0,900 0
10	0,099 6	0,195 9	0,200 0	0,200 0	0,118 5	19%	0,226 8	0,200 0	0,200 0
11	0,066 1	0,023 8	0,000 0	0,000 0	0,037 3	-44%	0,000 0	0,000 0	0,000 0
12	0,152 3	0,209 1	1,000 0	0,700 0	0,146 2	-4%	0,204 5	1,000 0	0,900 0
13	0,0008	0,0000	0,0000	0,0000	0,0018	<b>125%</b>	0,0000	0,0000	0,0000
14	0,0116	0,0000	0,0000	0,0000	0,0070	-40%	0,0000	0,0000	0,0000
15	0,0570	0,1000	0,2000	0,1000	0,0536	-6%	0,1000	0,2000	0,1000
16	0,1922	0,2606	0,8000	0,9000	0,2405	25%	0,3239	0,8000	0,9000
17	0,1319	0,2328	0,0000	0,2000	0,1534	16%	0,2845	0,2000	0,3000
18	0,1019	0,1959	0,4000	0,6000	0,0961	-6%	0,2113	0,4000	0,4000
19	0,0567	0,0857	0,4000	0,2000	0,0559	-1%	0,0857	0,4000	0,2000
20	0,1194	0,1429	0,2000	0,3000	0,1058	-11%	0,1224	0,4000	0,2000
21	0,2167	0,3333	0,4000	0,4000	0,2046	-6%	0,2941	0,4000	0,5000
22	0,0503	0,1127	0,6000	0,4000	0,0408	-19%	0,0915	0,6000	0,3000
23	0,1915	0,3179	0,4000	0,5000	0,3959	<b>107%</b>	0,4462	1,0000	1,0000
24	0,0784	0,1875	0,4000	0,3000	0,1341	<b>71%</b>	0,2500	0,4000	0,5000
25	0,0841	0,2037	0,2000	0,3000	0,1273	<b>51%</b>	0,1852	0,2000	0,4000
26	0,0159	0,0500	0,2000	0,1000	0,0211	33%	0,0500	0,2000	0,1000
27	0,0233	0,0658	0,0000	0,1000	0,0258	11%	0,1053	0,0000	0,1000
28	0,0065	0,0182	0,2000	0,1000	0,0011	-83%	0,0182	0,0000	0,0000
29	0,0928	0,1429	0,6000	0,4000	0,0988	6%	0,1667	0,6000	0,5000
30	0,0708	0,1633	0,4000	0,4000	0,0759	7%	0,1837	0,4000	0,3000
Moy.	0,1058	0,1755	0,3467	0,3200	0,1211	<b>14,5%</b>	0,1904	0,3867	0,3833

### 3.12 Positionnement des résultats de l'extension ICPU

L'objectif principal de ces expérimentations et l'utilisation du corpus de test de la TREC 2017 sont de pouvoir se donner un repère expérimental concret avec une très grande objectivité. Les résultats ne font pas appel à des requêtes ou des jugements de pertinence qui proviennent d'un effort isolé et empreint de risque de subjectivité. De cette conviction et malgré le fait que nous ne disposons pas de ressources dans le domaine de l'oncologie (comme les équipes qui ont participé ayant participé à la TREC 2017). Nous avons souhaité comparer les valeurs obtenues de nos expérimentations aux résultats globaux de ces 29 équipes participantes et le résultat de leurs 125 exécutions (Goodwin *et al.*, 2017; López-García *et al.*, 2017; Mahmood *et al.*, 2017; Noh et Kavuluru, 2017; Roberts *et al.*, 2017).

Le tableau 3.18 illustre les meilleures, les moyennes et les plus faibles valeurs pour chaque sujet. Ces valeurs obtenues pour P@10 et Rprec ainsi que d'autres mesures comme le infNDCG ont été publiés. Nous nous sommes limités à la comparaison avec la P@10 et la Rprec car le infNDCG ne peut pas être comparé au MAP de nos expérimentations. Nous avons pris les deux ensembles qui ont eu les meilleurs scores P@10 et la Rprec dans nos expérimentations en l'occurrence : « Qinit+ICPU(p) » pour les relations RO — InL2 et Qinit +ICPU(p) +PA pour les relations CHD, PAR, RB, RN — InL2. Le classement montre que les valeurs obtenues situent les résultats de nos approches au-dessus de la moyenne des moyennes des scores obtenus par les autres équipes. Ceci permet d'un côté de valider nos démarches d'expérimentations et la bonne utilisation des ressources, des prétraitements ainsi que les logiciels d'évaluation. Et d'un autre côté, de confirmer les progrès réalisés par cette expansion dans un domaine tout à fait nouveau qui est la médecine de précision en l'oncologie. Nous en concluons que le caractère universel de l'outil ICPU et les orientations PICO peuvent devenir des pistes sérieuses à l'amélioration de la recherche d'information.

Tableau 3.18 Comparaison aux résultats statistiques par sujet (29 groupes et plus de 125 exécutions) pour 30 sujets pour la recherche des articles dans la littérature.

Q	P @ 10					Rprec				
	Meil-leur	Moyen	Faible	Q+ICPU	Q+ICPU +PA	Meil-leur	Moyen	Faible	Q+ICPU	Q+ICPU +PA
1	1,0000	0,6000	0,0000	0,4000	0,3000	0,5000	0,3387	0,0000	0,3387	0,3387
2	1,0000	0,9000	0,0000	0,9000	<i>0,9000</i>	0,4404	0,2742	0,0000	0,2659	<b>0,2936</b>
3	0,9000	0,3000	0,0000	0,5000	<b>0,6000</b>	0,4118	0,1961	0,0000	0,1765	<b>0,2549</b>
4	1,0000	0,7000	0,0000	0,1000	0,3000	0,4157	0,2097	0,0000	0,2434	<b>0,2584</b>
5	0,5000	0,2000	0,0000	0,0000	0,0000	0,2111	0,1222	0,0000	0,0667	0,0778
6	0,9000	0,6000	0,0000	0,7000	<b>0,7000</b>	0,4219	0,3047	0,0000	0,3906	<b>0,3750</b>
7	1,0000	0,5000	0,0000	0,4000	<i>0,5000</i>	0,3786	0,1705	0,0000	0,1763	<b>0,1734</b>
8	0,9000	0,3000	0,0000	0,3000	<b>0,4000</b>	0,3417	0,2333	0,0000	0,1833	0,2083
9	1,0000	0,8000	0,1000	0,9000	<b>0,9000</b>	0,4612	0,3490	0,0020	0,3673	<b>0,3816</b>
10	0,6000	0,2000	0,0000	0,2000	<i>0,2000</i>	0,2680	0,1340	0,0000	0,2062	<b>0,2268</b>
11	0,9000	0,2000	0,0000	0,0000	0,0000	0,5952	0,1429	0,0000	0,0238	0,0000
12	1,0000	0,8000	0,0000	0,7000	<b>0,9000</b>	0,4591	0,2409	0,0000	0,2091	0,2045
13	0,5000	0,1000	0,0000	0,0000	0,0000	0,2400	0,0400	0,0000	0,0000	0,0000
14	0,8000	0,0000	0,0000	0,0000	0,0000	0,5484	0,0000	0,0000	0,0000	<i>0,0000</i>
15	0,3000	0,1000	0,0000	0,1000	<i>0,1000</i>	0,3000	0,1000	0,0000	0,1000	<i>0,1000</i>
16	1,0000	0,6000	0,0000	0,8000	<b>0,9000</b>	0,4648	0,2606	0,0000	0,2606	<b>0,3239</b>
17	1,0000	0,3000	0,0000	0,2000	<i>0,3000</i>	0,3707	0,2328	0,0000	0,2500	<b>0,2845</b>
18	1,0000	0,5000	0,0000	0,5000	0,4000	0,2990	0,1546	0,0000	0,2216	<b>0,2113</b>
19	0,9000	0,2000	0,0000	0,2000	<i>0,2000</i>	0,3429	0,0857	0,0000	0,0857	<i>0,0857</i>
20	0,7000	0,2000	0,0000	0,3000	<i>0,2000</i>	0,3469	0,1224	0,0000	0,1429	<i>0,1224</i>
21	0,9000	0,5000	0,0000	0,5000	<i>0,5000</i>	0,3922	0,2990	0,0049	0,3480	0,2941
22	0,9000	0,4000	0,0000	0,4000	0,3000	0,4155	0,1127	0,0000	0,1127	0,0915
23	1,0000	0,6000	0,0000	0,8000	<b>1,0000</b>	0,5744	0,3128	0,0000	0,4205	<b>0,4462</b>
24	1,0000	0,4000	0,0000	0,4000	<b>0,5000</b>	0,4375	0,1719	0,0000	0,2031	<b>0,2500</b>
25	0,7000	0,3000	0,0000	0,4000	<b>0,4000</b>	0,3704	0,1667	0,0000	0,1852	<b>0,1852</b>
26	0,8000	0,1000	0,0000	0,1000	<i>0,1000</i>	0,5000	0,0500	0,0000	0,0500	<i>0,0500</i>
27	0,8000	0,1000	0,0000	0,1000	<i>0,1000</i>	0,2105	0,0921	0,0000	0,0789	<b>0,1053</b>
28	0,9000	0,2000	0,0000	0,1000	0,0000	0,4000	0,0727	0,0000	0,0182	0,0182
29	1,0000	0,3000	0,0000	0,5000	<b>0,5000</b>	0,4048	0,1429	0,0000	0,1190	<b>0,1667</b>
30	0,9000	0,2000	0,0000	0,4000	<b>0,3000</b>	0,3265	0,1497	0,0000	0,1701	<b>0,1837</b>
Moy.		<b>0,3733</b>		0,3633	<b>0,3833</b>		0,1760		<b>0,1805</b>	<b>0,1904</b>

### 3.13 Conclusion et travaux futurs

Ce chapitre a été motivé par la question de recherche : peut-on améliorer la pertinence des résultats de la recherche d'information, et ce, par une extension de requête de recherche utilisant des concepts potentiellement pertinents issus d'UMLS et une pondération reflétant la centralité de ces nouveaux concepts ?

À la suite de nos développements, nous pouvons constater que la réponse est affirmative et que cette amélioration est significative sur des ensembles de sujets de recherche pas très informatifs au départ (très peu de termes). Cela a été possible grâce à l'outil d'exploration ICPU et une pondération représentant la centralité de ces concepts au sein de leur graphe d'exploration ontologique.

Pour l'amélioration de ces résultats, cela reste très possible au vu des différents paramètres de départs qui peuvent être modifiés et particulièrement :

- 1) Le choix des relations à explorer entre concepts par ICPU peut être bonifié par l'ajout d'autres types de relations comme celle : CHD, PAR, QB, RL, SY (Hiérarchiques — synonymie) et CHD, PAR, RO (Hiérarchiques – associatives).
- 2) Dans les relations associatives RO (sémantiques) elles-mêmes, une présélection de types de relation à explorer par l'ICPU peut avoir un impact différent sur les résultats. Ceci va limiter l'exploration de l'ICPU à un ensemble plus significatif de relations en lien avec une particularité comme l'anatomie du corps humain ou le diagnostic des maladies. Néanmoins, cela demande une bonne connaissance du domaine médical pour pouvoir choisir les relations susceptibles d'améliorer la détection des concepts pertinents.
- 3) La longueur de la marche d'exploration dans UMLS, le nombre d'itérations et le seuil de visite pour la sélection des concepts peuvent faire varier les résultats de la marche aléatoire de l'ICPU et modifier ainsi les résultats des extensions.

- 4) Une étude d'optimisation peut être effectuée sur le nombre de concepts à ajouter au moment de l'extension de la requête. Cette étude pourrait déterminer le nombre idéal de concepts pour parvenir à une plus haute MAP.
- 5) Un certain nombre de thésaurus ont été utilisés dans les expérimentations (MTH, MeSH, SNOMED\_CT, NCI) pour leur lien possible avec le domaine médical, mais il en existe une centaine. Ajouter ou supprimer des thésaurus peut aussi influencer sur la richesse des concepts rapportés par la marche aléatoire. Ceci constitue, en réalité une nouvelle piste de recherche.
- 6) Utiliser et intégrer de nouvelles méthodes d'identification des éléments PICO (apprentissage machine ou apprentissage profond) pour une meilleure classification des concepts retrouvés par ICPU.
- 7) Comme pour les équipes qui ont participé dans la TREC, il est possible d'utiliser d'autres ressources extérieures et spécialisées dans le domaine de l'oncologie et de faire d'autres traitements en aval et en amont de l'extension pour augmenter son efficacité.

Pour ce qui est de la deuxième question de recherche, nous voulions savoir si cette extension orientée par des concepts appartenant aux catégories PICO (« Problème », « intervention » et « outcome ») pourrait améliorer la recherche d'information d'articles scientifiques issus de la recherche clinique.

Nos expérimentations ont montré qu'il est possible de mettre à contribution ces concepts, notamment ceux PA et O. En effet, une amélioration significative a été constatée. Ceci nous donne des pistes solides pour continuer à travailler pour améliorer les approches afin de valoriser au mieux ces concepts. L'amélioration pourrait venir de réévaluation de la méthode de pondération proposée ainsi que du nombre de concepts à ajouter pour augmenter les scores obtenus.



## CHAPITRE IV

ÉVALUATION DE L'AIDE À LA DÉCISION RENFORCÉE PAR UN ACCÈS  
AUX RÉSUMÉS D'ARTICLES AU MOMENT DE LA PRISE DE DÉCISION  
CLINIQUE : UNE ÉTUDE COMPARATIVE DANS LE CONTEXTE DE LA  
MÉDECINE TRANSFUSIONNELLE PÉDIATRIQUE.

## Préambule

Cette enquête, qui a été menée auprès des praticiens, est primordiale pour jauger la pertinence de l'idée générale derrière le modèle que nous proposons afin de favoriser l'application de la MBE (section 1.5). Rappelons que ce modèle préconise une meilleure prise de décision avec l'apport d'un accès à des articles scientifiques au moment de la prise de décision. Cette enquête émule un déploiement d'un système doté de cette capacité de recherche d'information dans des conditions semblables à son utilisation par les praticiens. À noter, qu'un déploiement d'une solution complète nécessite un long cheminement d'approbation dans le milieu médical. Pour ce qui est de ce travail, deux facettes ont été analysées : 1) l'aspect médical et décisionnel concernant l'adhésion des médecins aux recommandations en matière de transfusion si un accès leur est offert au moment de la prise de décision, 2) les aspects cognitifs qui entourent cette prise de décision et les biais cognitifs engagés. Ceci permettra une meilleure compréhension des tenants et des aboutissants qui peuvent joncher le chemin du déploiement d'un tel modèle. Ce chapitre est présenté sous la forme d'un protocole d'enquête tel qu'il a été soumis pour approbation aux comités scientifique et éthique de l'hôpital Sainte Justine et de l'UQAM. Notre enquête ne présente aucun risque. Elle ne se fait que dans un contexte d'évaluation de la décision médicale liée à une pratique spécifique et n'implique aucune intervention sur des patients. Le protocole a été soumis au comité d'éthique de la recherche pour approbation. L'enquête suit les lignes directrices pour la bonne gestion des données (ICH 6<sup>45</sup>) et la confidentialité des participants a été maintenue.

---

<sup>45</sup> <https://www.canada.ca/en/health-canada/services/drugs-health-products/drug-products/applications-submissions/guidance-documents/international-conference-harmonisation/efficacy/good-clinical-practice-consolidated-guideline-topic.html>

## 4.1 Résumé du protocole

Tableau 4.1 Résumé du protocole.

Titre :	Impact de l'aide à la décision renforcée par un accès aux résumés d'articles de références au moment de la prise de décision : Une étude comparative dans le contexte de la médecine transfusionnelle pédiatrique
Description de l'étude :	L'étude est une enquête auprès des praticiens afin d'évaluer l'accès aux résumés (abstract) d'articles scientifiques particulièrement pertinents en plus des suggestions simulées d'un système d'aide à la décision améliore la confiance décisionnelle de médecins en pratique transfusionnelle sur des patients traités dans un service de soins intensifs pédiatriques
Objectifs :	Évaluer la conformité des décisions des praticiens dans la prise en charge transfusionnelle des patients pédiatriques à la suite de l'exposition (ou non) à des résumés d'articles de références).
Critères d'évaluation	Le taux de réponse en accord avec l'arbre de décision suivant les recommandations du groupe de « Transfusion and Anemia expertise Initiative. TAXI »
Population	Les soignants de patients traités aux soins intensifs pédiatriques du CHU Sainte-Justine
Description des sites participants :	CHU Ste-Justine
Description de l'intervention :	Enquête
Durée de l'étude :	3 mois
Durée de la participation :	20 minutes

## 4.2 Schéma

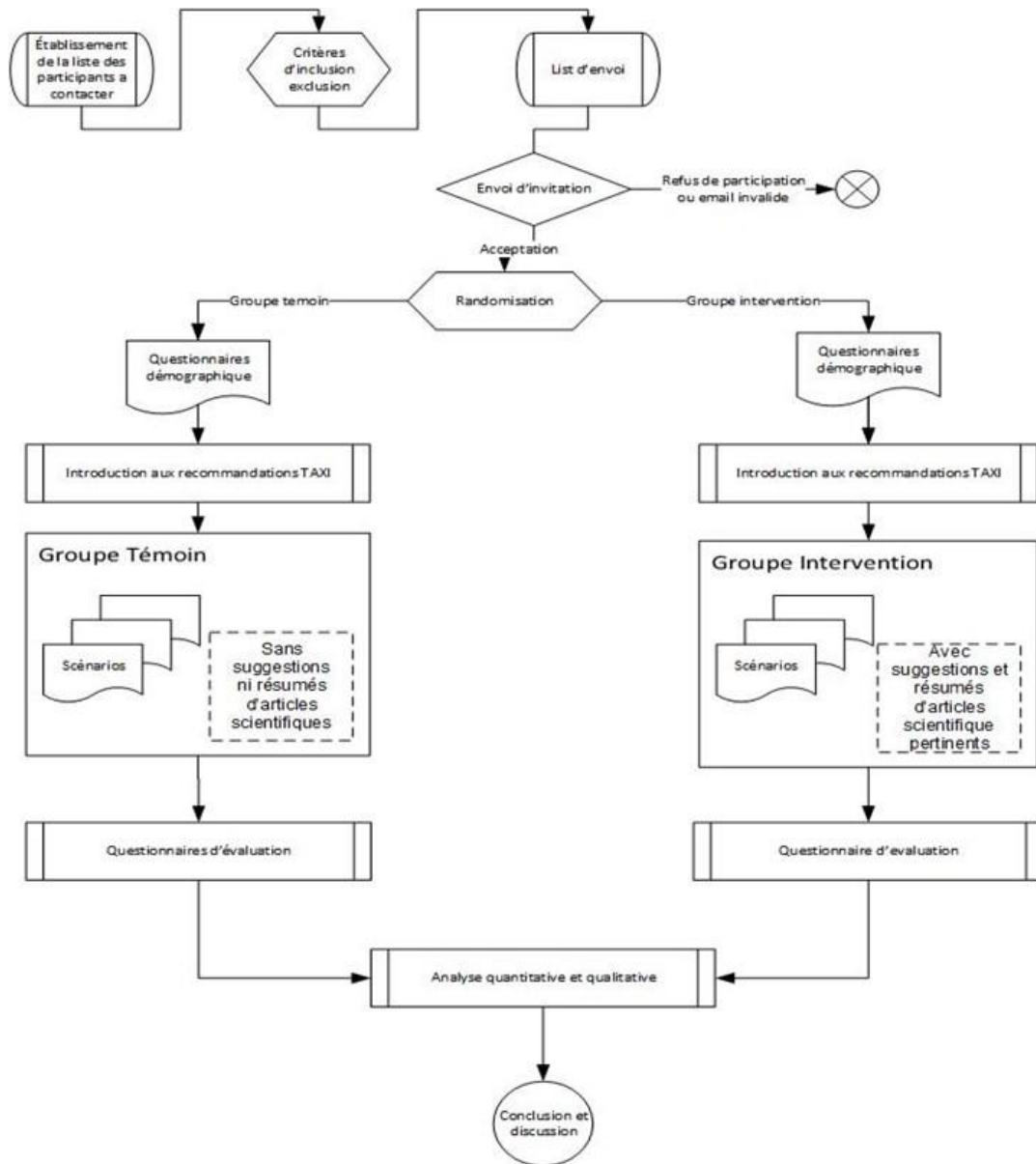


Figure 4.1 Schéma du déroulement de l'enquête.

### 4.3 Question de recherche

Se pourrait-il qu'un accès aux résumés (« *abstracts* ») d'articles scientifiques particulièrement pertinents, au moment de la prise de décision et en situation d'incertitude, permette d'améliorer l'adhérence des praticiens aux nouvelles recommandations dans la médecine transfusionnelle pour des patients traités dans un service de soins intensifs pédiatriques ? C'est la question à laquelle nous voudrions répondre.

Plusieurs données publiées dans la littérature suggèrent qu'un soutien supplémentaire en évidences (Benjamin et Gordon, 2017; Henry *et al.*, 1998; Sackett, David L. *et al.*, 1998) à la prise de décision médicale pourrait effectivement améliorer les pratiques dans le cadre de la médecine basée sur les évidences (MBE). Toutefois, ceci n'a pas fait l'objet de vérifications dans la pratique transfusionnelle pour des suggestions d'aide à la décision jumelée avec un accès à des résumés d'études scientifiques pertinentes en lien avec des scénarios. Par les résultats de cette enquête, nous serions en mesure d'évaluer la pertinence (au regard du niveau de conformité des praticiens par rapport aux propositions des résumés) de déployer un modèle de système d'aide à la avec une capacité de recherche d'articles (résumés).

Remarque : Cette habilité concernant un système qui peut retrouver d'une manière automatique des résumés articles liés au contexte de décision est en développement dans le volet informatique de notre thèse doctorale. L'évaluation de la pertinence des résumés d'articles associés aux scénarios de l'enquête a été opérée par des médecins spécialistes, ce qui simulera une situation optimale. L'objectif ultime de ce programme de recherche est d'améliorer la pratique médicale concernant la transfusion de globules rouges à des enfants traités en réanimation.

#### 4.4 Contexte

Une pratique optimale de la médecine basée sur les évidences (MBE) nécessite l'intégration des meilleurs résultats de la recherche clinique et de l'expertise clinique individuelle des praticiens (Greenhalgh *et al.*, 2014; Sackett, David L *et al.*, 1996). La MBE propose un nouveau paradigme, selon lequel un médecin ou un autre soignant appelé à prendre une décision médicale doit tenir compte non seulement de son expérience clinique et de ses connaissances de base, mais aussi des préférences du patient et (en priorité) des évidences disponibles dans la littérature scientifique médicale (Greenhalgh *et al.*, 2014; Sackett, David L *et al.*, 1996). Par exemple, l'application de la MBE s'est traduite par plusieurs succès dès ses débuts (Fineout-Overholt *et al.*, 2005; Greenhalgh *et al.*, 2014; Straus et McAlister, 2000). L'application du guide des directives pour le traitement de l'asthme établi par la société thoracique britannique est un bel exemple de succès lié à la MBE. Ce document qui a été mis au point en 1990, à la suite d'un consensus, se base principalement sur les résultats d'une combinaison d'essais randomisés et d'études observationnelles. Le suivi de l'application de ces directives a démontré que l'utilisation des stéroïdes inhalés pour l'asthme a fait chuter le taux de mortalité et la comorbidité (Sackett, David L *et al.*, 1996).

Par ailleurs, et selon (Fineout-Overholt *et al.*, 2005), des études ont pu démontrer que des patients ayant profité de soins basés sur des évidences ont présenté une amélioration de leur sort d'au moins 28 % comparé au sort de ceux dont les soins furent prodigués de façon habituelle. Pour Tonneli (2011), une MBE assure une pratique clinique plus uniforme et plus scientifique. L'incorporation des meilleurs résultats de recherche actuels dans la prise de décision clinique permet de raccourcir le délai traditionnel entre la production de données probantes et leur application et d'augmenter la proportion de patients auxquels le meilleur traitement actuel est offert. Cependant, l'intégration de la MBE au moment de la prise de décision clinique n'est pas chose simple. Selon Greenhalgh (2014), le volume de données disponibles est devenu si grand que cela constitue

en soi un problème majeur alors que cela devrait en réalité représenter une opportunité formidable. En particulier, le nombre de lignes directrices cliniques publiées est devenu à la fois ingérable et insondable. En 2005, un audit de 24 heures portant sur 18 patients hospitalisés avec 44 diagnostics a identifié 3679 pages de lignes directrices qu'il serait nécessaire de lire pour traiter ces cas selon les normes proposées, ce qui correspondrait à environ 122 heures de lecture. Dans la pratique, imposer une telle charge de travail aux praticiens irréaliste.

Quelle que soit la méthode de collectes de données (par exemple recherche d'information à titre individuel versus par l'entremise d'un outil particulier), le fait de mettre rapidement à la disposition des cliniciens des éléments de preuve relatifs à une pratique médicale donnée permet d'améliorer la prise de décision. Même l'aide d'un charriot pour dossiers documentaires (éléments de preuve écrite sur du papier) augmente l'intérêt des praticiens à rechercher et intégrer ces évidences dans la prise de décision relative à des soins particuliers (Sackett, David L *et al.*, 1996; Sackett, David L. *et al.*, 1998); fait intéressant, l'effet stimulant d'un tel charriot s'estompe après son retrait. D'autres études comme celle de (Henry *et al.*, 1998) se sont intéressées à une prise de décision soutenue par des résumés d'articles afin de mesurer l'impact d'une telle approche.

L'étude de Henry *et al.* (1998) conclut que le format (non structuré; structuré (contexte, conception de l'étude, population, collecte de données, résultats, conclusion); format IMRAD (introduction, méthodes, résultats et discussion) ou le format orienté patient avec uniquement les preuves qui comptent pour l'examen) de présentation d'un résumé n'a pas d'impact sur la prise de décision; en revanche, le fait de disposer d'informations valables a semblé influencer les décisions prises dans le contexte d'un scénario clinique. Il reste que ces études ne révèlent pas quel est l'impact d'un outil d'aide à la décision dont les conseils sont appuyés par des résumés, tout particulièrement si la recommandation est ouverte et n'est pas totalement consensuelle. Une autre étude

(Fontelo P *et al.*, 2015) constate que les cliniciens préfèrent les résumés succincts. Il ressort également de cette étude que les cliniciens considèrent que les preuves qui conduisent à de nouvelles compétences en diagnostic, à un nouveau test de diagnostic ou à un nouveau plan de traitement sont la principale raison de l'utilité des preuves. De ce qui précède, il est clair que l'accès rapide aux résumés d'articles reste un moyen intéressant pour appuyer la MBE au quotidien. Cet impact serait-il le même dans un domaine particulier comme la transfusion en soins intensifs ou la prise de décisions est assujettie à des conditions d'urgence, de risque et d'incertitude.

#### 4.5 Contribution

La principale contribution de ce chapitre est la mise en place d'une étude clinique randomisée pour étudier l'impact d'une intervention sur la prise de décision médicale. Cette approche nécessite une grande rigueur éthique et scientifique pour avoir des résultats crédibles. Les résultats auprès d'un ensemble de praticiens permettent d'analyser plusieurs facettes de cette décision médicale (transfusion) au vu de l'hypothèse que nous avons formulée.

#### 4.6 Rôle des technologies

Les progrès des technologies de l'information et de l'informatique ont grandement contribué au développement de la MBE (Georgiou, 2001). Plusieurs auteurs et institutions savantes (Bakken, 2001; Bakken *et al.*, 2004; Lisa, 2007) abordent le rôle de l'informatique dans la promotion et la diffusion de la médecine basée sur les évidences (MBE) en mettant l'accent sur l'aspect crucial et critique de ce rôle, sans lequel la MBE serait toujours à la traîne. Plus récemment, Les applications mobiles ou web peuvent être très utiles pour fournir aux professionnels de la santé un moyen rapide d'accéder à la littérature médicale en ligne pour soutenir la pratique de la médecine fondée sur des preuves (Djulgovic B et Guyatt GH, 2017; Fontelo P *et al.*, 2015). Grâce à ces applications,



les résumés accessibles via une recherche peuvent jouer un rôle important dans l'intégration des preuves médicales actuelles dans le flux de travail de la pratique clinique. Toutefois, convaincre les médecins de l'efficacité de ces systèmes constitue un défi qui peut être difficile à relever, étant donné que bon nombre de médecins sont réticents face au changement de leur pratique et s'y connaissent peu en matière d'informatique (Sim *et al.*, 2001).

Certains de ces systèmes informatiques et notamment les systèmes d'aide à la décision peuvent avoir des effets inverses et ce, par l'inhibition du raisonnement clinique des praticiens. En effet, il se pourrait aussi qu'un clinicien inexpérimenté s'aligne mécaniquement sur les suggestions d'un outil d'aide à la décision simplement afin d'éviter tout risque de litige subséquent (Lisa, 2007). Aussi et malgré leur apport, la plupart des systèmes d'aide à la décision contemporains sont basés sur des guides de pratique et des évidences formalisées au moment de la création du système (Afzal *et al.*, 2015). Ceci implique que ces systèmes restent figés en attendant une mise à jour, laquelle est souvent coûteuse, laborieuse et viendrait probablement en retard pour bon nombre de patients. Le modèle général que nous proposons (voir section 1.5) permet de remédier à cette situation et ce, en s'assurant que les praticiens aient accès à de la littérature récente et adapter au cas de leur patient.

#### 4.7 Aspect cognitif lié à la prise de décision médicale

Il ne faudrait pas négliger le processus cognitif que les praticiens utilisent lorsqu'ils doivent prendre une décision, que ce soit la décision de transfuser ou celle de lire ou non un article scientifique. La recherche en psychologie cognitive et en théorie de la décision comportementale a identifié plusieurs processus cognitifs utilisés par les décideurs lorsqu'ils traitent une situation complexe, ambiguë et incertaine (Jim Q. Chen, 2003). Suivant la perspective de la rationalité limitée introduite en sciences cognitives, en théorie de la décision et en économie par Hebert Simon (1957), ces processus sont

généralement interprétés comme des stratégies heuristiques (ou simplement « heuristiques »), c'est-à-dire des simplifications cognitives permettant de prendre des décisions suffisamment bonnes (*satisficing*) en situation d'incertitude. Ces stratégies permettent de prendre de bonnes décisions dans plusieurs circonstances. En revanche, la manière dont elles le font implique qu'elles feront nécessairement des erreurs dans d'autres situations, ce qu'on appelle des biais cognitifs (Kahneman *et al.*, 1982).

Nous introduisons ici quelques-unes de ces heuristiques ainsi que les biais qui leur sont associés, notamment la disponibilité, l'ajustement, l'hypothèse préalable et le raisonnement par analogie, lesquels nous permettront à la fois de mieux comprendre les processus humains de décision afin d'améliorer éventuellement notre algorithme ; mais également pour comprendre pourquoi certains médecins sont plus disposés à lire de nouveaux articles que d'autres, et pourquoi la lecture des articles est plus profitable à certains médecins qu'à d'autres, ce qui permettra éventuellement d'améliorer notre algorithme afin qu'il soit capable de tenir compte de ces différences.

Pour l'heuristique de la disponibilité, il s'avère que les informations rapidement disponibles en mémoire de travail, soit parce qu'elles reflètent des événements récents ou fréquents, ont un impact important sur nos raisonnements. L'heuristique de disponibilité reflète cet accès rapide à certaines informations aux dépens d'autres. Les preneurs de décision accorderont en général plus d'importance à des événements récents ou fréquents, car il leur est plus facile de se les rappeler de mémoire. Puisqu'en général ces événements récents ou fréquents ont de fait une importance accrue eu égard aux actions que l'on doit effectuer au moment présent, une telle heuristique favorise une bonne prise de décision, bien qu'il arrive à l'occasion que ces informations ne soient pas pertinentes, auquel cas l'heuristique de disponibilité causera divers biais cognitifs. Ainsi les gens auront tendance à surestimer la probabilité de décéder dans un accident d'avion, et sous-estiment celui de mourir dans un accident de voiture, car chaque accident d'avion est mentionné dans les médias alors que peu d'accidents de voiture le sont.

L'hypothèse préalable est un processus simplificateur qui décrit les préjugés sur une hypothèse antérieure. Les individus ont tendance à rechercher et à utiliser des informations cohérentes plutôt que des informations incohérentes avec leurs convictions. Ce biais explique pourquoi parfois des individus prennent de mauvaises décisions en se basant sur des hypothèses erronées, même si de nombreuses évidences montrent que ces hypothèses sont erronées (Jervis, 1976).

L'heuristique d'ancrage et d'ajustement (Kahneman *et al.*, 1982; Tversky, A et Kahneman) reflète aussi cette tendance à accorder trop d'importance à une information présente en mémoire de travail. Si on demande à un sujet de donner un chiffre au hasard dans une certaine plage et si on demande par la suite à ce même sujet d'estimer une valeur par exemple, si on demande à un sujet de donner un chiffre de 1 à 75 et qu'on lui demande d'estimer par la suite l'âge d'une personne sur une photo, celui-ci sera influencé par le premier chiffre donné : s'il est petit, la personne sera estimée plus jeune que si le chiffre initial est grand. Le chiffre initial ancre le jugement. Lors de la prise de décision, les preneurs de décision émettent souvent des jugements initiaux sur certaines variables de décision, lesquels ancrent de la même manière le reste du processus de prise de décision. Lorsque de nouvelles données deviennent disponibles, le preneur de décision pourra ajuster son estimation initiale. Cependant, les ajustements sont généralement insuffisants : le jugement final est souvent biaisé par l'estimation initiale qui a ancré tout le processus de décision.

Une autre heuristique fréquemment rencontrée, et responsable de plusieurs biais, est l'heuristique de représentativité, selon laquelle plus un objet ressemble à un cas prototypique d'une catégorie, plus il a de chance d'appartenir à cette catégorie. Cette heuristique est illustrée par le fameux cas de Linda, qui est jugée comme ayant plus de chance d'appartenir à la classe des caissières féministes qu'à la classe des caissières, violant ainsi une loi fondamentale de la théorie de la probabilité (la loi de conjonction), parce qu'une description d'elle la fait ressembler davantage à une caissière féministe

qu'à une simple caissière (Tversky, Amos et Kahneman, 1983). Cette heuristique est responsable de plusieurs biais cognitifs, dont l'ignorance des probabilités a priori et l'illusion de validité. L'ignorance des probabilités a priori (ou ignorance du taux de base) se manifeste lorsque des personnes ne tiennent compte que de la vraisemblance d'un événement (étant donné la description de Linda, il semble plus probable qu'elle soit une caissière féministe) et négligent une information préalable importante (il y a énormément plus de simples caissières que de caissières féministes). En termes bayésiens, ceci revient à confondre la probabilité a posteriori d'un événement avec sa vraisemblance (*likelihood*) parce qu'on néglige de tenir compte (multiplicativement) de la probabilité a priori de cet événement. Lorsque le résultat prédit est bien ajusté aux informations disponibles, l'heuristique de représentativité est également responsable de l'illusion de validité qui donne aux gens une tendance à tendance à être trop confiants dans leurs croyances et leurs jugements (Fichhoff *et al.*, 1977; Kahneman et Tversky, 1973). Un tel excès de confiance peut s'avérer contre-productif. Cela indique que souvent les gens ne savent pas combien ils savent et combien d'informations supplémentaires dont ils ont besoin.

Il faut tenir compte de ce qui précède et tenir compte de tous les facteurs qui entourent la prise de décision dans le contexte de la MBE. Il serait souhaitable de renforcer l'effet des systèmes d'aide à la décision en combinant les suggestions faites par ce système et la mise en disponibilité du résumé des articles les plus récents et les plus pertinents retrouvés dans la littérature. Après validation par les experts, ces articles permettront aux lecteurs d'avoir un aperçu rapide de l'information véhiculée par l'article intégral, ce qui permettrait une prise de décision mieux éclairée. L'efficacité escomptée d'une telle mesure est de faire en sorte que les praticiens se sentent plus en confiance pour adopter la suggestion proposée en présence de justificatifs énoncés dans des résumés dont le format leur serait familier.

#### 4.8 La MBE appliquée à la pratique transfusionnelle

En médecine, la pratique concernant la transfusion de globules rouges consiste à prendre en charge des patients souffrant de perte de sang aiguë ou d'une anémie grave. La décision de transfuser du sang doit être étayée par la nécessité de soulager des signes et symptômes cliniques et de prévenir une morbidité ou le décès. Toute transfusion doit être prescrite uniquement lorsque les avantages pour le patient sont susceptibles de l'emporter sur les risques.

Comme dans d'autres domaines, la transfusion n'est pas en reste en ce qui concerne le retard d'application de la MBE (Butler *et al.*, 2015). La littérature rapporte une augmentation des transfusions de globules rouges dans la pratique médicale. De plus, certaines études laissent entendre qu'une proportion importante de ces transfusions n'est pas justifiée (Rajbhandary *et al.*, 2014-15). Cet état de fait devient problématique, car en plus des risques que peut engendrer une transfusion pour le patient, toutes les banques de sang à travers le monde souffrent d'un manque de donneurs et de produits à transfuser (Goel *et al.*, 2018 Feb). Plusieurs groupes de travail et chercheurs militent pour une approche plus restrictive et mieux adaptée au besoin des patients en tentant d'optimiser le recours à la transfusion aux patients qui en ont vraiment besoin.

Un exemple dans l'effort d'appliquer la MBE, la « *Society for the Advancement of Blood Management* » (SABM<sup>46</sup>) (Susan M *et al.*, 2019), propose la mise en place d'indicateurs traitent des activités cliniques liées au programme de gestion du sang pour

---

<sup>46</sup> La Société pour l'avancement de la gestion du sang (*Society for the Advancement of Blood Management*) est une organisation professionnelle à but non lucratif composé d'une équipe multidisciplinaire de fournisseurs de soins de santé qui éduque les soignants sur les problèmes cliniques associés à la transfusion sanguine et leur impact sur les résultats pour les patients. (<https://www.sabm.org/publications>.)

les patients en pédiatrie et visent à optimiser les résultats cliniques, à améliorer la sécurité des patients et à réduire les coûts des soins de santé. Ces indicateurs (ex. 1- Les directives de transfusion pour tous les composants sanguins doivent être adaptées au poids et à l'âge de l'enfant ; 2- Le service de transfusion doit disposer d'une politique et de procédures pour garantir les culots de globules rouge frais et lavés... etc. <https://www.sabm.org/publications>) fournissent aux soignants un modèle des éléments requis qui doivent être traités collectivement dans les soins du patient pédiatrique afin de réduire au minimum l'anémie iatrogène et de limiter l'exposition allogénique à la transfusion de composants sanguins. Reste que ces normes ne sont pas destinées à fournir des indications strictes, des contre-indications ou d'autres critères pour la pratique de la médecine clinique et de la chirurgie. Les décisions cliniques doivent être fondées sur les pratiques acceptées localement et sur l'état clinique du patient pédiatrique.

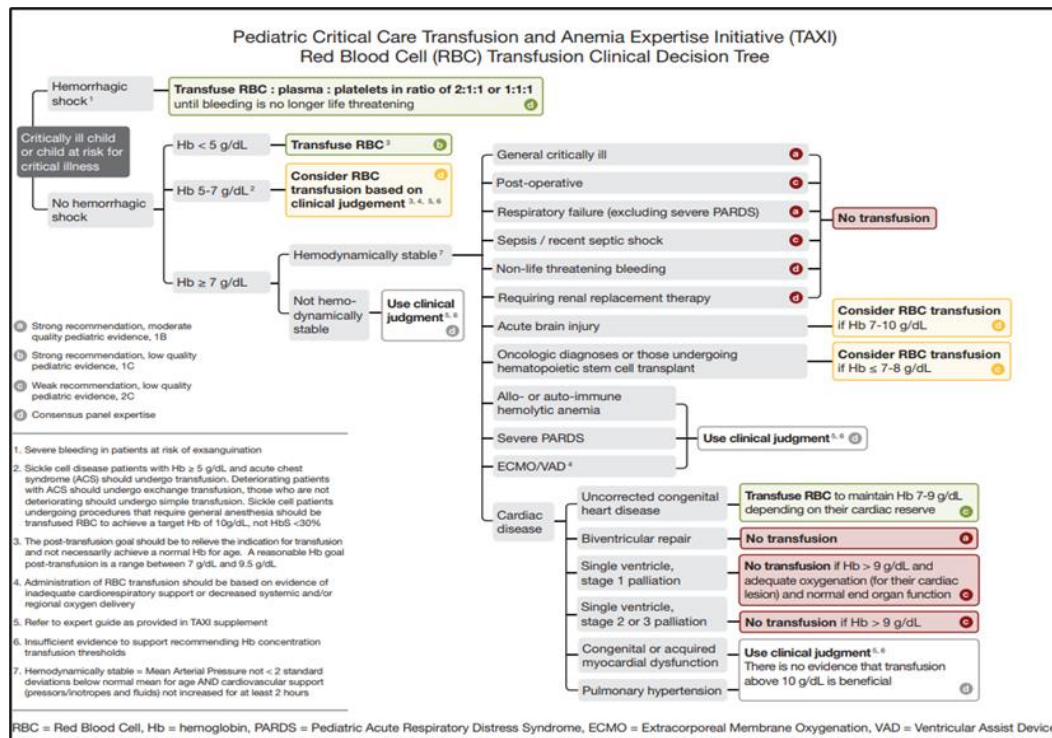


Figure 4.2 Arbre de décision selon le consensus de limitatif TAXI pour la transfusion.

Un groupe d'experts dénommé « Transfusion and Anemia eXpertise Initiative » (TAXI) (Bembea MM et al., 2018) a proposé un arbre de décision qui orienterait les praticiens sur la décision à prendre pour la transfusion de concentrés de globules rouges. Il est prévu de créer un système électronique d'aide à la décision - un système expert - se basant sur cet arbre de décision afin d'améliorer l'adhérence des praticiens aux recommandations du groupe TAXI, lesquelles sont résumées dans l'arbre de décision contenu dans la figure 4.2.

Toutefois, cela peut prendre plusieurs années avant que le développement et la mise en place d'un tel système puissent se faire et qu'on puisse démontrer son efficacité et sa

sécurité. Néanmoins, notre étude, tout en focalisant sur le processus décisionnel, pourrait contribuer à donner un aperçu préliminaire d'un tel déploiement.

Il existe des données suggérant qu'un tel système d'aide à la décision médicale puisse améliorer la pratique transfusionnelle. (Butler *et al.*, 2015) a étudié l'impact d'un système d'alerte qui réagissait à toute prescription d'une transfusion en proposant une stratégie restrictive (en application de la MBE) quand cela était approprié. Cette étude a démontré qu'il y'a eu réduction des transfusions de globules rouges de 69.0% à 43.4% ( $p \leq 0.005$ ) et des transfusions de plaquettes de 41.9% à 31.2% ( $p = 0.16$ ). Ce système ne couvrait qu'une fraction de l'arbre de décision ci-dessus et n'a pas été implanté en milieu hospitalier jusqu'à présent. Cette étude ne s'attarde pas sur l'aspect de prise de décision des praticiens, mais renforce l'idée que les systèmes d'alertes et d'aide à la décision sont des outils qui favorisent l'application de la MBE.

Dans une approche complémentaire, une autre équipe de notre groupe de recherche travaille sur le développement d'un outil d'aide à la décision médicale qui favoriserait l'application des recommandations contenues dans l'arbre de décision produit par le groupe TAXI (Valentine *et al.*, 2018) : il s'agira d'un système ouvert qui proposera une conduite médicale sans l'imposer.

Dans l'enquête que nous avons réalisée, nous avons vérifié, en utilisant un questionnaire basé sur des scénarios pré-validés, si l'accès à des résumés de références améliore le transfert des connaissances. Nous avons eu recours à un devis particulier en réalisant notre enquête : en effet, nous avons envoyé le même formulaire à tous les répondants, mais nous avons randomisé les répondants à recevoir ou à ne pas recevoir d'information supplémentaires sous forme d'abstracts. Notre hypothèse était que les participants pouvant profiter de références auront de meilleurs résultats au questionnaire que le groupe témoin. Le paramètre du critère de jugement principal — meilleur résultat au questionnaire — sera la proportion de réponses adéquates, c'est-à-dire conformes aux



recommandations de la littérature, en particulier celles proposées par le groupe TAXI dans l'arbre de décision contenu dans la figure 4.2.

Le critère de jugement principal de notre étude randomisée sera le taux de réponse en accord avec l'arbre de décision de la figure 4.3. Ce questionnaire comprendra plusieurs scénarios cliniques standardisés inspirés par l'arbre de décision TAXI. Ces scénarios mettent les praticiens devant des situations où la décision n'est pas évidente à première vue. Nous avons repris certains scénarios déjà employés par notre équipe lors d'enquêtes précédentes menées en 1997 et en 2010 auprès des intensivistes de plusieurs centres aux États-Unis et en France (Du Pont-Thibodeau *et al.*, 2014; Laverdière *et al.*, 2002). Dans ces deux études, les participants ont été exposés à quatre scénarios de cas de soins intensifs pédiatriques. Les questions étaient conçues pour évaluer si la pathologie principale dans chaque scénario changeait le niveau d'hémoglobine (Hb) qui inciterait les répondants à prescrire une transfusion de globules rouges.

En comparant les résultats obtenus en 1997 et en 2010, il a été constaté que la pratique transfusionnelle chez l'échantillon des intensivistes pédiatriques participant avait évolué vers un régime plus restrictif probablement à la suite de la publication de l'étude « *Transfusion Requirement in PICU* » (TRIPICU) dans le *New England Journal Med*

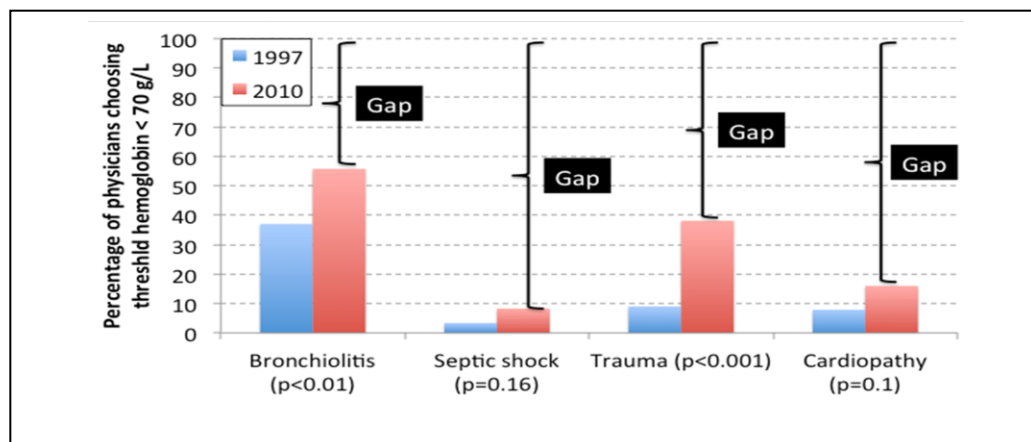


Figure 4.3 Écart entre la pratique réelle et les évidence ((Du Pont-Thibodeau *et al.*, 2014).

en 2007 (Lacroix *et al.*, 2007). Cependant, l'enquête menée en 2010 (Du Pont-Thibodeau *et al.*, 2014) a aussi démontré qu'il persistait un écart important entre les connaissances acquises grâce à l'étude TRIPICU et leur application (Figure 4.3).

De plus, plusieurs études épidémiologiques faites après l'année 2010 ont démontré que les praticiens se posaient encore des questions vis-à-vis des patients plus ou moins gravement malades que ceux qui avaient été recrutés dans l'étude TRIPICU. On peut donc s'attendre à ce que l'application des recommandations du groupe TAXI se fasse lentement, d'où l'intérêt de recourir à la fois à un outil d'aide à la décision médicale qui intègre un accès aux références les plus pertinentes.

## 4.9 Projet de démonstration

### 4.9.1 Objectifs

L'objectif principal de notre programme de recherche est d'améliorer la pratique médicale concernant la transfusion de globules rouges à des enfants traités en réanimation par les évidences (Tableau 4.1). L'étude évalue le niveau d'adhésion des praticiens au moment de leur prise de décision aux suggestions contenues dans l'arbre de décision proposé par le groupe TAXI, un consensus d'experts médicaux dans le domaine (Figure 4.2) ; l'aspiration est que l'adhésion à ces recommandations soit augmentée par un accès immédiat à des résumés d'articles scientifiques de référence. Ce travail donne l'occasion d'apprécier l'incidence d'un éventuel système d'aide à la décision médicale qui fournirait un accès facile, rapide et automatique aux prescripteurs de transfusion de globules rouges à des résumés d'articles scientifiques particulièrement pertinents.

Les articles de références ont été sélectionnés manuellement suivant leur pertinence et leurs liens avec les recommandations de TAXI. Notre perspective est de pouvoir doter un SADC de mécanismes (voir les chapitres précédents) capables retrouver ces articles de références d'une façon automatique, et ce, en se servant d'une formulation PICO adéquate (cas optimal qui concorde avec les scénarios) et des règles de décisions imbriquées dans un SADC. D'un autre côté, les résultats de cette étude permettent de vérifier d'un point de vue clinique la disparité (ou non) des décisions prises par des praticiens devant faire face à des situations de transfusion pédiatriques similaires à ce qu'ils peuvent rencontrer dans leur pratique clinique quotidienne ou occasionnelle.

Tableau 4.2 Objectifs de l'étude

Objectifs	Critères d'évaluation
Objectif principal	
Évaluer la conformité des décisions des praticiens dans la prise en charge transfusionnelle des patients pédiatriques à la suite de l'exposition (ou non) à des résumés d'articles pertinents de références.	<p>Proportion de prescriptions de transfusion de globules rouges conformes aux recommandations du groupe TAXI.</p> <p>Pour chaque scénario les participants auront à répondre par « en fonction de différents seuils (concentration hémoglobine variant de 50 à 130 g/l) », « pas de transfusion » ou « je ne sais pas ».</p> <p>Numérateur = l'ensemble des réponses conformes à TAXI des participants.</p> <p>Dénominateur = l'ensemble de scénarios multiplié par le nombre de participant.</p> <p>Dans les deux groupes</p>
Objectifs secondaires	
<p>Vérifier si les recommandations de TAXI renforcées par des résumés de références ont un impact positif sur la prise de décision.</p> <p>Déterminer la proportion et les caractéristiques des praticiens réfractaires et comprendre les raisons de leur résistance (condition médicale, risque ou conviction divergente).</p> <p>Comprendre le processus décisionnel pour éventuellement améliorer les outils d'aide à la décision en les alimentant avec de nouvelles publications scientifiques en lien avec les affections médicales du patient.</p>	<p>Proportion de participants ayant consulté les résumés d'articles.</p> <p>Numérateur= nombre de réponses positives.</p> <p>Dénominateur = nombre de participants.</p> <ul style="list-style-type: none"> <li>• J'ai lu à l'un article</li> <li>• Je n'ai pas lu d'articles</li> <li>• Je n'ai pas eu besoin de lire d'articles</li> </ul> <p>Dans le groupe « Intervention »</p> <p>Appréciation des participants de l'accès aux résumés lors de la prise de décision :</p> <p>Noter sur une échelle de 1 à 5 quel été l'apport de ces « abstracts » :</p> <p>Très pertinent (2) Pertinent (3) peu pertinent (4) pas du tout pertinent (5) ne sait pas / pas de réponse</p> <p>Impact des résumés lors de la prise de décision :</p> <p>Noter sur une échelle de 1 à 5 quel été l'apport de ces « abstracts » lors de la prise de décision :</p> <p>Décisif à la prise de décision (2) contribuant à la prise de décision (3) peu contribuant à la prise de décision (4) pas du tout contribuant à la prise de décision (5) ne sait pas / pas de réponse</p> <p>Dans le groupe « Intervention »</p>

#### 4.9.2 Hypothèse de base

L'accès à un outil d'aide à la décision et à l'information supplémentaire provenant de résumés d'articles scientifiques au moment de la prise de décision améliore la conformité des praticiens (indépendamment de leur nombre d'années d'expérience) aux recommandations préconisées présentement par des experts qui se sont basés sur les évidences pour les créer.

#### 4.9.3 Type d'étude

Nous avons réalisé une étude randomisée. Cette dernière est un type d'étude scientifique utilisé souvent en médecine pour tester et comparer l'efficacité de plusieurs approches thérapeutiques ou d'interventions dans une population donnée de patients. Les études randomisées doivent impliquer au moins deux groupes auxquels sont alloués les participants : un groupe expérimental (traitement ou intervention) et un groupe de référence (autre traitement ou placebo, par exemple). La procédure est accompagnée par des mesures spécifiques permettant de bien documenter les résultats qui serviront à évaluer le traitement ou l'intervention étudiée (voir le schéma de la figure 4.1).

Nous avons obtenu la liste de tous les soignants de patients admis aux soins intensifs pédiatriques du CHU Sainte-Justine, ce qui comprend : les moniteurs cliniques et médecins résidents en formation. Nous avons inclus tous les médecins traitants de l'hôpital Sainte-Justine membre des départements ou des services dont les patients sont à risque Population d'être hospitalisés aux soins intensifs. La liste d'envoi aux moniteurs cliniques et aux médecins résidents a été établie en collaboration avec le département universitaire en fonction à l'hôpital Sainte-Justine. La liste des médecins traitants a été fournie par le bureau du Conseil des médecins, dentistes et pharmaciens (CMDP) de l'hôpital Sainte-Justine.

Une invitation a été transmise par courriel aux participants par l'entremise d'un outil d'enquête clinique « REDCap<sup>47</sup> ». Tous les soignants éligibles ont été contactés afin d'éviter tout biais de sélection. Les participants qui ont accepté de participer ont été affectés d'une manière aléatoire à un des deux groupes de l'étude. Suivant le calcul de la taille d'échantillon (Section 2.11.2, voire ci-dessous) qui détermine un nombre requis 154 participant (répartis en deux groupes de 77 participants chacun).

#### 4.9.4 Critères d'inclusion

- Résidents et moniteurs cliniques figurant sur les listes fournies par les départements universitaires impliqués dans le soin d'enfants hospitalisés aux soins intensifs pédiatriques du CHU Sainte-Justine ;
- Médecins figurant sur la liste des médecins traitants pratiquant dans différents départements du CHU Sainte-Justine.

#### 4.9.5 Critères d'exclusion a priori

- Médecins retraités ou médecins n'ayant pas un statut de médecin actif au CHU Sainte-Justine.
- Externes ou stagiaires de recherche.
- Médecins ou résidents des services ou des départements suivants (ordre alphabétique) : biochimie, néonatalogie, microbiologie, radiologie, pathologie. Remarque : nous avons exclus ces départements parce que les médecins qui y travaillent ne prescrivent pas de transfusion de globules rouges (Ex. biochimie) ou encore, parce qu'ils traitent une population qui n'est pas visée par notre étude (ex. prématurés en néonatalogie).

---

47 REDCap : Logiciel pour la collecte de données en recherche clinique (<https://www.project-redcap.org/>)

- Personnel de recherche.
- Autres professionnels de la santé que des médecins (infirmières, pharmaciens, etc.).
- Fausse adresse courriel ou adresse courriel introuvable.

#### 4.9.6 Critères d'exclusion a posteriori

- Refus de participer.

#### 4.9.7 Recrutement et randomisation des sujets

- Un courriel d'invitation et d'explication a été envoyé dans un premier temps à tous les participants possibles afin de les informer sur l'étude en cours.
- Si le participant ne répond pas au courriel d'invitation, un rappel lui a été envoyé après un intervalle raisonnable.
- Si acceptation, alors une randomisation aurait été effectuée et un lien vers la version du logiciel de présentation des cas et d'interaction correspondant lui aurait été envoyé avec un accès garanti par son adresse courriel.

#### 4.9.8 Randomisation

La randomisation consiste à attribuer les participants, d'une façon aléatoire, afin d'éviter un biais de sélection et de permettre la comparaison de deux groupes strictement identiques pour des facteurs connus et inconnus. La seule différence entre les 2 groupes est liée à l'intervention. Dans cette étude, la randomisation a été réalisée juste après l'acceptation de l'invitation par les participants et elle a affecté la nature du questionnaire que le participant a reçu.

La liste de randomisation a été établie par un biostatisticien. Cette dernière a été établie avant le démarrage de l'étude et avant les inclusions, et ce, pour un nombre supérieur

au nombre total de participants estimé par le calcul de la taille d'échantillon (en prévention des perdus de vue). La randomisation a été établie avec un rapport (1 : 1) et aucune stratification n'a été faite.

#### 4.9.9 Intervention

Notre étude vise à évaluer l'impact d'un accès à des résumés d'articles de références au moment de la prise de décision. Cette évaluation de la pratique transfusionnelle se déroule auprès d'une population de praticiens provenant de différentes spécialités pédiatriques et ayant des années d'expérience variée.

#### 4.9.10 Scénarios

Les différents scénarios utilisés et approuvés par le comité d'éthique de sainte Justine sont présentés dans l'appendice « B » de cette thèse.

Pour chaque scénario les participants auront à répondre par « en fonction de différents seuils (concentration hémoglobine variant de 50 à 130 g/l) », « pas de transfusion » ou « je ne sais pas ». Scénario #1 : hémorragie massive ; Scénario #2 : drépanocytose ; Scénario #3 : bronchiolite ; Scénario #4 : polytraumatisme ; Scénario #5 : choc septique ; Scénario #6 : postopératoire de chirurgie non cardiaque ; Scénario #7 : tétralogie de Fallot ; Scénario #8 : postopératoire de chirurgie cardiaque, physiologie de cœur uni ventriculaire.

#### 4.9.11 Critères de jugement

##### 4.9.11.1 Critère de jugement principal

Les répondants des deux groupes ont été invités à répondre à une série de scénarios liés à des cas d'anémie justifiant d'envisager de prescrire une transfusion. Nous espérons documenter un changement de comportement qui tiendra compte du consensus TAXI qui prend lui-même en considération plusieurs déterminants du besoin de recevoir une



transfusion, comme la concentration d'hémoglobine. Pour chaque scénario les participants auront à répondre par :

- (1) *Transfusion en fonction de différents seuils (concentration hémoglobine variant de 50 à 130 g/l)*
- (2) *Pas de transfusion*
- (3) *Je ne sais pas.*

Pour répondre aux scénarios, les participants du groupe « intervention » ont accès aux suggestions faites par le consensus TAXI (Valentine *et al.*, 2018). Contrairement au groupe contrôle, le groupe « intervention » a aussi accès à un ensemble de résumés de références en lien avec ces scénarios, ce qui permet de vérifier si un accès rapide et facile à ces résumés influence la prise de décision de prescrire ou non une transfusion de globules rouges.

#### 4.9.11.2 Critères de jugement secondaires

Les critères de jugements secondaires porteront sur les trois points suivants :

- La proportion de participants ayant consulté les résumés d'articles au moment de la prise de décision. (Numérateur= NB de réponses positives, dénominateur = NB de participants).
  - (1) J'ai lu au moins un article
  - (2) Je n'ai pas lu d'articles
  - (3) Je n'ai pas eu besoin de lire d'articles

Dans le groupe « Intervention »

- Appréciation des participants de l'accès aux résumés lors de la prise de décision :

Noter sur une échelle de 1 à 5 quelle été l'apport de ces « abstracts » :

- (1) Très pertinent
- (2) Pertinent
- (3) Peu pertinent
- (4) Pas du tout pertinent
- (5) Ne sait pas / pas de réponse

- Impact des résumés lors de la prise de décision :

Noter sur une échelle de 1 à 5 quel été l'apport de ces « abstracts » lors de la prise de décision :

- (1) Décisif à la prise de décision
- (2) Contribuant à la prise de décision
- (3) Peu Contribuant à la prise de décision
- (4) Pas du tout Contribuant à la prise de décision
- (5) Ne sait pas / pas de réponse

Dans le groupe « Intervention »

#### 4.9.12 Gestion des données (*Data management*)

Trois types de données seront collectés dans le cadre de cette étude à différentes phases de l'enquête :

- (1) Données démographiques
- (2) Données liées à la transfusion de globules rouges en réanimation pédiatrique
- (3) Données liées à la prise de décision et à l'impact de l'accès à des résumés.

L'instrument d'enquête pour chaque groupe comporte une collecte d'informations démographiques concernant, entre autres, le type de pratique du praticien après son diplôme MD. Une collecte d'informations pour des réponses à une série de scénarios liés

à des cas d'anémie justifiant d'envisager prescrire une transfusion. De plus, et pour le groupe intervention, une collecte supplémentaire d'informations a été réalisée par des questions en lien avec la décision de prescrire une transfusion (au vu de la suggestion selon les recommandations de TAXI et les résumés en accompagnement).

Ces données résident sur un système de collectes de données « REDCap<sup>48</sup> » conçu spécialement pour la recherche clinique. Le suivi des données manquantes a été effectué en temps réel (à la saisie des données par le répondant) pour maximiser le nombre de réponses.

#### 4.9.13 Déroulement de l'étude

L'étude est une enquête par questionnaire (auto-administré) utilisant des scénarios prédéfinis et validés. L'enquête a été intégrée dans un projet de collecte de données créé à l'aide du logiciel « REDCap ». Les questionnaires ont été envoyés par courriel. Les participants randomisés au groupe expérimental ont eu droit à un accès immédiat à des résumés d'article jugés pertinents grâce à section intégrée au projet de collecte de données « REDCap » alors que les participants du groupe contrôle n'ont pas eu droit à un tel accès (Voir la figure 4.1 au début du chapitre).

##### 4.9.13.1 Préparation de l'étude

- Mise en place de l'entente avec la direction de l'enseignement du CHU Sainte-Justine pour récupérer les adresses courriel des moniteurs et des médecins résidents en formation.

---

<sup>48</sup> REDCap (*Research Electronic Data Capture*) est une solution logicielle d'acquisition électronique de données en recherche clinique. Il est largement utilisé dans le milieu de la recherche universitaire. REDCap a été développé par une équipe informatique à l'Université Vanderbilt avec l'appui continu de subventions du NCCR (*National Center for Research Resources*) et du NIH. <https://www.project-redcap.org/>

- Récupération des adresses courriel des médecins praticiens actifs du CHU Sainte-Justine sur la liste transmise par le secrétariat du collège des médecins.
- Rédaction des scénarios en s’inspirant de l’arbre de décision TAXI.

#### 4.9.13.2 Envoi et réception des questionnaires

Toutes les personnes comprises dans les différentes listes ont été invitées à répondre aux questionnaires. L’accès à ces derniers était précédé par des questions portant sur des données démographiques de base. Nous avons considéré que la participation d’un répondant était terminée une fois les dernières questions du formulaire complétées. Le suivi des participants a été faite sur la base de la détection du statut de la réponse de la participation aux courriels envoyés, que ce soit pour l’invitation à la participation ou pour l’enquête proprement dite. Ce suivi a été effectué sous la forme d’une relance automatique sur le courriel du participant pour l’inviter à compléter l’enquête (Tableau 4.3). La relance automatique par courriel (à des intervalles de 2 ou 3 semaines) est le moyen privilégié pour essayer de retenir les participants au sein de l’étude.

Tableau 4.3 Envoi et réception des questionnaires

Étapes	Dates ou fréquences envisagées
Envois des courriels d’invitations	Janvier 2020
Réception des réponses et randomisation	Au fur et à mesure
Relance pour les invitations	14 jours après 1er envoi et 28 jours après le deuxième envoi
Envoi de la série des questionnaires (groupe témoins et groupe intervention)	Automatique après chaque randomisation
Relance pour les questionnaires	14 jours si pas de réponses et 28 après de deuxièmes envois.

#### 4.9.14 Considérations statistiques

##### 4.9.14.1 Calcul de taille de l'échantillon

Une taille d'échantillon de 77 dans le groupe un (1) et de 77 dans le groupe deux (2) permettrait d'atteindre une puissance de 90 % pour détecter une différence absolue entre les proportions des deux groupes de 14 %. Les proportions estimées sont de 85 % et 99 % pour le groupe contrôle et intervention, respectivement. Nous avons fait nos analyses statistiques à l'aide du test de probabilité exacte de Fisher avec un niveau de signification bilatérale (valeur p) fixée à 0,05.

##### 4.9.14.2 Analyse statistique

Les données continues sont été présentées (section 4.10) par des moyennes et des écarts-types et les données catégoriques sont été présentées par des totaux et des fréquences (pourcentages). L'analyse principale de l'étude a été effectuée par le test exact de Fisher pour comparer les proportions entre le groupe d'intervention et le groupe contrôle. Toutes les analyses statistiques ont été réalisées avec le logiciel SAS 9,4 (Cary, NC, USA).

#### 4.9.15 Considérations éthiques

Nous avons considéré que les personnes qui répondent au questionnaire ont accordé leur consentement à participer à l'étude. Pour ce qui est des risques, cette étude ne présente aucun risque pour les participants. L'intervention se résume à une réflexion intellectuelle sur un ensemble de scénarios de cas cliniques susceptibles de survenir en temps normal lors de la pratique courante de la population concernée. Il se pourrait que l'accès à des résumés augmente les bénéfices du groupe « intervention ».

## 4.10 Résultats

### 4.10.1 Taux de réponse et caractéristiques démographiques des répondants

L'enquête a été réalisée entre janvier et mars 2020. Le taux de réponse fut de 49 % (75/154) (voir section 4.8.15, considération statistique). Sur les 75 répondants, 42 ont été randomisés dans le groupe intervention (avec un accès à des résumés de références) et 33 dans le groupe contrôle (sans accès aux résumés), cette randomisation semble débalancée et l'explication peut être attribuer au hasard, car le processus de randomisation et le suivi des envois des questionnaires a été vérifié par des parties tiers. Parmi ces répondants et dans les deux groupes, certains n'ont pas répondu à tous les scénarios présentés, ce qui n'était pas inattendu, car nous savions que certains répondants ne traitaient pas les patients décrits dans certains scénarios ; les réponses ont été prises en considération lors de l'analyse même quand il manquait des réponses à certains scénarios. La taille d'échantillon initialement requise pour l'enquête était de 77 individus par groupe, cette taille d'échantillon permettant de détecter une différence de 14% entre le groupe intervention et le groupe de référence avec une puissance de 90%. Avec 33 individus (groupe intervention) et 42 (groupe référence) la puissance actuelle calculée a posteriori est de seulement de 59% et ne permet plus de détecter des différences de cet ordre (de 14%). Le tableau 4.4 présente les différentes caractéristiques démographiques des répondants selon leur groupe de randomisation ; caractéristiques démographiques. La proportion des répondants selon leur genre est presque identique dans les deux groupes. La plupart des répondants sont francophones et les tranches d'âge sont comparables. La pédiatrie est la spécialité de la plupart des répondants dans les deux groupes. Pour ce qui est du rang académique, les répondants sont majoritairement des professeurs (adjoint, agrégé ou titulaire). Dans les deux groupes. Plus de 80 % des participants n'ont pas assisté à un cours sur la transfusion durant l'année précédente.

Tableau 4.4 Caractéristiques des répondants par groupe de randomisation

Caractéristique des répondants		Groupe 1 † (intervention)	Groupe 2 † (groupe témoin)	Total
Nombre de répondants		42	33	75
Genre Q1	Femme	23 (54.7 %)	19 (57.6 %)	42
	Homme	19 (45.3 %)	14 (42.4 %)	33
Âge des répondants Q2	20-29	5 (11.9 %)	4 (12.1 %)	9
	30-39	10 (23.8 %)	13 (39.3 %)	23
	40-49	10 (23.8 %)	3 (9.1 %)	13
	50-59	10 (23.8 %)	7 (21.2 %)	17
	60 ans et plus	7 (16.6 %)	6 (18.1 %)	13
Langue principale Q3 — Q4	Anglais	5 (11.9 %)	2 (6.1 %)	7
	Français	37 (88.1 %)	31 (93.9 %)	68
Spécialité principale (autre que les soins intensifs) Q5— Q6	Anesthésie		1 (3.0 %)	1
	Chirurgie cardiaque		1 (3.0 %)	1
	Chirurgie générale (+ la chirurgie pédiatrique)	1 (2.38 %)	2 (6.1 %)	3
	Pédiatrie	24 (57.14 %)	14 (42.42 %)	38
	Pneumologie		1 (3.03 %)	1
	Autre	17 (40.4 %)	14 (42.4 %)	31
Rang académique Q7	Professeur titulaire	4 (9.5 %)	7 (21.21 %)	11
	Professeur plein temps universitaire	1 (2.4 %)		1
	Professeur agrégé	13 (30.9 %)	6 (18.1 %)	19
	Professeur adjoint	15 (35.7 %)	12 (36.3 %)	27
	Chargé d'enseignement		1 (3.0 %)	1
	Chef de clinique	1 (2.3 %)	3 (9.1 %)	4
	Résident en réanimation		1 (3.0 %)	1
	Résident	8 (3.0 %)	3 (9.1 %)	11
Année d'obtention du diplôme de médecin (M.D) Q8	Avant 1970	1 (2.3 %)		1
	1970-1979	4 (9.5 %)	5 (15.2 %)	9
	1980-1989	6 (14.2 %)	5 (15.2 %)	11
	1990-1999	8 (19.0 %)	4 (12.1 %)	12
	2000-2009	11 (26.2 %)	7 (21.2 %)	18
	>= 2010	12 (28.5 %)	12 (36.4 %)	24
Cours sur la transfusion /dernière année	Oui	7 (16.67 %)	3 (9.09 %)	10
	Non	35 (83.33 %)	30 (90.91 %)	65

#### 4.10.2 Déterminants de la transfusion de globule rouge par scénario (groupe intervention et groupe contrôle).

L'analyse des réponses obtenues auprès des répondants a été effectuée en comparant les seuils d'hémoglobine choisis par les répondants en-deçà duquel ils prescriraient une transfusion au(x) seuil(s) recommandé(s) par TAXI (voir la section 4.8) pour chacun des scénarios :

- Le scénario #2 (*Drépanocytose*) porte sur un cas de drépanocytose, ce qui est particulier. La réponse correcte selon TAXI est de ne pas donner une transfusion si  $Hb \geq 120$  g/L. Toute transfusion en-deçà de ce seuil est acceptable.
- Le scénario #8 (chirurgie de *Norwood*) porte sur un cas de cardiopathie cyanogène. Il faut considérer que la réponse correcte est de ne pas donner de transfusion si l'Hb est  $\geq 90$  g/L. Toute transfusion en deçà de ce seuil est acceptable.
- Pour tous les autres scénarios, la réponse correcte est la même : pas de transfusion si  $Hb \geq 70$  g/L. Toute transfusion en deçà de ce seuil est acceptable.
- Par ailleurs, il faut transfuser dans tous les scénarios si  $Hb < 50$  g/L.
- Les réponses « je ne transfuse pas » et « je ne sais pas » ont été comptabilisées et considérées comme des réponses non correctes.

Dans le tableau 4.5 qui suit, nous présentons les réponses obtenues auprès de tous les répondants pour les différents scénarios (sans distinction du groupe d'appartenance). Ces informations sont complétées par la moyenne globale du seuil d'hémoglobine (ainsi que l'écart-type) choisi par les répondants pour prendre la décision de transfuser le patient pour chacun des scénarios.



Tableau 4.5 Seuil d'hémoglobine (Hb) dans chaque scénario.

Numéro du scénario	1	2	3	4	5	6	7	8
Conditions	<i>Hémorragie</i>	<i>Drépanocytose</i>	<i>Bronchiolite</i>	<i>Polytraumatisme</i>	<i>Choc septique</i>	<i>Post-opératoire de chirurgie non cardiaque</i>	<i>Tétralogie de Fallot</i>	<i>Norwood</i>
Description générale des réponses								
Répondants (n)	63	61	60	59	58	57	57	56
Nombre de répondants par seuil d'Hb :								
Hb < 50		1 2 %	2 3 %		1 2 %			
Hb < 60	1 2 %	5 8 %	7 12 %	1 2 %	2 3 %	3 5 %		1 2 %
Hb < 70	10 16 %	18 30 %	34 57 %	32 54 %	15 25 %	28 49 %	17 30 %	2 4 %
Hb < 80	8 13 %	7 11 %	4 7 %	5 8 %	8 13 %	5 9 %	8 14 %	4 7 %
Hb < 90	15 24 %	5 8 %	1 2 %	7 12 %	15 25 %	7 12 %	14 25 %	16 29 %
Hb < 100	11 17 %	5 8 %		3 5 %	7 12 %	4 7 %	4 7 %	12 21 %
Hb < 110	3 5 %	1 2 %		1 2 %				3 5 %
Hb < 120	2 3 %				1 2 %		1 2 %	2 4 %
Hb < 130	4 6 %							2 4 %
Pas de transfusion		10 16 %	8 13 %	1 2 %	2 3 %	3 5 %	2 4 %	
Je ne sais pas	9 14 %	9 15 %	4 7 %	9 15 %	7 12 %	7 12 %	11 19 %	14 25 %
Hb moyenne (écart-type)	88.4 (14.0)	76.9 (13.7)	69.0 (6.9)	76.3 (10.9)	82.2 (13.6)	76.0 (11.0)	82.0 (11.9)	93.3 (12.3)

Le tableau 4.6 présente plus en détail les choix du seuil de transfusion par groupe et par scénario. Dans la rangée du bas de ce tableau, nous rapportons la « valeur  $p$  » de la comparaison de l'impact de la présence ou de l'absence d'exposition des répondants aux articles de références. Le premier constat est que l'impact est significatif pour le scénario #8 d'une façon claire ( $p=0,004$ ). D'autres cas présentent une tendance comme les cas #2 ( $p=0.21$ ) et #5 ( $p=0.18$ ).

Tableau 4.6 Seuil d'hémoglobine (Hb) sélectionné dans chaque scénario par les deux groupes (Khi-2 ou Test Exact de Fisher).

Numéro du scénario	1		2		3		4		5		6		7		8	
Conditions	Hémorragie		Drépanocytose		Bronchiolite		Polytraumatisme		Choc septique		Post-opératoire de chirurgie non cardiaque		Tétralogie de Fallot		Norwood	
Groupe 1*	34		33		32		32		31		31		31		31	
Groupe 2*	29		28		28		27		27		26		26		25	
Total	63		61		60		59		58		57		57		56	
	Gr 1	Gr 2	Gr 1	Gr 2	Gr 1	Gr 2	Gr 1	Gr 2	Gr 1	Gr 2	Gr 1	Gr 2	Gr 1	Gr 2	Gr 1	Gr 2
Hb < 50 g/L			1 (3.0 %)		2 (6.2 %)				1 (3.2 %)							
Hb < 60 g/L	1 (2.9 %)		4 (12.1 %)	1 (3.5 %)	5 (15.6 %)	2 (7.1 %)		1 (3.7 %)	2 (6.4 %)			3 (11.5 %)				1 (4 %)
Hb < 70 g/L	5 (14.7 %)	5 (17.2 %)	9 (27.3 %)	9 (32.1 %)	17 (53.1 %)	17 (60.7 %)	19 (59.3 %)	13 (48.1 %)	9 (29.0 %)	6 (22.2 %)	17 (54.8 %)	11 (42.3 %)	11 (35.4 %)	6 (23.0 %)	2 (6.4 %)	
Hb < 80 g/L	4 (11.8 %)	4 (13.7 %)	4 (12.1 %)	3 (10.7 %)	1 (3.1 %)	3 (10.7 %)	3 (9.3 %)	2 (7.4 %)	4 (12.9 %)	4 (14.8 %)	4 (12.9 %)	1 (3.85 %)	7 (22.5 %)	1 (3.8 %)	3 (9.6 %)	1 (4 %)
Hb < 90 g/L	9 (26.4 %)	6 (20.6 %)	3 (9.0 %)	2 (7.1 %)	1 (3.1 %)		5 (15.6 %)	2 (7.4 %)	7 (22.5 %)	8 (29.6 %)	3 (9.6 %)	4 (15.3 %)	4 (12.9 %)	10 (38.4 %)	13 (41.9 %)	3 (12 %)
Hb < 100 g/L	5 (14.7 %)	6 (20.6 %)	3 (9.0 %)	2 (7.1 %)			2 (6.2 %)	1 (3.7 %)	3 (9.6 %)	4 (14.8 %)	2 (6.4 %)	2 (7.6 %)	3 (9.6 %)	1 (3.8 %)	3 (9.6 %)	9 (36 %)
Hb < 110 g/L	3 (8.8 %)		1 (3.0 %)					1 (3.7 %)							2 (6.4 %)	1 (4 %)
Hb < 120 g/L	2 (5.8 %)								1 (3.2 %)				1 (3.2 %)		2 (6.4 %)	
Hb < 130 g/L	1 (2.9 %)	3 (10.3 %)													1 (3.2 %)	1 (4 %)
Pas de transfusion			4 (12.1 %)	6 (21.4 %)	5 (15.6 %)	3 (10.7 %)		1 (3.7 %)	2 (6.4 %)		2 (6.4 %)	1 (3.8 %)	1 (3.2 %)	1 (3.8 %)		
Je ne sais pas	4 (11.7 %)	5 (17.2 %)	4 (12.1 %)	5 (17.8 %)	1 (3.1 %)	3 (10.7 %)	3 (9.3 %)	6 (22.2 %)	2 (6.4 %)	5 (18.5 %)	3 (9.6 %)	4 (15.3 %)	4 (12.9 %)	7 (26.9 %)	5 (16.1 %)	9 (36 %)
Valeur p	0.97		0.21		0.54		0.56		0.18		0.94		0.31		<b>0.004</b>	

Groupe de randomisation (1) : exposition automatique à la littérature médicale. Groupe de randomisation (2) : pas d'exposition automatique à la littérature médicale.

Dans le tableau 4.7, nous présentons le résultat de l'analyse des réponses par rapport au seuil de transfusion de référence de « 70 g/L » d'Hb (en excluant les réponses « je ne transfuse pas » et « Je ne sais pas »). La lecture de ce tableau fait ressortir le même constat que le tableau précédant (4.4) avec une valeur significative ( $p=0.016$ ) pour le scénario #8 avec cette nouvelle variable basée sur le seuil de référence (seuil de Hb  $>70$  et Hb  $< 70$ ). La valeur-p n'est pas significative pour les autres scénarios.

Le tableau 4.8 présente les résultats d'une analyse faite en tenant compte d'une nouvelle variable qui regroupe les répondants selon le choix de « transfuser » versus le choix de « non transfuser » (ceci exclut les réponses du type : « je ne sais pas »). Si on tient compte de cette variable, aucune différence n'est constatée entre les deux groupes (pour l'ensemble des scénarios). L'accès aux articles n'influence pas le choix de transfuser ou pas. Toutefois, on observe une tendance pour les cas #4, #7 et #8 qui sont très proche du seuil de confiance ( $\alpha=0,05$ ) avec une valeur p respectivement de 0,09, 0,18 et 0,09

Tableau 4.7 Seuil d'hémoglobine choisi en fonction du seuil recommandé par TAXI pour les affections des différents scénarios et en fonction de la randomisation (groupe 1 / groupe 2)

Scénario	1		2		3		4		5		6		7		8	
Condition	Hémorragie		Drépanocytose		Bronchiolite		Polytraumatisme		Choc septique		Post-opératoire non cardiaque		Tétralogie de Fallot		Norwood	
Hb (g/L)	Gr 1	Gr 2	Gr 1	Gr 2	Gr 1	Gr 2	Gr 1	Gr 2	Gr 1	Gr 2	Gr 1	Gr 2	Gr 1	Gr 2	Gr 1	Gr 2
Hb < 70	6 20.0%	5 20.8%			24 92.3%	19 86.4%	19 65.5%	19 86.4%	12 44.4%	6 27.2%	17 65.3%	14 66.7%	11 42.3%	6 33.3%		
Hb > 70	24 80.0%	19 79.2%			2 7.7%	3 13.6%	10 34.5%	6 30.0%	15 55.7%	16 72.7%	9 34.6%	7 33.3%	15 57.6%	12 66.7%		
Valeur p	0.94				0.50		0.74		0.21		0.93		0.55			
	Hb < 120		25 100%	17 100%									Hb < 90		18 69.2%	5 31.2%
	Hb > 120		0 0%	0 0%									Hb > 90		8 30.7%	11 68.7%
	Valeur p		NA										Valeur p		<b>0.016</b>	

Hb : hémoglobine. Groupe de randomisation (1) : exposition automatique à la littérature médicale. Groupe de randomisation (2) : pas d'exposition automatique à la littérature médicale.

Tableau 4.8 Seuil Hb (choisi) vs (non choisi) par groupe.

Numéro du scénario	1		2		3		4		5		6		7		8	
Condition	Hémorragie		Drépanocytose		Bronchiolite		Polytraumatisme		Choc septique		Post-opératoire non cardiaque		Tétralogie de Fallot		Norwood	
	Gr 1	Gr 2	Gr 1	Gr 2	Gr 1	Gr 2	Gr 1	Gr 2	Gr 1	Gr 2	Gr 1	Gr 2	Gr 1	Gr 2	Gr 1	Gr 2
Hb choisi	30 88.2%	24 82.8%	25 75.8%	17 60.7%	26 81.2%	22 78.5%	29 90.6%	20 74.0%	27 87.1%	22 81.5%	26 83.9%	21 80.8%	26 83.9%	18 69.2%	26 83.9%	16 64.0%
Hb pas choisi	4 11.8%	5 17.2%	8 24.2%	11 39.2%	6 18.7%	6 21.5%	3 9.4%	7 26%	4 12.9%	5 18.5%	5 16.1%	5 19.2%	5 16.1%	8 30.8%	5 16.1%	9 36.0%
Valeur p	0.54		0.21		0.80		0.09		0.56		0.76		0.19		0.09	

Le tableau 4.9 présente la moyenne d'hémoglobine choisie par chaque groupe pour chacun des scénarios. Il ressort de cette analyse qu'aucune différence significative n'est observée concernant la moyenne d'hémoglobine globale des deux groupes.

Tableau 4.9 Hémoglobine moyenne et écart-type par groupe.

Numéro du scénario	1	2	3	4	5	6	7	8
Conditions	Hémorragie	Drépanocytose	Bronchiolite	Polytraumatisme	Choc septique	Post-opératoire de chirurgie non cardiaque	Tétralogie de Fallot	Norwood
Groupe 1 Moyenne (Écart-type) (g/L)	90.0 (15.6)	76.8 (15.2)	67.7 (8.15)	76.6 (10.1)	80.3 (15.3)	76.1 (9.8)	81.1 (13.0)	92.4 (12.7)
Groupe 2 Moyenne (Écart-type) (g/L)	86.2 (11.6)	77.0 (11.6)	70.4 (4.8)	76.0 (12.3)	84.5 (11.0)	75.7 (12.5)	83.3 (10.3)	94.7 (11.9)
Valeur p (1 vs. 2)	0.35	0.95	0.17	0.86	0.29	0.89	0.56	0.58

#### 4.10.3 Déterminants pour la transfusion de globule rouge suivant les caractéristiques démographiques et académiques des répondants (analyses exploratoires).

Les tableau 4.10, tableau 4.11 et tableau 4.12 présentent le résultat de l'analyse de la corrélation entre les différentes catégories de répondants pour chaque caractéristique démographique. Les plus importants constats concernent :

- L'âge des répondants : la catégorie (B) 20-29 ans et la catégorie (D) 40-49 ans influencent significativement ( $p = 0.033$ ) le seuil d'hémoglobine pour la transfusion.
- Certaines spécialités des répondants influencent la moyenne d'hémoglobine choisie (Anesthésie et pédiatrie  $p = 0.012$  ; Anesthésie et pneumologie  $p = 0.021$ ) indépendamment du groupe d'appartenance.

- Le rang académique apparaît aussi comme un élément qui influence le seuil d'hémoglobine utilisé pour la décision de transfuser. Un croisement entre les différentes catégories est présenté dans le tableau 4.10, notamment entre les professeurs et les résidents, ainsi que les chargés de cours et les professeurs.
- L'année de diplomation aussi présente, un lien significatif notamment entre les diplômés d'avant 1970 et les autres plages d'années de diplomation (voir le tableau 4.11).

Tableau 4.10 Caractéristiques démographiques des répondants de chaque groupe par rapport au seuil d'hémoglobine moyen.

Caractéristiques des répondants		Groupe 1 † (inter) Hb (g/L)	Groupe 2 † (cont.) Hb (g/L)	Global Hb (g/L)	(Valeur <i>p</i> entre catégories)
Tous les répondants					
Genre Q1	Femme (F)	79.0	79.8	79.6 (76.9 - 82.4)	
	Homme (H)	81.3	81.5	81.2 (78.5 - 84.0)	
	P-value (F vs. H)			0.40	
Âge des répondants Q2	Moins de 20 ans (A)				
	20-29 (B)	80.9	86.3	83.8 (80.0 - 87.5)	<b>B vs D (<i>p</i> = 0.033)</b>
	30-39 (C)	83.7	80.0	82.2 (78.3 - 86.1)	
	40-49 (D)	79.2	70.9	77.7 (73.6 - 81.8)	
	50-59 (E)	78.7	80.3	79.7 (75.7 - 83.7)	
	60 ans et plus (F)	77.3	80.0	78.4 (73.3 - 83.6)	
Langue principale Q3 — Q4	Anglais (An)	74.0	80.6	77.5 (73.9 - 81.1)	
	Français (Fr)	80.8	80.5	80.7 (78.6 - 82.8)	
	Autre (Au)				
	P-value (An, Fr, Au)			0.12	
Spécialité principale (autre que les soins intensifs) Q5 — Q6	Anesthésie (A)		83.8	83.8 (81.5 - 86.1)	A vs Pe ( <i>p</i> =0.0206) A vs P ( <i>p</i> =0.0118)
	Chirurgie cardiaque (Cc)				
	Chirurgie générale*	76.7	80.7	80.5 (71.9 - 89.0)	
	Neurochirurgie (N)				
	Pédiatrie (P)	79.7	80.2	80.0 (77.7 - 82.3)	
	Pneumologie (Pe)		70.0	81.4 (78.4 - 84.4)	
	Autre (Au)	81.8	80.8	81.0 (76.7 - 85.2)	

(\*) y compris la chirurgie pédiatrique

Tableau 4.11 Seuil d'hémoglobine moyen par rapport au rang académique des répondants de chaque groupe

Caractéristique des répondants		Groupe 1 † (inter) Hb (g/L)	Groupe 2 † (cont.) Hb (g/L)	Global Hb (g/L)	Contrastes*(valeur <i>p</i> entre catégories)
Rang académique Q7	Professeur titulaire (Pt)	75.5	79.7	77.2 (72.3 - 82.0)	Pt vs Rd (0.046)
	Professeur plein temps universitaire (Pu)	73.8		73.8 (72.1 - 75.5)	Pu vs Rd (<.0001) Pu vs Rr (0.0004)
	Professeur agrégé (Pg)	78.8	77.5	79.1 (75.6 - 82.6)	Pg vs Pu (0.0059)
	Professeur adjoint (Pa)	82.9	79.5	81.3 (77.9 - 84.8)	Pa vs Pu (0.0002)
	Chargé d'enseignement (Ce)		83.8	83.7 (81.3 - 86.0)	Ce vs Cc (<.0001) Ce vs Pu (<.0001) Ce vs Pt (0.0158)
	Chef de clinique (Cc)		78.6	79.6 (77.2 - 82.0)	Cc vs Pt (0.0058) Cc vs Rd (<.0001)
	Résident en réanimation (Rr)		81.3	81.2 (78.8 - 83.5)	
	Résident (Rd)	82.0	88.4	83.9 (79.3 - 88.4)	
	Sans affiliation universitaire (Sa)				
Année d'obtention du diplôme de médecin (M.D) Q8	Avant 1970 (A)	68.8		68.6 (67.1—70.3)	
	1970-1979 (B)	73.8	79.5	77.3 (70.5 - 84.1)	B vs A (0.0180)
	1980-1989 (C)	81.3	76.2	79.5 (75.2 - 83.7)	C vs A (<.0001)
	1990-1999 (D)	76.7	83.8	79.4 (74.9 - 84.0)	D vs A (<.0001)
	2000-2009 (E)	80.8	77.7	79.9 (76.1 - 83.7)	E vs A (<.0001)
	>= 2010 (F)	83.2	83.5	83.4 (80.0 - 86.7)	F vs A (<.0001)

\*Seuls les contrastes significatifs au seuil nominal  $p=0.05$  sont reportés ici

Pour ce qui est du suivi d'un cours sur les transfusions dans l'année précédant l'administration du questionnaire, l'analyse montre que la moyenne d'hémoglobine globale (sans distinction du groupe) pour la transfusion ne présente aucune différence statistiquement significative ( $p=0.55$ ) (Tableau 4.12).

Tableau 4.12 Suivi du cours sur la transfusion et seuil d'hémoglobine moyen.

Caractéristique des répondants		Groupe 1 † (Inter.) Hb (g/L)	Groupe 2 † (Cont.) Hb (g/L)	Global Hb (g/L)
Cours suivi sur la transfusion ? Q9	Oui	84.1	77.5	80.9
	Non	79.4	81.0	81.7
	P-value (Oui vs Non)			0.55
† Groupe de randomisation 1 : exposition automatique à la littérature médicale ; groupe de randomisation 2 : pas d'exposition automatique à la littérature médicale.				



#### 4.10.4 Déterminants de la prise de décision du groupe avec articles de référence fournis (groupe intervention)

Trente et un (31) participants ont répondu au questionnaire de suivi sur les 42 participants randomisé dans le groupe intervention. Pour rappel, ce questionnaire avait comme objectif de sonder l'appréciation de ce groupe par rapport à l'accès aux résumés d'articles de références.

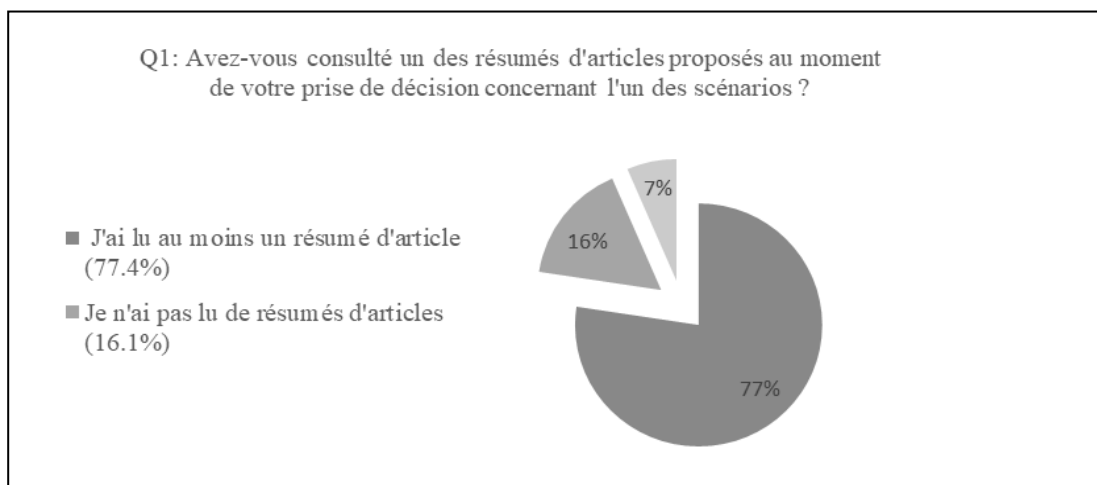


Figure 4.4 Informations sur la consultation des résumés d'articles.

La première question cherche à confirmer la lecture des résumés au moment de la prise de décision. La figure 4.4 ci-dessus nous donne un aperçu des appréciations avec une majorité de réponses favorables (77 %). Toutefois 16 % n'ont lu aucun des résumés. Les 6.7 % des répondants restant n'ont pas ressenti le besoin de lire ces résumés. Les questions suivantes furent adressées uniquement aux 24 répondants qui ont lus au moins un article.

La figure 4.5 résume les différents niveaux d'intérêt des répondants vis-à-vis des résumés présentés au moment de la prise de décision. Il en ressort que plus de 60 % considèrent que les articles sont intéressants (très intéressants, 3.6 % et intéressants, 57.1 %).

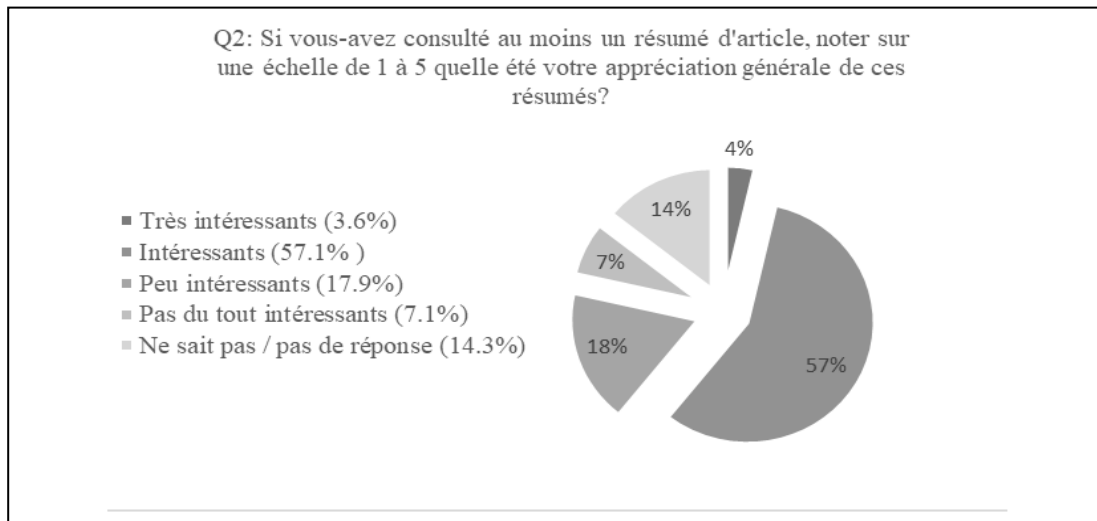


Figure 4.5 Appréciations générales de résumés fournis.

Concernant la troisième question qui cherche à avoir directement une réponse sur l'impact possible des résumés d'article sur la décision des participants, la figure 4.6 nous apprend que plus de 42 % pensent que les résumés contribuent à leur prise de décision, tandis que 32 % considèrent que l'impact est faible. Le reste des participants se divisent entre, ceux qui pensent que les résumés ne contribuent en rien à cette prise de décision (11%) et ceux (14 %) qui ne peuvent pas se prononcer.

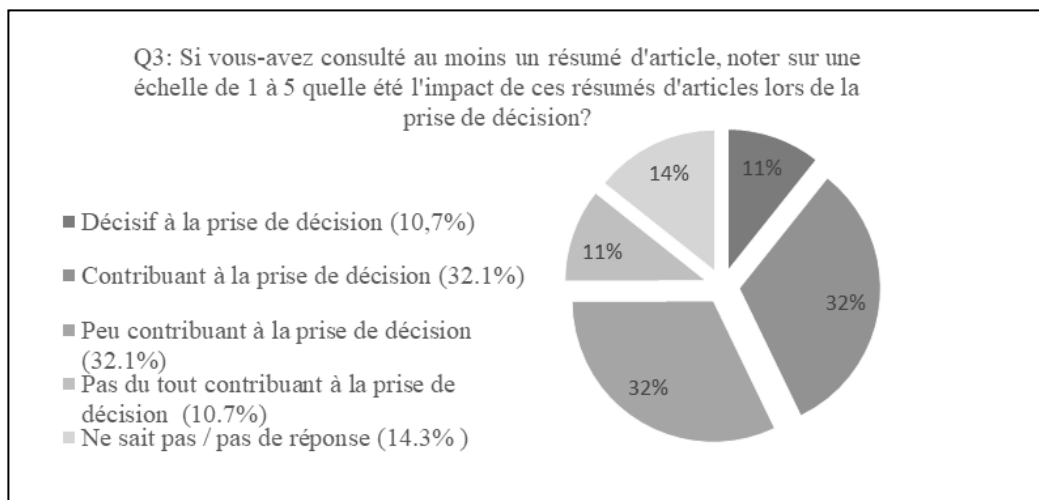


Figure 4.6 Impact des articles fournis sur la prise de décision.

Les résultats de l'analyse statistique portant sur la moyenne globale de la concentration d'Hb pré-transfusion en fonction des différentes catégories de réponses pour chaque question sont rapportés dans le tableau 4.13.

Tableau 4.13 Analyse entre les catégories de réponses au questionnaire de suivi et le seuil d'hémoglobine moyen du groupe d'intervention

Questions	Groupe inter- vention (répon- dants: n, %)	Hb moyenne (IC à 95 %) † (g/L)	Valeur <i>p</i>
Q1 : Avez-vous consulté un des résumés d'articles proposés au moment de votre prise de décision concernant l'un des scénarios ?			
Nombre de répondants à la question Q1 (N)		31	
1. J'ai lu au moins un résumé d'article	24 (57.1 %)	80.0 (78.0 - 82.0)	0.63
2. Je n'ai pas lu de résumés d'articles	5 (11.9 %)	78.1 (71.8 - 84.5)	
3. Je n'ai pas eu besoin de lire de résumés d'articles	2 (4.7 %)	81.7 (77.5 - 85.8)	
Q2 : Si vous avez consulté au moins un résumé d'article, noter sur une échelle de 1 à 5 quelle est votre appréciation générale de ces résumés :			
Nombre de répondants à la question Q2 (N)		24	
1. Très intéressants	1 (4.17 %)	79.4 (77.1 - 81.7)	0.63
2. Intéressants	16 (66.6 %)	82.8 (76.2 - 89.3)	
3. Peu intéressants	5 (20.8 %)	79.8 (75.7 - 83.9)	
4. Pas du tout intéressants	2 (8.3 %)	84.0 (75.0 - 92.9)	
5. Ne sait pas/pas de réponse			
Q3 : Si vous avez consulté au moins un résumé d'article, noter sur une échelle de 1 à 5 quelle été l'impact de ces résumés d'articles lors de la prise de décision :			
Nombre de répondants à la question Q3 (N)		24	
1. Décisif à la prise de décision	3 (12.5 %)	80.5 (77.5 - 83.5)	0.010
2. Contribuant à la prise de décision.	9 (37.5 %)	83.4 (78.3 - 88.5)	
3. Peu contribuant à la prise de décision	9 (37.5 %)	84.7 (80.0 - 89.4)	
4. Pas du tout contribuant à la prise de décision	3 (12.5 %)	76.1 (73.0 - 79.2)	
5. Ne sait pas/pas de réponse			

Malgré l'intérêt exprimé dans les réponses aux questions # 2 et #3, le fait d'être exposé aux résumés ne se reflète pas dans le seuil moyen d'Hb choisi effectivement par les répondants. Cependant et pour la question #3, nous avons trouvé une association significative entre les réponses des participants et le seuil restrictif (moyenne globale de l'Hb) des décisions prises pour transfuser ( $p$ -value = 0.010). Nous pouvons aussi remarquer que le groupe de répondants qui considère que les articles sont « Peu contribuant à la prise de décision » se caractérisent par la valeur la plus élevée du seuil de transfusion, ce qui pourrait refléter une certaine « décision acquise » sans prises en compte des évidences présentés.

#### 4.11 Discussion

Les résultats de cette enquête ne permettent pas d'observer une différence notable dans la décision de transfuser des participants, qu'ils soient exposés ou non à un accès rapide à des résumés, sauf pour le scénario #8 qui portait sur les soins post-opératoires d'un cas de chirurgie cardiaque présentant la physiologie d'un cœur uni ventriculaire. Cependant, l'enquête démontre qu'une forte proportion de répondants ont apprécié avoir cet accès. Une association significative existe entre l'appréciation positive de l'accès aux résumés (question #3 du questionnaire de suivi) et la moyenne du seuil d'hémoglobine observé de ces mêmes répondants.

L'enquête démontre qu'une proportion significative des médecins œuvrant au CHU Sainte-Justine continue de prescrire une transfusion de globules rouges au-delà de seuil d'hémoglobine recommandé (Tableau 4.5). Cela est particulièrement notable dans les cas de sepsis graves chez qui la majorité des répondants ont choisi un seuil  $> 80$  g/L alors que le seuil recommandé est de 70 g/L.

L'enquête démontre en outre que les caractéristiques académiques ainsi que l'année de diplomation des répondants influencent leur stratégie transfusionnelle sur le comportement restrictif des praticiens (Tableau 4.10, tableau 4.11 et tableau 4.12). Il apparaît dans l'enquête que les praticiens les plus expérimentés choisissent une stratégie restrictive de transfusion plus souvent que les plus jeunes qui devraient être plus à jour étant donné leur diplomation récente. Des questionnements peuvent être soulevés pour essayer de comprendre pourquoi les praticiens récemment diplômés sont « probablement » moins au courant des nouvelles recommandations dans le domaine de la transfusion que leurs aînés.

Plusieurs répondants prétendent ne pas avoir besoin de lire les résumés d'articles alors qu'ils prennent leur décision de prescrire une transfusion sur une concentration d'hémoglobine plus élevée que les autres répondants (Tableau 4.13). Autrement dit, il semble que ce soit les répondants qui en profiteraient le plus qui choisissent de ne pas consulter la littérature médicale. Il serait intéressant de comprendre pourquoi les répondants les plus déviants vis-à-vis les recommandations ne profitent pas de l'occasion qui leur est offerte pour mieux éclairer leur prise de décision, mais les données récoltées lors de l'enquête que nous avons menée et de la littérature ne permettent pas de découvrir la ou les raisons qui conduisent ces répondants à se comporter ainsi.

Plusieurs causes pourraient expliquer l'absence de différence quant à l'application des recommandations du groupe TAXI et à l'intérêt d'offrir un accès rapide à une littérature médicale solide, bien ciblée et très courte :

4. La puissance statistique de l'étude est peut-être insuffisante. En effet, le nombre de répondants avoisine la moitié de la taille d'échantillon nécessaire pour permettre une validation statistique des observations.
5. Le choix des répondants pourrait aussi être en cause. L'enquête a été envoyée aux médecins du CHU Sainte-Justine. Une grande partie des recommandations du groupe TAXI découle des résultats de l'étude TRIPICU (Lacroix *et al.*, 2007). L'étude TRIPICU a été faite à l'origine au CHU Sainte-Justine. Ses résultats sont donc mieux connus au sein de ce centre hospitalier qu'ailleurs. Une étude de cohorte publiée en 2014 démontre que la proportion de patients traités dans le service de soins intensifs pédiatriques du CHU Sainte-Justine est de 17 %, ce qui constitue la proportion la plus basse en Amérique du Nord (Demaret *et al.*, 2014). Nous avons choisi d'envoyer quand même l'enquête exclusivement aux médecins du CHU Ste-Justine pour des raisons de faisabilité : étant donné que les médecins du CHU Ste-Justine sont particulièrement sensibilisés à la médecine transfusionnelle, nous avons pensé que cela

augmenterait les chances qu'un nombre conséquent d'entre eux acceptent de remplir le questionnaire. En revanche, nous étions conscients qu'il y avait un risque qu'on ne puisse prouver qu'un accès rapide à des résumés améliore la conformité des décisions de transfuser ou non des cas de soins intensifs pédiatriques si les répondants connaissent et approuvent déjà les recommandations de l'étude TRIPICU. Les résultats du groupe contrôle de l'enquête démontrent effectivement une très grande conformité aux recommandations en question, ce qui diminuait d'autant nos chances de trouver une différence.

6. Certains médecins ont suivi un cours en médecine transfusionnelle quelques mois ou quelques années avant de répondre au questionnaire. Il se pourrait que ce cours ait été si efficace qu'il ait permis à ces répondants particulièrement studieux de bien savoir ce qu'il fallait répondre.
7. Il se pourrait finalement que l'absence de différence soit due au hasard.

Le nombre relativement petit de répondants, le manque possible de puissance de l'étude et le fait que l'étude ait été faite dans un seul centre hospitalier sont les principales faiblesses de l'étude. L'étude compte tout de même 75 répondants, ce qui n'est pas rien, mais le nombre de répondants que nous avions prévus recruter était de 140. Ce sont surtout les médecins des départements d'anesthésie et de chirurgie qui ont omis de répondre à l'enquête, ce qui suggère qu'ils se sentent moins concernés par la médecine transfusionnelle ; ces résultats suggèrent que ce sont peut-être les médecins appartenant à ces disciplines qu'il faudrait cibler avant tout si l'on veut améliorer la pratique transfusionnelle en milieu hospitalier pédiatrique. Il faut aussi reconnaître que l'étude manque probablement de puissance. Malgré tout, nous avons découvert que le fait de fournir un accès immédiat à un arbre de décision et à un résumé scientifique améliorerait

la pratique transfusionnelle dans un cas de soins post-opératoires de cœur uni ventriculaire ( $p=0.004$ ), un échantillon plus grand aurait peut-être permis de trouver une différence statistiquement significative dans les cas de drépanocytose ( $p=0.21$ ) et de soins post-opératoires d'un cœur à physiologie bi ventriculaire ( $p=0.18$ ). Finalement, il faut souligner que l'enquête a été menée dans un seul centre hospitalier, ce qui en limite l'aspect généralisable ; il faudrait refaire cette enquête auprès d'un certain nombre de centres hospitaliers communautaires et d'autres CHU afin de vérifier si l'accès rapide à des données scientifiques améliore ou non la pratique transfusionnelle pédiatrique.

#### 4.11.1 Aspect cognitif

Nous avons constaté que 16,6 % (Tableau 4.13) des médecins dont le taux de transfusion est élevé (81,7 et 71,8 g/L respectivement), et donc qui gagneraient à prendre connaissance des articles proposés par notre système ne se prévalent pas des propositions qui lui sont offertes par celui-ci. Il s'agit là d'une question associée à la prise de décision de la part des médecins. Pourquoi certains décident-ils de lire les articles proposés et d'autres non ? En particulier, pourquoi ceux qui gagneraient à lire les articles prennent-ils la décision de ne pas les consulter ? Si la prise de décision humaine se fonde sur des heuristiques comme nous l'avons présenté au début du chapitre (section 4.7), alors on devrait pouvoir déterminer, au moins comme hypothèse, quelles sont les heuristiques qui sont à l'œuvre dans cette prise de décision ? Cette hypothèse pourrait peut-être nous permettre de proposer des modifications éventuelles à notre algorithme pour augmenter la probabilité que les médecins qui ne lisent pas les articles suggérés les lisent davantage.

Rappelons que les heuristiques sont classées en quatre grandes catégories : 1) trop d'information, 2) pas assez de temps, 3) pas assez de signification pour déterminer l'importance et 4) impossibilité de tout mémoriser. Une étude complète du rôle des heuristiques entrant dans ces différentes catégories (Rappelons qu'il y a plus d'une centaine) dépasse les objectifs de cette thèse. L'heuristique de la disponibilité stipule que nous



aurons tendance à surestimer la probabilité d'une information immédiatement disponible en mémoire. Lorsqu'un médecin prend la décision de lire ou non un article proposé par le système, on peut faire l'hypothèse qu'un jeu de telles heuristiques a été impliqué. Pour comprendre exactement lesquelles sont de fait utilisées demanderait d'effectuer des expériences psychologiques. La nature de la situation de décision en salle d'urgence suggère cependant l'hypothèse que des heuristiques associées à la prise de décision se mettent en branle lorsqu'il y a une pression temporelle. En effet, les heuristiques suggèrent que, lorsque nous subissons des pressions temporelles, nous devons avoir le sentiment que ce que nous faisons est important, nous favorisons les choses immédiates et tangibles devant nous, nous avons tendance à compléter les choses dans lesquelles nous avons déjà investi du temps et de l'énergie, et enfin nous favorisons les options qui paraissent simples et l'information complète plutôt que les options ambiguës et complexes. Les modifications éventuelles à notre système devraient ainsi tenir compte de ces facteurs psychologiques, par exemple en n'offrant pas l'option de ne pas lire l'article, mais simplement de reporter la lecture à un moment où les pressions temporelles sont moindres (par exemple en envoyant l'article par courriel). Pour donner le sentiment que la lecture de l'article est importante, le courriel pourrait rappeler les méfaits de la « sur-transfusion ». Une stratégie complète, comme nous l'avons vu, devrait considérer toutes les heuristiques impliquées dans la prise de décision, le but étant de prendre cela en considération lors d'une prochaine itération du système.

#### 4.12 Conclusion

Suivant les résultats de cette enquête randomisée, un accès immédiat à des résumés d'articles de références au moment de la prise de décision a été apprécié par un bon nombre de répondants (57.1 %) (Tableau 4.13). Ces derniers ont pris la peine de les lire et de statuer par après sur la décision à prendre. Nous n'avons pas réussi à démontrer qu'un accès immédiat à une littérature pertinente améliorerait la stratégie transfusionnelle des répondants, mais cette absence de différence pourrait être attribuable à la faible puissance statistique et à la décision d'envoyer ce questionnaire exclusivement aux médecins du CHU Sainte-Justine, lesquels sont probablement les plus au courant des résultats de l'étude TRIPICU. Il serait intéressant de reprendre cette enquête en impliquant cette fois des médecins œuvrant en milieu hospitalier communautaire et dans d'autres CHU afin de savoir si un accès immédiat à des résumés améliore ou non la conformité à ces recommandations-là où les recommandations de TRIPICU et de TAXI sont moins connues. Il serait intéressant aussi, de réaliser une étude semblable, mais avec plus de questionnements à l'intention des praticiens qui n'ont pas ressenti le besoin de consulter les articles en particulier. Cette étude pourrait répondre à des aspects liés aux heuristiques et aux biais cognitifs qu'ils ont expérimentés lors de cette prise de décision. Nous pouvons aussi déceler les distractions possibles ainsi que d'autres facteurs comme les formats de présentation des résumés et l'impact de la qualité des sources et des revues de publication. Le résultat serait utilisé pour comprendre les spécifications de cette catégorie de professionnels et ainsi moduler les attentes et les objectifs pour une meilleure intégration de ce type de modèle pour la MBE.

## CONCLUSION

La médecine basée sur les évidences (MBE) est un paradigme qui fait la promotion d'une utilisation consciencieuse, explicite et judicieuse des meilleures preuves scientifiques disponible dans le domaine médical. L'intégration de ce paradigme favorise une meilleure prise de décision et une prise en charge clinique de cas individuels mieux éclairée. Cependant, l'intégration d'une MBE au moment de la prise de décision clinique n'est pas chose simple. Les praticiens qui font face à un problème/dilemme et qui nécessitent une alternative thérapeutique à la pratique habituelle ou une meilleure approche pour le bien-être du patient doivent être capables de trouver le temps nécessaire pour chercher de nouvelles évidences publiées, identifier les évidences les plus pertinentes pour le cas de son patient et analyser et juger la probable application de ces évidences.

De ce fait, la question principale à la base de ce travail est de savoir comment peut-on améliorer l'intégration de cette MBE au plus près de ces praticiens et au moment même de la prise de décision clinique ?

Exploitant l'avantage des systèmes d'aide à la décision qui sont des outils très proches des moments de prise de décision des praticiens, notre réponse est un modèle de système d'aide à la décision capable de faire une recherche d'information ciblée à partir des termes qui apparaissent dans les règles de décisions. Cette recherche permet la récupération d'articles scientifiques en lien avec le cas du patient traité, ce qui peut aider les praticiens à prendre des décisions éclairées. De ce fait, ce modèle fera passer ces

SDAC d'un mode de fonctionnement rigide vers une forme plus souple et plus ouverte sur les dernières publications et recommandations scientifiques dans le domaine.

Partant du questionnement général sur la possibilité d'améliorer l'intégration de la MBE par le modèle que nous avons proposé précédemment, plusieurs axes de recherche se sont avérés nécessaires pour parvenir à cet objectif.

Tout d'abord, il faut identifier des concepts pertinents pour cette recherche d'information à partir d'un ensemble restreint de termes de départs. Pour se faire, nous avons proposé une méthode et réalisé un outil générique pour l'exploration des termino-ontologies médicales d'UMLS. Cet outil s'appuie sur une approche innovante d'une marche aléatoire biaisée par la cooccurrence entre concepts dans un graphe ontologique. Ce processus permet l'identification des concepts pertinents au regard de leur centralité dans le voisinage termino-ontologique des termes de départs. Cette exploration se fera dans un rayon défini par un nombre de marches, une longueur de la marche (fenêtre), un choix des relations interconcepts et un seuil de co-occurrence des concepts dans les publications. Nous avons évalué la pertinence de cette approche en comparant le nombre de concepts du domaine médical identifié par cette dernière aux résultats d'une marche aléatoire d'exploration sans co-occurrence. À partir des résultats obtenus, il apparaît évident que l'utilisation de la cooccurrence pour choisir les prochains nœuds (concepts) à visiter augmente sensiblement l'efficacité d'identification des concepts pertinents par rapport à une marche totalement aléatoire.

La tâche suivante à attaquer est la recherche d'information proprement dite. Pour cela, nous avons élaboré une nouvelle approche qui conjugue l'extension de requête par de nouveaux concepts pertinents issus de l'exploration ontologique, une pondération qui reflète la centralité termino-ontologique des concepts et une discrimination qui exploite le paradigme PICO pour retrouver les publications scientifiques les plus pertinentes en recherche clinique. Pour évaluer nos approches, nous avons choisi d'utiliser des corpus de test de la TREC 2017 en oncologie pour confronter nos hypothèses à des sources de

jugements fiables et reconnues. À partir des résultats obtenus lors de nos expérimentations, il apparaît clairement que notre approche réalise la plus grande MAP avec l'extension de requête par les concepts retrouvés par l'ICPU par rapport une requête de base. Nos expérimentations ont démontré aussi qu'il est possible de mettre à contribution les concepts catégorisés comme PICO. Les résultats d'évaluation de l'ajout de ces concepts ont permis d'établir les meilleurs scores MAP de toutes nos expérimentations. De plus et à titre de comparaison, le classement de nos valeurs obtenues situe nos résultats au-dessus de la moyenne des moyennes des scores obtenus par les autres équipes participantes à ce concours. Cela reste encourageant au vu des possibilités d'amélioration de notre approche.

Le troisième et dernier axe de recherche concerne l'évaluation du modèle dans un milieu clinique en conditions réelles. Cette étude a été réalisée auprès des praticiens en s'inspirant du modèle de recherche clinique avec un protocole et une randomisation des participants. Ceci a permis d'étudier l'adhésion des médecins aux recommandations récentes en matière de transfusion si un accès leur a été fourni à des articles de référence au moment de la prise de décision. L'enquête a démontré qu'une forte proportion de répondants ont apprécié d'avoir cet accès. Une association significative existe entre l'appréciation positive de l'accès aux résumés et la moyenne du seuil d'hémoglobine observé de ces mêmes répondants.

Pour ce qui est de la facette cognitive de ce travail de recherche, celle-ci a été au centre de notre réflexion que ce soit pour l'utilisation des termino-ontologies, considérées comme une représentation des connaissances ou pour ce qui est de l'implémentation de la marche aléatoire qui prélude une possible application comme une la propagation d'activation à un niveau conceptuel. Les heuristiques et les biais cognitifs liés à la prise de décision ont été aussi attentivement pris en compte et reportés tout au long de l'enquête auprès des praticiens. Parmi les résultats obtenus, il ressort que certains nombres de médecins résistent aux propositions faites au moment de la prise de décision et qu'il

serait opportun d'appliquer la théorie des heuristiques et biais cognitifs pour identifier des méthodes qui pourraient aider à réduire ce nombre.

### Travaux futurs

Pour ce qui est de l'exploration termino-ontologique et l'identification des concepts pertinents, un prétraitement plus efficace des termes retrouvés dans les clauses des règles de décision et notamment les abréviations permettra de fournir une meilleure assise pour cette navigation et beaucoup plus de chance de retrouver des concepts pertinents au domaine. On peut aussi influencer l'exploration de concepts par une pondération des différents arcs du graphe au lieu de les considérer comme identiques (équivalant à 1). Cette pondération peut être ajusté dynamiquement suivant un retour d'appréciation des médecins envers les concepts pertinents retrouvés par ICPU. Ceci implémenterait une forme d'apprentissage qui permettrait, par exemple, d'ajuster le système aux préférences locales des médecins. Ceci pourrait donner lieu à une notion/valeur différente de la centralité des concepts retrouvés et impactera indéniablement le poids de ces derniers au moment de la recherche d'information.

Une autre amélioration intéressante serait de lancer la marche aléatoire sous la forme d'une « propagation d'activation » avec une certaine profondeur. Ceci reviendrait à faire la propagation au début à partir des concepts initiaux, lorsque la première cohorte de nouveaux concepts pertinents est sélectionnée une autre propagation est effectuée à partir de ce nouvel ensemble et ainsi de suite. La profondeur représente le nombre des différentes couches des nouveaux concepts pertinents successives. Dans ce cas de figure, la cooccurrence entre concepts n'est plus liée uniquement aux termes de départ, mais elle se déplacera elle aussi à chaque couche. L'utilisation des relations associatives pourrait permettre d'étendre efficacement cette propagation. L'objectif est de

faire avancer l'exploration et la découverte d'autres zones termino-ontologiques plus éloignées.

Pour ce qui est de l'axe de recherche d'information et sachant que les concepts candidats à l'extension de requête sont le résultat de l'exploration termino-ontologique, le choix des relations à explorer entre concepts peut être bonifié par l'ajout d'autres types de relations comme les relations (Hiérarchiques - synonymie) ou les relations (Hiérarchiques – associatives). Aussi, une présélection de types de relations (notamment les relations sémantiques) à explorer peut avoir un impact différent sur les résultats de recherche, et ce, en limitant l'exploration à un ensemble plus significatif de relations en lien avec une particularité comme une partie de l'anatomie du corps humain ou le diagnostic des maladies. En parallèle, une étude d'optimisation peut être effectuée sur le nombre de concepts à ajouter au moment de l'extension de la requête. Cette étude pourrait déterminer le nombre idéal de concepts pour parvenir à une plus haute MAP. Il est possible aussi d'optimiser la pondération utilisée pour booster les concepts de la requête candidate pour la recherche d'information en utilisant des techniques d'apprentissage à classer (*Learning to Rank*) (Deveaud *et al.*, 2018; Liu, T.-Y., 2011). Ceci revient à optimiser les paramètres dans les modèles de pondération, tels que BM25 ou PL2 à partir de paires requête-document grâce à un entraînement discriminant. Il serait intéressant d'examiner l'impact des nouvelles méthodes comme celle du modèle d'apprentissage en profondeur pour la reconnaissance des éléments PICO basée sur la mémoire bidirectionnelle à long terme bi-LSTM de Jin et Szolovits (2018) afin d'accroître la qualité de la classification des concepts suivant les catégorisations PICO.

Pour ce qui est de l'axe d'évaluation en milieu médicale, Il serait intéressant de reprendre cette enquête plus vaste en impliquant cette fois des médecins œuvrant en milieu hospitalier communautaire et dans d'autres CHU afin d'atteindre une puissance d'échantillon suffisante pour établir l'impact de ce type de modèle. Aussi, il serait intéressant de réaliser cette enquête, mais avec plus de questionnements à l'intention des

praticiens qui n'ont pas ressenti le besoin de consulter les articles en particulier. Cette étude pourrait répondre à des aspects liés aux heuristiques et aux biais cognitifs qu'ils ont expérimentés lors de cette prise de décision. Le résultat serait utilisé pour comprendre les spécifications de cette catégorie de professionnels et ainsi moduler les attentes et les objectifs pour une meilleure intégration de ce type de modèle pour la MBE.



## ANNEXE A

### PRÉSENTATION DU LOGICIEL ICPU

Le logiciel que nous avons conçu est mis en ligne à l'adresse (<https://www.icpu-umls.ml/>) pour une diffusion plus large. Le développement de cette application s'appuie sur quatre composantes principales :

- 1) Une Interface développée en PHP<sup>49</sup>, JavaScripts<sup>50</sup> et jQuery<sup>51</sup> ;
- 2) Une Base de données MySQL qui contient l'essentiel des tables d'UMLS utilisé (Taille = 40 Gig) ;
- 3) L'intégration du module MetaMapLite (voir les sections 2.9.2 et 2.9.3) ;
- 4) L'intégration d'une API<sup>52</sup> pour l'affichage des graphes résultats.

---

<sup>49</sup> PHP : Ce langage est principalement utilisé pour produire un site web dynamique. Il est courant que ce langage soit associé à une base de données, tel que MySQL. Exécuté du côté serveur (l'endroit où est hébergé le site) il n'y a pas besoin aux visiteurs d'avoir des logiciels ou plug-ins particuliers installés localement (sur leurs ordinateurs).

<sup>50</sup> JavaScript est un langage de programmation de scripts principalement employé dans les pages web interactives

<sup>51</sup> JQuery est une bibliothèque JavaScript libre et multiplateforme créée pour faciliter l'écriture de scripts côté client dans le code HTML des pages web

<sup>52</sup> Application programming interface (API ou Interface de programmation): est un ensemble normalisé de classes, de méthodes, de fonctions et de constantes qui sert de façade par laquelle un logiciel offre des services à d'autres logiciels. Elle est offerte par une bibliothèque logicielle ou un service web, le

## Interface graphique

L'interface logicielle de notre outil a été conçue dans un souci de permettre une intuitivité visuelle des utilisateurs pour comprendre rapidement le fonctionnement du logiciel et les différentes étapes pour l'exploration d'UMLS. L'ensemble des fonctionnalités est projeté sur un seul écran subdivisé en trois zones distinctes dans la figure A.1. À gauche : Première zone pour l'insertion des textes bruts et les boutons de commande pour lancer les différents processus de l'exploration ; 2) Au milieu : une deuxième zone pour afficher graphiquement les résultats et 3) à droite : une dernière zone dynamique

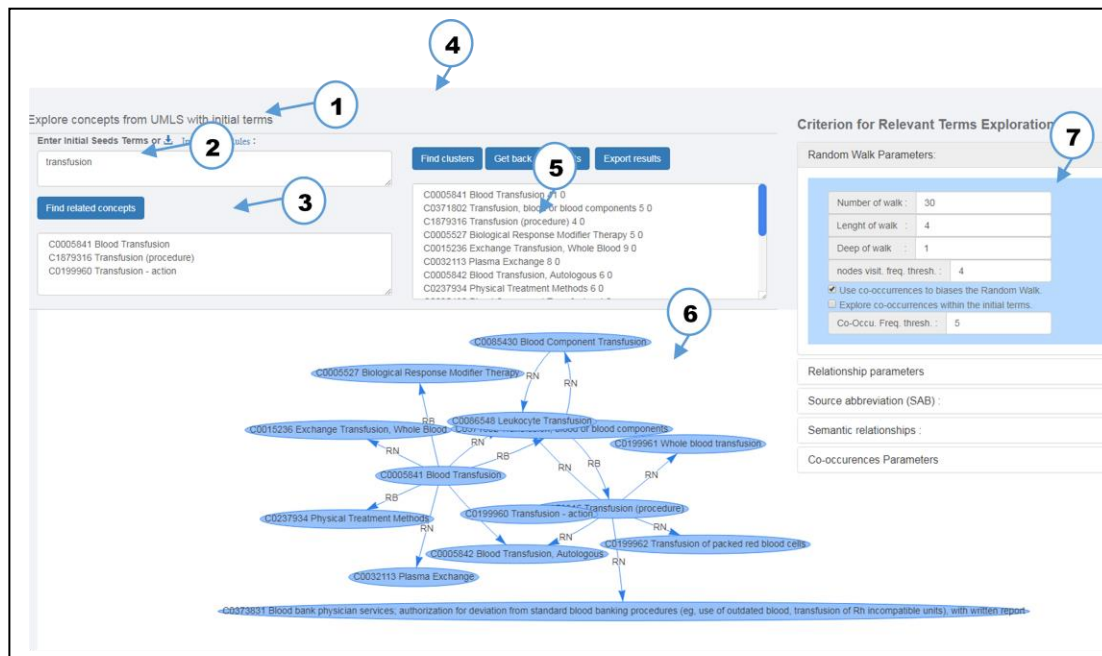


Figure A.1. Interface graphique (web) du logiciel ICPU

plus souvent accompagnée d'une description qui spécifie comment des programmes consommateurs peuvent se servir des fonctionnalités du programme fournisseur.

pour les menus de paramétrages de l'algorithme. La plus de détails sur les différents composants de chaque zone figure A.1 :

- 1) Zone texte pour écrire ou importer le texte des règles sous forme de « bag of words ».
- 2) Bouton pour lancer le traitement le « Mapping » de MetaMapLite.
- 3) Zone d'affichage des résultats du premier du traitement avec MetaMapLite. Le résultat est sous forme d'un tuple (code Cui, description du concept) ;
- 4) Bouton pour lancer la recherche des clusters à partir des concepts retrouvés dans la phase précédente et se trouvant dans la Zone (3).
- 5) Zone de texte pour afficher les résultats de l'algorithme de la marche aléatoire selon les paramètres dans le point (7). Ce résultat sera présenté sous forme de triplet (code Cui, nombre de visites et nombre de cooccurrences avec le concept de départ).
- 6) Zone d'affichage graphique pour les clusters de concepts résultant de l'exploration ainsi que les relations qui les lient.
- 7) Menu pour permettre le paramétrage de l'extraction : Ce menu permet aux utilisateurs de modifier plusieurs paramètres selon leur besoin (longueur de la marche, le seuil minimum de visite, le seuil de cooccurrence...).

### Intégration de MetaMap Lite

D'un point de vue technique, MetaMapLite est un répertoire qui comporte différents fichiers et programmes. Des fichiers de type « batch » (avec extension de fichier. bat pour Windows et. sh pour Linux), des fichiers d'indexation et les tables de références ainsi que les fichiers de classes Java et les fichiers Java compilés (les vrais exécutables). Deux tâches sont nécessaires pour pouvoir intégrer un tel programme (dans notre cas, c'est une intégration avec un langage de programmation différent : PHP-jQuery), 1) configurer le fichier de propriétés de MetaMapLite pour identifier les dictionnaires de

références allégés et les librairies nécessaires ; 2) fournir le texte initial à MetaMapLite et récupérer le résultat du traitement.

### Base de données MySQL

Nous avons utilisé une base de données MySQL, recommandée par UMLS afin de déployer les différentes tables utilisées dans notre application. La taille de cette base avoisine les 40 giga-octets. C'est une base de données relationnelle communautaire et en utilisation libre très stable et dotée de plusieurs fonctionnalités avancées (<https://www.mysql.com/fr/>).

### API (application *programming interface*) pour l'affichage des graphes résultats

Pour l'affichage graphique, nous avons intégré la bibliothèque JavaScript « Vis.js » qui est une bibliothèque de visualisation dynamique basée sur un navigateur. La bibliothèque est conçue pour gérer de grandes quantités de données dynamiques et permettre la manipulation de ces données (<https://visjs.org/>).

## ANNEXE B

### SCÉNARIOS UTILISÉS DANS L'ENQUÊTE (CHAPITRE VI).

#### Scénario #1 : hémorragie massive.

Un garçon de 2 mois est ventilé mécaniquement pour soigner une bronchiolite. Les paramètres du respirateur comprennent une fréquence respiratoire de 40 respirations/minute, une pression inspiratoire maximale (PIP) de 30 cm H<sub>2</sub>O et une valeur de PEEP de 2 cm H<sub>2</sub>O. Les autres paramètres du respirateur et les résultats de laboratoire sont rapportés précédemment. L'état neurologique et métabolique est normal, mais le garçon saigne abondamment et sa fréquence cardiaque augmente rapidement depuis quelques minutes.

#### Articles de références :

- (1) Consensus Recommendations for RBC Transfusion Practice in Critically Ill Children from the Pediatric Critical Care Transfusion and Anemia Expertise Initiative. (Pediatric Critical Care Medicine- 2018 Septembre).
- (2) Recommendations on RBC Transfusion in General Critically Ill Children Based on Hemoglobin and/or Physiologic Thresholds from the Pediatric Critical Care Transfusion and Anemia Expertise Initiative. (Pediatric Critical Care Medicine- 2018 September).

#### Scénario #2 : drépanocytose.

Un garçon âgé de 8 mois est ventilé mécaniquement après une chirurgie non cardiaque ; tout va bien, mais il est admis aux soins intensifs pour surveillance, car il souffre d'une anémie à hématies falciformes (drépanocytose). Les paramètres du respirateur comprennent une fréquence respiratoire de 40 respirations/minute, une pression inspiratoire maximale (PIP) de 30 cm H<sub>2</sub>O et une valeur de PEEP de 2 cm H<sub>2</sub>O. Les autres paramètres du respirateur et les résultats de laboratoire sont rapportés précédemment. L'état cardiovasculaire, neurologique et métabolique est normal. Actuellement, le garçon ne saigne pas.

#### Articles de références :

- (1) Consensus Recommendations for RBC Transfusion Practice in Critically Ill Children from the Pediatric Critical Care Transfusion and Anemia Expertise Initiative. (Pediatric Critical Care Medicine- 2018 Septembre)
- (2) Recommendations on Red Blood Cell Transfusion Support in Children with Hematologic and Oncologic Diagnoses from the Pediatric Critical Care Transfusion and Anemia Expertise Initiative (Pediatric Critical Care Medicine- 2018 September).

#### Scénario #3 : bronchiolite.

Un garçon de 2 mois est ventilé mécaniquement pour soigner une bronchiolite. Les paramètres du respirateur comprennent une fréquence respiratoire de 40 respirations/minute, une pression inspiratoire maximale (PIP) de 30 cm H<sub>2</sub>O et une valeur de PEEP de 2 cm H<sub>2</sub>O. Les autres paramètres du respirateur et les résultats de laboratoire sont rapportés précédemment. L'état cardiovasculaire, neurologique et métabolique est normal. Actuellement, le garçon ne saigne pas.

#### Articles de références :

- (1) Consensus Recommendations for RBC Transfusion Practice in Critically Ill Children from the Pediatric Critical Care Transfusion and Anemia Expertise Initiative. (Pediatric Critical Care Medicine- 2018 September)
- (2) Recommendations for red blood cell transfusions in critically ill children with acute respiratory failure from the Pediatric Critical Care Transfusion and Anemia Expertise Initiative (TAXI). (Pediatric Critical Care Medicine- 2018 September)
- (3) Transfusion strategies for patients in pediatric intensive care units. (New England Journal of Medicine 2007)

#### Scénario #4 : polytraumatisme.

Un garçon âgé de 2 mois est soumis à une ventilation mécanique après un polytraumatisme. Les paramètres du respirateur sont : fréquence respiratoire de 40 respirations/minute, pression inspiratoire maximale (PIP) de 30 cm H<sub>2</sub>O et une PEEP de 6 cm H<sub>2</sub>O ; les autres paramètres du respirateur et les résultats de laboratoire sont rapportés précédemment. La rate du patient est déchirée, mais ne saigne pas activement. Un pneumothorax est évacué par un drain thoracique. Le syndrome de détresse respiratoire aiguë (SDRA, ARDS en anglais) est stabilisé. Pression artérielle de 80/50 mmHg, fréquence cardiaque de 110/minute. Les 2 jambes sont fracturées. Le bébé ne saigne pas. Il reçoit midazolam, fentanyl, pancuronium et antibiotiques.

#### Articles de références :

- (1) Consensus Recommendations for RBC Transfusion Practice in Critically Ill Children from the Pediatric Critical Care Transfusion and Anemia Expertise Initiative. (Pediatric Critical Care Medicine- 2018 September)
- (2) Recommendations on RBC Transfusion in Critically Ill Children with Acute Brain Injury from the Pediatric Critical Care Transfusion and Anemia Expertise Initiative. (Pediatric Critical Care Medicine- 2018 September).

- (3) Recommendations on RBC Transfusion in General Critically Ill Children Based on Hemoglobin and/or Physiologic Thresholds from the Pediatric Critical Care Transfusion and Anemia Expertise Initiative. (Pediatric Critical Care Medicine- 2018 September)
- (4) Transfusion strategies for patients in pediatric intensive care units. (New England Journal of Medicine 2007)

#### Scénario #5 : choc septique.

Un garçon âgé de 2 mois bénéficie d'une ventilation mécanique pour traiter un choc septique et un syndrome de détresse respiratoire aiguë (SDRA, ARDS en anglais). Les paramètres du respirateur sont : fréquence respiratoire de 40 respirations/minute, pression inspiratoire maximale (PIP) de 30 cm H<sub>2</sub>O et PEEP de 6 cm H<sub>2</sub>O ; les autres paramètres du respirateur et de laboratoire sont rapportés précédemment. Pression artérielle de 45/25 mmHg, fréquence cardiaque de 140/minute. Une infusion de dopamine (20 µg/kg/minute) et de dobutamine (10 µg/kg/minute) sont administrés. Actuellement, le bébé ne saigne pas activement. Il reçoit midazolam, fentanyl, pancuronium et antibiotiques.

#### Articles de références :

- (1) Consensus Recommendations for RBC Transfusion Practice in Critically Ill Children from the Pediatric Critical Care Transfusion and Anemia Expertise Initiative. (Pediatric Critical Care Medicine- 2018 September)
- (2) Recommendations on RBC Transfusions for Critically Ill Children with No hemorrhagic Shock from the Pediatric Critical Care Transfusion and Anemia Expertise Initiative. (Pediatric Critical Care Medicine- 2018 September)
- (3) Transfusion strategies for patients in pediatric intensive care units. (New England Journal of Medicine 2007)

#### Scénario #6 : post-opératoire de chirurgie non cardiaque.

Un garçon de 2 mois revient de la salle d'opération après une transplantation hépatique. Les paramètres du respirateur comprennent une fréquence respiratoire de 40 respirations/minute, une pression inspiratoire maximale (PIP) de 30 cm H<sub>2</sub>O et une valeur de PEEP de 2 cm H<sub>2</sub>O. Les autres paramètres du respirateur et les résultats de laboratoire sont rapportés précédemment. L'état cardiovasculaire, neurologique et métabolique est normal. Actuellement, le garçon ne saigne pas.

#### Articles de références :

- (1) Consensus Recommendations for RBC Transfusion Practice in Critically Ill Children from the Pediatric Critical Care Transfusion and Anemia Expertise Initiative. (Pediatric Critical Care Medicine- 2018 September)
- (2) Recommendations on RBC Transfusion in General Critically Ill Children Based on Hemoglobin and/or Physiologic Thresholds from the Pediatric Critical Care Transfusion and Anemia Expertise Initiative. (Pediatric Critical Care Medicine- 2018 September)

- (3) Transfusion strategies for patients in pediatric intensive care units. (New England Journal of Medicine 2007)

#### Scénario #7 : tétralogie de Fallot.

Un garçon âgé de 2 mois arrive dans l'unité de soins intensifs après une chirurgie correctrice pour une tétralogie de Fallot. Les paramètres du respirateur incluent une fréquence respiratoire de 40 respirations/minute, une pression inspiratoire maximale (PIP) de 20 cm H<sub>2</sub>O et une PEEP de 3 cm H<sub>2</sub>O. D'autres paramètres du respirateur et des résultats de laboratoire sont rapportés précédemment. Une infusion de dopamine (5 µg/kg/minute) et une perfusion de dobutamine (10 µg/kg/minute) sont données. La fréquence cardiaque est de 140/min et la pression artérielle est de 75/50 mmHg. Actuellement, le garçon ne saigne pas. Il reçoit midazolam, fentanyl, pancuronium et antibiotiques.

#### Articles de références :

- (1) Consensus Recommendations for RBC Transfusion Practice in Critically Ill Children from the Pediatric Critical Care Transfusion and Anemia Expertise Initiative. (Pediatric Critical Care Medicine- 2018 September).
- (2) Recommendations on RBC Transfusion in Infants and Children with Acquired and Congenital Heart Disease from the Pediatric Critical Care Transfusion and Anemia Expertise Initiative. (Pediatric Critical Care Medicine- 2018 September).

#### Scénario #8 : post-opératoire de chirurgie cardiaque, physiologie de cœur uni ventriculaire.

Un nourrisson maintenant âgé de 2 semaines de vie vient d'arriver aux soins intensifs à la suite d'une chirurgie cardiaque (procédure de Norwood + Sano). Les paramètres du respirateur comprennent une fréquence respiratoire de 40 respirations/minute, une pression inspiratoire maximale (PIP) de 30 cm H<sub>2</sub>O et une valeur de PEEP de 2 cm H<sub>2</sub>O. La saturation (SaO<sub>2</sub>) varie de 70 % à 80 %. Il reçoit 20 PPM de NO, de l'épinephrine à 0,1 mcg/kg/min et de la millonine à 0,5 mcg/kg/hre.

#### Articles de références :

- (1) Consensus Recommendations for RBC Transfusion Practice in Critically Ill Children from the Pediatric Critical Care Transfusion and Anemia Expertise Initiative. (Pediatric Critical Care Medicine- 2018 September)
- (2) Recommendations on RBC Transfusion in Infants and Children with Acquired and Congenital Heart Disease from the Pediatric Critical Care Transfusion and Anemia Expertise Initiative. (Pediatric Critical Care Medicine- 2018 September)



Pour tous les scénarios, la question suivante a été posée :

Quelle concentration d'hémoglobine vous inciterait  $HB < 50$  g/L à transfuser le patient du scénario ?

Les réponses possibles sont :

- $HB < 50$  g/L
- $HB < 60$  g/L
- $HB < 70$  g/L
- $HB < 80$  g/L
- $HB < 90$  g/L
- $HB < 100$  g/L
- $HB < 110$  g/L
- $HB < 120$  g/L
- $HB < 130$  g/L
- Pas de transfusion
- Je ne sais pas



Le 16 janvier 2020

Docteur Jacques Lacroix

CHU Sainte-Justine

Objet	Autorisation de réaliser la recherche
	2020-2354 Impact de l'aide à la décision renforcée par un accès aux résumés d'articles de références au moment de la prise de décision: Une étude comparative dans le contexte de la médecine transfusionnelle pédiatrique. Cochercheurs: Redha El Taani; Pierre Poirier; Nancy Robitaille; Marisa Tucci; Petcko Valtchev

Bonjour,

Il nous fait plaisir de vous autoriser à réaliser la recherche identifiée en titre dans notre établissement et/ou sous ses auspices.

Cette autorisation vous est accordée sur la foi des documents que vous avez déposés auprès de notre établissement afin de compléter l'examen de convenance ainsi que la lettre du CER évaluateur. Si ce CER vous informe pendant le déroulement de cette recherche d'une décision négative portant sur l'acceptabilité éthique de cette recherche, vous devrez considérer que la présente autorisation de réaliser la recherche dans notre établissement est, de ce fait, révoquée à la date que porte l'avis du CER évaluateur.

Notre établissement a reçu une copie de la version finale des documents se rapportant à la recherche, approuvée par le CER évaluateur. Cette autorisation de réaliser la recherche suppose également que vous vous engagez :

- 1) à vous conformer aux demandes du CER évaluateur, notamment pour le suivi éthique continu de la recherche;

- 2) à rendre compte au CER évaluateur et à la signataire de la présente autorisation du déroulement du projet, des actes de votre équipe de recherche, s'il en est une, ainsi que du respect des règles de l'éthique de la recherche;
- 3) à respecter les moyens relatifs au suivi continu qui ont été fixés par le CER évaluateur;
- 4) à conserver les dossiers de recherche pendant la période fixée par le CER évaluateur, après la fin du projet, afin de permettre leur éventuelle vérification;
- 5) à respecter les modalités arrêtées au regard du mécanisme d'identification des sujets de recherche dans notre établissement, à savoir la tenue à jour et la conservation de la liste à jour des participants de recherche recrutés dans notre établissement. Cette liste devra nous être fournie sur demande.

La présente autorisation peut être suspendue ou révoquée par notre établissement en cas de non-respect des conditions établies. Le CER évaluateur en sera alors informé.

Vous consentez également à ce que notre établissement communique aux autorités compétentes des renseignements personnels qui sont nominatifs au sens de la loi en présence d'un cas avéré de manquement à la conduite responsable en recherche de votre part lors de la réalisation de cette recherche.

Je vous invite à entrer en communication avec moi pendant le déroulement de cette recherche dans notre établissement, si besoin est. Vous pouvez aussi contacter notre CER en vous adressant au Bureau de l'éthique de la recherche ([ethique@recherche-ste-justine.qc.ca](mailto:ethique@recherche-ste-justine.qc.ca), poste 4040). En terminant, je vous demanderais de toujours mentionner dans votre correspondance au sujet de cette recherche le numéro attribué à votre demande par notre établissement ainsi que le numéro attribué au projet de recherche par le CER évaluateur.

Veuillez accepter mes sincères salutations.



Marc Girard, M.D.  
Directeur des services professionnels (DAMU)

Personne formellement mandatée au CHU Sainte-Justine pour autoriser la réalisation des projets de recherche

## BIBLIOGRAPHIE

- Abadi, M. J. S. et Zamanifar, K. (2011). Producing complete modules in ontology partitioning IEEE.
- Abbes, S. B., Scheuermann, A., Meilender, T. et d'Aquin, M. (2012). Characterizing modular ontologies
- Abdou, S. et Savoy, J. (2007). Considérations sur l'évaluation de la robustesse en recherche d'information ARIA (Association Francophone de Recherche d'Information et Applications).
- Afzal, M., Hussain, M., Ali, T., Hussain, J., Khan, W. A., Lee, S. et Kang, B. H. (2015). Knowledge-based query construction using the CDSS knowledge base for efficient evidence retrieval. *Sensors*, 15(9), 21294-21314.
- Agirre, E., López de Lacalle, O. et Soroa, A. (2014). Random walks for knowledge-based word sense disambiguation. *Computational Linguistics*, 40(1), 57-84.
- Aitchison, J. (2012). *Words in the mind: An introduction to the mental lexicon* John Wiley & Sons.
- Aitchison, J., Bawden, D. et Gilchrist, A. (2003). *Thesaurus construction and use: a practical manual* Routledge.
- Alani, H. et Brewster, C. (2005). Ontology ranking based on the analysis of concept structures. *Proceedings of the 3rd international conference on Knowledge capture* (p. 51-58). 1088633 : ACM. doi: 10.1145/1088622.1088633

- Alani, H., Brewster, C. et Shadbolt, N. (2006). Ranking ontologies with AKTiveRank Springer.
- Amati, G., Carpineto, C. et Romano, G. (2004). Query difficulty, robustness, and selective application of query expansion Springer.
- Amati, G. et Van Rijsbergen, C. J. (2002). Probabilistic models of information retrieval based on measuring the divergence from randomness. ACM Transactions on Information Systems (TOIS), 20(4), 357-389.
- Anderson, J. R. (1983). A spreading activation theory of memory. Journal of verbal learning and verbal behavior, 22(3), 261-295.
- Aronson, A. R. et Lang, F.-M. (2010). An overview of MetaMap: historical perspective and recent advances. Journal of the American Medical Informatics Association, 17(3), 229-236.
- Aronson, A. R. et Rindflesch, T. C. (1997). Query expansion using the UMLS Metathesaurus American Medical Informatics Association.
- Arsenault, C. (2006). L'utilisation des langages documentaires pour la recherche d'information. Documentation et bibliothèques, 52(2), 139-148.
- Asim, M. N., Wasim, M., Khan, M. U. G., Mahmood, W. et Abbasi, H. M. (2018). A survey of ontology learning techniques and applications. Database, 2018.

- Aslam, J. A., Yilmaz, E. et Pavlu, V. (2005). A geometric interpretation of r-precision and its correlation with average precision
- Aslam, S. et Emmanuel, P. (2010). Formulating a researchable question: A critical step for facilitating good clinical research. *Indian journal of sexually transmitted diseases and AIDS*, 31(1), 47.
- Baeza-Yates, R. et Ribeiro-Neto, B. (1999). *Modern information retrieval* (vol. 463) ACM press New York.
- Baeza-Yates, R. et Ribeiro-Neto, B. (2011). *Modern Information Retrieval: the concepts and technology behind search* 2nd Ed. Harlow, England: Pearson Education Ltd.. Google Scholar Google Scholar Digital Library Digital Library.
- Bakken, S. (2001). An informatics infrastructure is essential for evidence-based practice. *Journal of the American Medical Informatics Association*, 8(3), 199-201.
- Bakken, S., Cimino, J. J. et Hripcsak, G. (2004). Promoting patient safety and enabling evidence-based practice through informatics. *Medical care*, 42(2), II-49-II-56.
- Balota, D. A. et Lorch, R. F. (1986). Depth of automatic spreading activation: Mediated priming effects in pronunciation but not in lexical decision. *Journal of Experimental Psychology: Learning, memory, and cognition*, 12(3), 336.
- Beckage, N. M. et Colunga, E. (2016). Language networks as models of cognition: Understanding cognition through language. Dans *Towards a theoretical framework for analyzing complex linguistic networks* (p. 3-28). Springer.

Bembea MM, Valentine SL, Bateman ST, Wilson LM, Anton B, Harger N, . . .  
Pediatric Critical Care Blood Research Network (BloodNet), a. t. P. A. L. I. a.  
S. I. P. N. (2018). The Pediatric Critical Care Transfusion and Anemia  
Expertise Initiative Consensus Conference Methodology. *Pediatr Crit Care  
Med*, 19(9S Suppl 1):S93-S97. doi: 10.1097/PCC.0000000000001593

Benjamin, D. et Gordon, H. G. (2017). Progress in evidence-based medicine: a  
quarter century on. *The lancet*.

Benkoussas, C. et Bellot, P. (2015). Information retrieval and graph analysis  
approaches for book recommendation. *The Scientific World Journal*, 2015.

Bernard, C. (1865). *Medecine Experimentale*. Paris: Bailliere et Fils.

Berner, E. S. (2007). *Clinical decision support systems* : Springer Science+ Business  
Media, LLC New York.

Bethesda, M. ( 2009 -a). UMLS® Reference Manual : Semantic Network. Récupéré  
de Available from: <https://www.ncbi.nlm.nih.gov/books/NBK9679/>

Bethesda, M. ( 2009 -b). UMLS® Reference Manual : SPECIALIST Lexicon and  
Lexical Tools. Récupéré de Available from:  
<https://www.ncbi.nlm.nih.gov/books/NBK9680/>

Bhogal, J., MacFarlane, A. et Smith, P. (2007). A review of ontology based query  
expansion. *Information processing & management*, 43(4), 866-886.

- Białecki, A., Muir, R., Ingersoll, G. et Imagination, L. (2012). Apache lucene 4
- Bodenreider, O. (2001). An object-oriented model for representing semantic locality in the UMLS. *Studies in health technology and informatics*, 84(0 1), 161.
- Boldi, P. et Vigna, S. (2005). MG4J at TREC 2005 Citeseer.
- Boudin, F., Nie, J.-Y., Bartlett, J. C., Grad, R., Pluye, P. et Dawes, M. (2010). Combining classifiers for robust PICO element detection. *BMC medical informatics and decision making*, 10(1), 29.
- Boudin, F., Nie, J.-Y. et Dawes, M. (2010). Clinical information retrieval using document and PICO structure *Association for Computational Linguistics*.
- Boudin, F., Shi, L. et Nie, J.-Y. (2010). Improving medical information retrieval with PICO element detection Springer.
- Bourdé, A., Cuggia, M., Ouazine, T., Turlin, B., Zékri, O., Bohec, C. et Duvauferrier, R. (2012). Vers la définition automatique des éléments de données des fiches RCP en cancérologie à partir d'une ontologie. *Systèmes d'information pour l'amélioration de la qualité en santé: Comptes rendus des quatorzièmes Journées francophones*, 121.
- Brin, S. et Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Computer networks and ISDN systems*, 30(1-7), 107-117.
- Brownson, R. C., Kreuter, M. W., Arrington, B. A. et True, W. R. (2006). From the schools of public health. *Public health reports*, 121(1), 97-103.



- Butler, C. E., Noel, S., Hibbs, S. P., Miles, D., Staves, J., Mohagheh, P., . . .  
Murphy, M. F. (2015). Implementation of a clinical decision support system  
improves compliance with restrictive transfusion policies in hematology  
patients. *Transfusion*, 55(8), 1964-1971.
- Butt, A. S., Haller, A. et Xie, L. (2016). DWRank: Learning concept ranking for  
ontology search. *Semantic Web*, 7(4), 447-461.
- Carpineto, C. et Romano, G. (2012). A survey of automatic query expansion in  
information retrieval. *Acm Computing Surveys (CSUR)*, 44(1), 1-50.
- Cartright, M.-A., Huston, S. et Feild, H. (2012). Galago: A Modular Distributed  
Processing and Retrieval System
- Chowdhury, G. G. (2010). Introduction to modern information retrieval Facet  
publishing.
- Cieslewicz, A., Dutkiewicz, J. et Jedrzejek, C. (2018). Baseline and extensions  
approach to information retrieval of complex medical data: Poznan's approach  
to the bioCADDIE 2016. *Database*, 2018.
- Codling, E. A., Plank, M. J. et Benhamou, S. (2008). Random walk models in  
biology. *Journal of the Royal society interface*, 5(25), 813-834.
- Collins, A. M. et Loftus, E. F. (1975). A spreading-activation theory of semantic  
processing. *Psychological review*, 82(6), 407.

- Crimp, R. et Trotman, A. (2018). Refining Query Expansion Terms using Query Context
- Dallmann, A., Niebler, T., Lemmerich, F. et Hotho, A. (2016). Extracting semantics from random walks on wikipedia: Comparing learning and counting methods
- Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological review*, 93(3), 283.
- Demaret, P., Tucci, M., Ducruet, T., Trottier, H. et Lacroix, J. (2014). Red blood cell transfusion in critically ill children (CME). *Transfusion*, 54(2), 365-375.
- Demner-Fushman, D. et Lin, J. (2007). Answering clinical questions with knowledge-based and statistical techniques. *Computational Linguistics*, 33(1), 63-103.
- Demner-Fushman, D., Rogers, W. J. et Aronson, A. R. (2017). MetaMap Lite: an evaluation of a new Java implementation of MetaMap. *Journal of the American Medical Informatics Association*, 24(4), 841-844.
- Després, S. et Szulman, S. (2008). Réseau terminologique versus Ontologie
- Deveaud, R., Mothe, J., Ullah, M. Z. et Nie, J.-Y. (2018). Learning to adaptively rank document retrieval system configurations. *ACM Transactions on Information Systems (TOIS)*, 37(1), 1-41.
- Dillon, M. (1983). Introduction to modern information retrieval: G. Salton and M. McGill. McGraw-Hill, New York (1983). xv+ 448 pp., \$32.95 ISBN 0-07-054484-0 : Pergamon.

Djulgovic B et Guyatt GH. (2017). Progress in evidence-based medicine: a quarter century on. *The Lancet*.

Dong, L., Srimani, P. K. et Wang, J. Z. (2011). Ontology graph based query expansion for biomedical information retrieval *IEEE*.

Doran, P., Tamma, V. et Iannone, L. (2007). Ontology module extraction for ontology reuse: an ontology engineering perspective

Dramé, K., Diallo, G., Delva, F., Dartigues, J. F., Mouillet, E., Salamon, R. et Mougin, F. (2014, 2014/04/01/). Reuse of termino-ontological resources and text corpora for building a multilingual domain ontology: An application to Alzheimer's disease. *Journal of Biomedical Informatics*, 48, 171-182. doi: <https://doi.org/10.1016/j.jbi.2013.12.013>

Dridi, O. (2008). Ontology-based information retrieval: Overview and new proposition *IEEE*.

Du Pont-Thibodeau, G., Tucci, M. et Ducruet, T. (2014). Survey on stated transfusion practices in PICUs. *Pediatr Crit Care Med* 15:409–16.

Dupret, G. (2011). *Discounted cumulative gain and user decision models* Springer.

Ebell, M. (1999). Information at the point of care: answering clinical questions. *The Journal of the American Board of Family Practice*, 12(3), 225-235.

- Egozi, O., Markovitch, S. et Gabrilovich, E. (2011). Concept-based information retrieval using explicit semantic analysis. *ACM Transactions on Information Systems (TOIS)*, 29(2), 1-34.
- Fava, G. A., Guidi, J., Rafanelli, C. et Sonino, N. (2015). The clinical inadequacy of evidence-based medicine and the need for a conceptual framework based on clinical judgment. *Psychotherapy and psychosomatics*, 84(1), 1.
- Fichhoff, B., Slovic, P. et Lichtenstein, S. (1977). Knowing with certainty: the appropriateness of extreme confidence. *Journal of Experimental Psychology: Human Perception and Performance* 3, 552 – 564.
- Fineout-Overholt, E., Melnyk, B. M. et Schultz, A. (2005). Transforming health care from the inside out: advancing evidence-based practice in the 21st century. *Journal of Professional Nursing*, 21(6), 335-344.
- Fontelo P, Fang Liu et Raymonde C. (2015). How does Evidence Affect Clinical Decision-making? *Evid Based Med*, 20(25): 156–161.
- Frey, L. J., Bernstam, E. V. et Denny, J. C. (2016, Jul). Precision medicine informatics. *J Am Med Inform Assoc*, 23(4), 668-670. doi: 10.1093/jamia/ocw053
- Gąsieniec, L. et Radzik, T. (2008). *Memory efficient anonymous graph exploration* Springer.
- Georgiou, A. (2001). Health informatics and evidence-based medicine-more than a marriage of convenience? *Health Informatics Journal*, 7(3-4), 127-130.

- Ghafourian, S., Rezaeian, A. et Naghibzadeh, M. (2013). Modularization of graph-structured ontology with semantic similarity
- Godfrey, M. D., Granger, C. W. et Morgenstern, O. (1964). THE RANDOM - WALK HYPOTHESIS OF STOCK MARKET BEHAVIOR a. *Kyklos*, 17(1), 1-30.
- Goel, R., Chappidi, M., Patel, E., Ness, P., Cushing, M., Frank, S. et Tobian, A. (2018 Feb ). Trends in red blood cell, plasma, and platelet transfusions in the United States 1993-2014  
*JAMA* 319(8):825-7.
- Gómez-Pérez, A., Martínez-Romero, M., Rodríguez-González, A., Vázquez, G. et Vázquez-Naya, J. M. (2013). Ontologies in medicinal chemistry: Current status and future challenges. *Current topics in medicinal chemistry*, 13(5), 576-590.
- Goodwin, T. R., Skinner, M. A. et Harabagiu, S. M. (2017). UTD HLTRI at TREC 2017: Precision Medicine Track
- Graves, A., Adali, S. et Hendler, J. (2008). A method to rank nodes in an RDF graph  
CEUR-WS. org.
- Greenhalgh, T., Howick, J. et Maskrey, N. (2014). Evidence based medicine: a movement in crisis? *Bmj*, 348, g3725.
- Griffa, A. et Van den Heuvel, M. P. (2018). Rich-club neurocircuitry: function, evolution, and vulnerability. *Dialogues in clinical neuroscience*, 20(2), 121.

Gruber, T. R. (1995). Towards principles for the design of ontologies used for knowledge sharing,. in *International Journal of Human-Computer Studies*, 43, pp. 907- 928.

Guarino, N., Oberle, D. et Staab, S. (2009). What Is an Ontology? Handbook on Ontologies, 10.1007/1978-1003-1540-92673-92673\_92670.

Hacene, M. R., Napoli, A., Valtchev, P., Toussaint, Y. et Bendaoud, R. (2008). Ontology learning from text using relational concept analysis IEEE.

Harpring, P. (2010). What Are Controlled Vocabularies. Introduction to Controlled Vocabularies: Terminology for Art, Architecture, and Other Cultural Works. Los Angeles, CA: Getty Publications.

Haynes, B. et Haines, A. (1998). Barriers and bridges to evidence based clinical practice. *BMJ (Clinical research ed.)*, 317(7153), 273-276. doi: 10.1136/bmj.317.7153.273

Haynes, R. B., Devereaux, P. J. et Guyatt, G. H. (2002). Clinical expertise in the era of evidence-based medicine and patient choice. *BMJ Evidence-Based Medicine*, 7(2), 36-38.

Hedden, H. (2010). Taxonomies and controlled vocabularies best practices for metadata. *Journal of Digital Asset Management*, 6(5), 279-284.

Henry, C., Barry, M. M., Mark, H., Ebell, M. M., Allen F. Shaughnessy, P. et David C. Slawson, M., MS, and Fern Neitzke, BS. ( 1998). Family Physicians' Use

of Medical Abstracts To Guide Decision Making: Style or Substance? Forth  
American Primary Research

Group Annual Scientific Meeting in Montreal,.

Hiemstra, D. (2009a). Information Retrieval Models.

Hiemstra, D. (2009b). Information retrieval models. *Information Retrieval: searching in the 21st Century*, 1-17.

Huang, K.-C., Chiang, I.-J., Xiao, F., Liao, C.-C., Liu, C. C.-H. et Wong, J.-M. (2013). PICO element detection in medical text without metadata: Are first sentences enough? *Journal of biomedical informatics*, 46(5), 940-946.

Huang, X., Lin, J. et Demner-Fushman, D. (2006). PICO as a Knowledge Representation for Clinical Questions

Jaspers, M. W., Smeulers, M., Vermeulen, H. et Peute, L. W. (2011). Effects of clinical decision-support systems on practitioner performance and patient outcomes: a synthesis of high-quality systematic review findings. *Journal of the American Medical Informatics Association*, 18(3), 327-334.

Jervis, R. (1976). *Perception and Misperception in International Politics*. Princeton Univ. Press, Princeton.

Jim Q. Chen, S. M. L. (2003). An exploratory cognitive DSS for strategic decision making,. *Decision Support Systems*, Volume 36(Issue 2,), Pages 147-160,. doi: [https://doi.org/10.1016/S0167-9236\(02\)00139-2](https://doi.org/10.1016/S0167-9236(02)00139-2).

- Jin, D. et Szolovits, P. (2018). Pico element detection in medical text via long short-term memory neural networks
- Jin, D. et Szolovits, P. (2020). Advancing PICO Element Detection in Biomedical Text via Deep Neural Networks. *Bioinformatics*.
- Kahneman, D., Slovic, S. P., Slovic, P. et Tversky, A. (1982). Judgment under uncertainty: Heuristics and biases Cambridge university press.
- Kahneman, D. et Tversky, A. (1973). On the psychology of prediction. *Psychological review*, 80(4), 237.
- Kashyap, V. et Borgida, A. (2003). Representing the UMLS® semantic network using OWL Springer.
- Kharbat, F. et El-Ghalayini, H. (2008). Building ontology from knowledge base systems. Dans *Data Mining in Medical and Biological Research*. IntechOpen.
- Kim, H., Bian, J., Mostafa, J., Jonnalagadda, S. et Del Fiol, G. (2016). Feasibility of extracting key elements from ClinicalTrials. gov to support clinicians' patient care decisions, vol. 2016. American Medical Informatics Association.
- Kim, Y., Zerfos, P., Sheinin, V. et Greco, N. (2017). Ranking the importance of ontology concepts using document summarization techniques *IEEE*.
- Koehrsen, W. (2018). Beyond accuracy: Precision and recall. *Towards Data Science*.



- Komurov, K., White, M. A. et Ram, P. T. (2010). Use of data-biased random walks on graphs for the retrieval of context-specific networks from genomic data. *PLoS Comput Biol*, 6(8), e1000889.
- Konev, B., Lutz, C., Walther, D. et Wolter, F. (2009). Formal properties of modularisation. Dans *Modular Ontologies* (p. 25-66). Springer.
- Kosowski, A. (2013). Graph Exploration: How to do better than the random walk? . Récupéré de <https://www.irif.fr/~displexity/Larochelle/Displexity-Adrian.pdf>
- L'homme, M.-C. (2004). *La terminologie: principes et techniques* : Pum.
- L'homme, M.-C. (2005). Sur la notion de «terme». *Meta: journal des traducteurs/Meta: Translators' Journal*, 50(4), 1112-1132.
- Lacroix , J., Hébert, P., Hutchison, Hume, H., Tucci, M., Ducruet, T., . . . Network, P. A. L. I. a. S. I. (2007). Transfusion strategies for patients in pediatric intensive care units. *N Engl J Med*, 19;356(16):1609-19.
- Lambin, P., Roelofs, E., Reymen, B., Velazquez, E. R., Buijsen, J., Zegers, C. M., . . . Oberije, C. (2013). Rapid Learning health care in oncology'—an approach towards decision support systems enabling customised radiotherapy. *Radiotherapy and Oncology*, 109(1), 159-164.
- Laverdière, C. M. F., France Gauvin, M., FRCPC FAAP, Paul C. Hébert, M., FRCPC, MHSc, Claire Infante-Rivard, M., FRCPC, PhD, Heather Hume, M., FRCPC, Baruch J. Toledano, M., FRCPC, FAAP, . . . Group, f. t. C. C. C. T. (2002). Survey of transfusion practices in pediatric intensive care units. *Pediatr Crit Care Med*, 3:335-40.

Li, J., Xia, F., Wang, W., Chen, Z., Asabere, N. Y. et Jiang, H. (2014). Acrec: a co-authorship based random walk model for academic collaboration recommendation ACM.

Lin, J., Crane, M., Trotman, A., Callan, J., Chattopadhyaya, I., Foley, J., . . . Vigna, S. (2016). Toward reproducible baselines: The open-source IR reproducibility challenge Springer.

Lin, J. et Trotman, A. (2015). Anytime ranking for impact-ordered indexes

Lisa, W. (2007). Role of information technology in evidence based medicine: Advantage and limitations. *International Journal of Healthcare Administration*, 4(2), 5.

Liu, T.-Y. (2011). Learning to rank for information retrieval.

Liu, Y., McInnes, B. T., Pedersen, T., Melton-Meaux, G. et Pakhomov, S. (2012). Semantic relatedness study using second order co-occurrence vectors computed from biomedical corpora, UMLS and WordNet

López-García, P., Oleynik, M., Kasác, Z. et Schulz, S. (2017). TREC 2017 Precision Medicine-Medical University of Graz

Lovász, L. (1993). Random walks on graphs: A survey. *Combinatorics, Paul erdos is eighty*, 2(1), 1-46.

- Luhn, H. P. (1957). A statistical approach to mechanized encoding and searching of literary information. *IBM Journal of research and development*, 1(4), 309-317.
- Macdonald, C., He, B., Plachouras, V. et Ounis, I. (2005). University of Glasgow at TREC 2005: Experiments in Terabyte and Enterprise Tracks with Terrier
- Maedche, A. et Staab, S. (2001). Ontology learning for the semantic web. *IEEE Intelligent systems*, 16(2), 72-79.
- Maedche, A. et Staab, S. (2004). Ontology learning. Dans *Handbook on ontologies* (p. 173-190). Springer.
- Mahmood, A. A., Li, G., Rao, S., McGarvey, P. B., Wu, C. H., Madhavan, S. et Vijay-Shanker, K. (2017). UD\_GU\_BioTM at TREC 2017: Precision Medicine Track
- McCray, A. T. (1989). The UMLS Semantic Network
- McCray, A. T., Burgun, A. et Bodenreider, O. (2001). Aggregating UMLS semantic types for reducing conceptual complexity. *Studies in health technology and informatics*, 84(0 1), 216.
- McInnes, B., Pedersen, T., Pakhomov, S., Liu, Y. et Melton-Meaux, G. (2013). Umls:: similarity: Measuring the relatedness and similarity of biomedical concepts
- Medaglia, J. D., Lynall, M.-E. et Bassett, D. S. (2015). Cognitive network neuroscience. *Journal of cognitive neuroscience*, 27(8), 1471-1491.

- Merabti, T., Letord, C., Abdoune, H., Lecroq, T., Joubert, M. et Darmoni, S. J. (2009). Projection and inheritance of SNOMED CT relations between MeSH terms
- Metzler, D. et Croft, W. B. (2004). Combining the language model and inference network approaches to retrieval. *Information processing & management*, 40(5), 735-750.
- Mikolov, T., Chen, K., Corrado, G. et Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.
- Mitra, M., Singhal, A. et Buckley, C. (1998). Improving automatic query expansion
- Mitra, P., Wiederhold, G. et Kersten, M. L. (1999). A Graph-Oriented Model for Articulation of Ontology.
- Mogotsi, I. (2010). Christopher d. manning, prabhakar raghavan, and hinrich schütze: Introduction to information retrieval : Springer.
- Mougin, F. (2019). Enjeux liés aux ressources termino-ontologiques en santé.
- Naik, V. A., Kotur, P. et Goudar, S. S. (2002). Evidence based medicine and its impact on medical education. *Indian Journal of Anaesthesia*, 46(2), 96-103.
- Neoh, H. L. et Na, J. (2005). Effectiveness of UMLS semantic network as a seed ontology for medical domain ontology building. Nanyang Technological University, School of Communication and Information.

Newell, B. R. et Shanks, D. R. (2014). Unconscious influences on decision making: A critical review. *Behavioral and brain sciences*, 37(1), 1-19.

Niebler, T., Schlör, D., Becker, M. et Hotho, A. (2016). Extracting semantics from unconstrained navigation on wikipedia. *KI-Künstliche Intelligenz*, 30(2), 163-168.

NIH. (2019). knowledge source Récupéré de [https://www.nlm.nih.gov/research/umls/knowledge\\_sources/metathesaurus/source\\_faq.html](https://www.nlm.nih.gov/research/umls/knowledge_sources/metathesaurus/source_faq.html)

Noh, J. et Kavuluru, R. (2017). Team UKNLP at TREC 2017 Precision Medicine Track: A Knowledge-Based IR System with Tuned Query-Time Boosting

Ogilvie, P., Voorhees, E. et Callan, J. (2009). On the number of terms used in automatic query expansion. *Information Retrieval*, 12(6), 666.

Ounis, I., Amati, G., Plachouras, V., He, B., Macdonald, C. et Lioma, C. (2006). Terrier: A high performance and scalable information retrieval platform

Ounis, I., Lioma, C., Macdonald, C. et Plachouras, V. (2007). Research directions in terrier: a search engine for advanced retrieval on the web. *CEPIS Upgrade Journal*, 8(1).

Pal, D., Mitra, M. et Datta, K. (2013). Query expansion using term distribution and term association. *arXiv preprint arXiv:1303.0667*.

- Park, H.-J. et Friston, K. (2013). Structural and functional brain networks: from connections to cognition. *Science*, 342(6158).
- Park, T. K. (1993). The nature of relevance in information retrieval: An empirical study. *The library quarterly*, 63(3), 318-351.
- Pathak, J., Johnson, T. M. et Chute, C. G. (2009). Survey of modular ontology techniques and their applications in the biomedical domain. *Integrated computer-aided engineering*, 16(3), 225-242.
- Paz-Trillo, C., Wassermann, R. et Braga, P. P. (2005). An information retrieval application using ontologies. *Journal of the Brazilian Computer Society*, 11(2), 17-31.
- Pedersen, T., Pakhomov, S. V., Patwardhan, S. et Chute, C. G. (2007). Measures of semantic similarity and relatedness in the biomedical domain. *Journal of biomedical informatics*, 40(3), 288-299.
- Pinel-Sauvagnat, K. et Mothe, J. (2013). Mesures de la qualité des systèmes de recherche d'information. *Ingénierie des systèmes d'information*, 18(3), 11-38.
- Plachouras, V., He, B. et Ounis, I. (2004). University of Glasgow at TREC 2004: Experiments in Web, Robust, and Terabyte Tracks with Terrier
- Pouriyeh, S., Allahyari, M., Kochut, K. et Arabnia, H. R. (2018). A comprehensive survey of ontology summarization: measures and methods. *arXiv preprint arXiv:1801.01937*.

- Rajbhandary, S., Whitaker, B. et Perez, G. (2014-15 ). Blood collection, utilization, and patient blood management survey report. AABB, 91p.
- Reymonet, A., Thomas, J. et Aussenac-Gilles, N. (2007). Modélisation de ressources termino-ontologiques en OWL
- Richardson, W. S., Wilson, M. C., Nishikawa, J. et Hayward, R. S. (1995). The well-built clinical question: a key to evidence-based decisions. *Acp j club*, 123(3), A12-13.
- Roberts, K., Demner-Fushman, D., Voorhees, E. M., Hersh, W. R., Bedrick, S., Lazar, A. J. et Pant, S. (2017). Overview of the TREC 2017 Precision Medicine Track
- Rocchio, J. (1971, 1971). Relevance feedback in information retrieval. *The Smart Retrieval System-Experiments in Automatic Document Processing*, 313-323.
- Roche, C. (2005a). Terminologie et ontologie. *Langages*, 157(1), 48-62. doi: 10.3917/lang.157.0048
- Roche, C. (2005b). Terminologie et ontologie. *Langages*, (1), 48-62.
- Roche, C. (2006). L'ontologie comme principe terminologique. *Terminologie et ontologie: description du réel*, 21-36.
- Rodrigues, R. J. (2000). Information systems: the key to evidence-based health practice. *Bulletin of the world health organization*, 78, 1344-1351.

- Roussey, C., Pinet, F., Kang, M. A. et Corcho, O. (2011). An introduction to ontologies and ontology engineering. Dans *Ontologies in Urban development projects* (p. 9-38). Springer.
- Ruotsalo, T. et Mäkelä, E. (2009). A comparison of corpus-based and structural methods on approximation of semantic relatedness in ontologies. *International Journal on Semantic Web and Information Systems (IJSWIS)*, 5(4), 39-56.
- Sackett, D. L., Rosenberg, W. M., Gray, J. M., Haynes, R. B. et Richardson, W. S. (1996). *Evidence based medicine: what it is and what it isn't* : British Medical Journal Publishing Group.
- Sackett, D. L., Straus, S. E. et *Medicine, f. F. A. o. t. N. D. o.* (1998). Finding and Applying Evidence During Clinical RoundsThe "Evidence Cart". *JAMA*, 280(15), 1336-1338. doi: 10.1001/jama.280.15.1336
- Saeed, G. (2005). *Fundamentals of Probability with Stochastic Processes* : Upper Saddle River, NJ: Pearson Education.
- Safaeifard, F., Shariatpanahi, S. P. et Goliaei, B. (2018). A survey on random walk-based stochastic modeling in eukaryotic cell migration with emphasis on its application in cancer. *Multidisciplinary Cancer Investigation*, 2(1), 1-12.
- Samimi, P. et Ravana, S. D. (2014). Creation of reliable relevance judgments in information retrieval systems evaluation experimentation through crowdsourcing: a review. *The Scientific World Journal*, 2014.
- Saracevic, T. (1995). *Evaluation of evaluation in information retrieval* Citeseer.



- Schardt, C., Adams, M. B., Owens, T., Keitz, S. et Fontelo, P. (2007). Utilization of the PICO framework to improve searching PubMed for clinical questions. *BMC medical informatics and decision making*, 7(1), 16.
- Schütze, H., Manning, C. D. et Raghavan, P. (2008). *Introduction to information retrieval*, vol. 4.
- Siew, C. S., Wulff, D. U., Beckage, N. M. et Kenett, Y. N. (2019). Cognitive Network Science: A review of research on cognition through the lens of network representations, processes, and dynamics. *Complexity*, 2019.
- Sim, I., Gorman, P., Greenes, R. A., Haynes, R. B., Kaplan, B., Lehmann, H. et Tang, P. C. (2001). Clinical decision support systems for the practice of evidence-based medicine. *Journal of the American Medical Informatics Association*, 8(6), 527-534.
- Simon, H. A. (1957). *Models of man: social and rational; mathematical essays on rational human behavior in society setting* New York: Wiley.
- Singer, P., Niebler, T., Strohmaier, M. et Hotho, A. (2013). Computing semantic relatedness from human navigational paths: A case study on wikipedia. *International Journal on Semantic Web and Information Systems (IJSWIS)*, 9(4), 41-70.
- Sony, P. et Sureshkumar, N. (2019). Concept-Based Electronic Health Record Retrieval System in Healthcare IOT. Dans *Cognitive Informatics and Soft Computing* (p. 175-188). Springer.
- Sparck Jones, K. et Van Rijsbergen, C. (1975). Report on the Need for and Provision of an "Ideal. *Information Retrieval Test Collection*.

- Straus, S. E. et McAlister, F. A. (2000). Evidence-based medicine: a commentary on common criticisms. *Canadian Medical Association Journal*, 163(7), 837-841.
- Strohman, T., Metzler, D., Turtle, H. et Croft, W. B. (2005). *Indri: A language model-based search engine for complex queries*, vol. 2. Citeseer.
- Stuckenschmidt, H. et Klein, M. (2004). *Structure-based partitioning of large concept hierarchies* Springer.
- Stuckenschmidt, H., Parent, C. et Spaccapietra, S. (2009). *Modular ontologies: concepts, theories and techniques for knowledge modularization* (vol. 5445) Springer.
- Susan M, G., Trudi, G., Irwin, G. et Aryeh, S. (2019). Society for the advancement of blood management administrative and clinical standards for patient blood management programs. (pediatric version). *Pediatric Anesthesia* 29(23):231-236.
- Tchechmedjiev, A. et Jonquet, C. (2017). *Enrichment of French Biomedical Ontologies with UMLS Concepts and Semantic Types for Biomedical Named Entity Recognition Through Ontological Semantic Annotation*
- Thomas, F. (2013). *Ontologie, Thesaurus et Taxonomie sur le web de données*. Récupéré de <http://blog.sparna.fr/2013/12/07/ontologie-thesaurus-taxonomie-web-de-donnees/>
- Tonelli, M. R. (2011). Integrating clinical research into clinical decision making. *Annali dell'Istituto superiore di sanità*, 47(1), 26-30.

- Tonelli, S., Rospocher, M., Pianta, E. et Serafini, L. (2011). Boosting collaborative ontology building with key-concept extraction IEEE.
- Tran, L.-T. T., Divita, G., Carter, M. E., Judd, J., Samore, M. H. et Gundlapalli, A. V. (2015). Exploiting the UMLS Metathesaurus for extracting and categorizing concepts representing signs and symptoms to anatomically related organ systems. *Journal of biomedical informatics*, 58, 19-27.
- Trotman, A., Jia, X. et Crane, M. (2012). Towards an Efficient and Effective Search Engine
- Turpin, A. et Scholer, F. (2006). User performance versus precision measures for simple search tasks
- Tversky, A. et Kahneman, D. Judgment under Uncertainty: Heuristics and Biases. *science*. 1974; 185 (4157): 1124–31.
- Tversky, A. et Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological review*, 90(4), 293.
- Valentine, S., Bateman, S., Bembea, M., Muszynski, J., Spinella, P., Steiner, M., . . . Lacroix, J. (2018). on behalf of the Pediatric Critical Care Transfusion and Anemia Expertise Initiative, the Pediatric Critical Care Blood Research Network (BloodNet), and the Pediatric Acute lung Injury and Sepsis Investigators (PALISI) Network. Consensus recommendations for red blood cell transfusion practice in critically ill children from the Pediatric Critical Care Transfusion and Anemia Expertise Initiative (TAXI). *Pediatr Crit Care Med* 19:884-98.

- Voorhees, E. et Harman, D. (2000). Overview of the ninth text retrieval conference (TREC-9), National Institute of Standards and Technology. report, Gaithersburg, available at: [http://trec.nist.gov/pubs/trec9/papers/overview\\_9.pdf](http://trec.nist.gov/pubs/trec9/papers/overview_9.pdf) (accessed February 2014).
- Voorhees, E. M. (1994). Query expansion using lexical-semantic relations Springer.
- Voorhees, E. M. (2000). Variations in relevance judgments and the measurement of retrieval effectiveness. *Information processing & management*, 36(5), 697-716.
- Voorhees, E. M. et Harman, D. K. (2000). The eighth text retrieval conference (TREC-8).
- Wallace, B. C., Kuiper, J., Sharma, A., Zhu, M. et Marshall, I. J. (2016). Extracting PICO sentences from clinical trial reports using supervised distant supervision. *The Journal of Machine Learning Research*, 17(1), 4572-4596.
- Weigert, S. (2014). From Random Walks to Personalized PageRank. Récupéré de <https://www.r-bloggers.com/from-random-walks-to-personalized-pagerank/>
- Wenger, D. R. (2012). Limitations of evidence-based medicine: the role of experience and expert opinion. *Journal of Pediatric Orthopaedics*, 32, S187-S192.
- Wilkie, C. et Azzopardi, L. (2014). Best and fairest: An empirical analysis of retrieval system bias Springer.

- Willis, C. et Losee, R. M. (2013). A random walk on an ontology: Using thesaurus structure for automatic subject indexing. *Journal of the American Society for Information Science and Technology*, 64(7), 1330-1344.
- Wright, T. B., Ball, D. et Hersh, W. (2017). Query expansion using MeSH terms for dataset retrieval: OHSU at the bioCADDIE 2016 dataset retrieval challenge. *Database*, 2017.
- Wu, B. et Chellapilla, K. (2007). Extracting link spam using biased random walks from spam seed sets *ACM*.
- Wu, G., Li, J., Feng, L. et Wang, K. (2008). Identifying potentially important concepts and relations in an ontology *Springer*.
- Wu, W., Liu, T., Hu, H. et Du, X. (2012). Extracting domain-relevant term using Wikipedia based on random walk model *IEEE*.
- Wulff, D. U., De Deyne, S., Jones, M. N., Mata, R. et Consortium, A. L. (2019). New perspectives on the aging lexicon. *Trends in cognitive sciences*.
- Xu, Y. et Chen, Z. (2006). Relevance judgment: What do information users consider beyond topicality? *Journal of the American Society for Information Science and Technology*, 57(7), 961-973.
- Yeh, E., Ramage, D., Manning, C. D., Agirre, E. et Soroa, A. (2009). WikiWalk: random walks on Wikipedia for semantic relatedness *Association for Computational Linguistics*.

- Yusuf, S., Wittes, J., Probstfield, J. et Tyroler, H. A. (1991). Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. *Jama*, 266(1), 93-98.
- Zacklad, M. (2007). Classification, thésaurus, ontologies, folksonomies: comparaisons du point de vue de la recherche ouverte d'information (ROI)
- Zhang, Z., Zeng, D., Abbasi, A. et Peng, J. (2013). A random walk model for item recommendation in folksonomies. arXiv preprint arXiv:1310.7957.
- Zhou, B. et Yao, Y. (2010). Evaluating information retrieval system performance based on user preference. *Journal of Intelligent Information Systems*, 34(3), 227-248.
- Zhu, W., Xu, X., Hu, X., Song, I.-Y. et Allen, R. B. (2006). Using UMLS-based re-weighting terms as a query expansion strategy IEEE.
- Zobel, J. (1998). How reliable are the results of large-scale information retrieval experiments? Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval, 307-314.
- Zweigenbaum, P. (2004). L'UMLS entre langue et ontologie: une approche pragmatique dans le domaine médical. *Revue d'Intelligence Artificielle*, 18(1), 111-137.