HAL
open science

Contribution à l'analyse et l'évaluation des requêtes expertes : cas du domaine médical

Eya Znaidi

To cite this version:

Eya Znaidi. Contribution à l'analyse et l'évaluation des requêtes expertes : cas du domaine médical. Recherche d'information [cs.IR]. Université Paul Sabatier - Toulouse III, 2016. Français. NNT : 2016TOU30054 . tel-01515377

HAL Id: tel-01515377

https://theses.hal.science/tel-01515377

Submitted on 27 Apr 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par : *l'Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)*
Cotutelle internationale *Université de Tunis El Manar*

Présentée et soutenue le 30/06/2016 **par :**

Eya ZNAIDI

Contribution à l'analyse et l'évaluation des requêtes expertes : cas du domaine médical

---

JURY

Chantal Soulé-Dupuy	Professeur, Université Toulouse 1, Capitole	Présidente
Catherine Berrut	Professeur, Université Joseph Fourier	Rapporteur
Ikram Amous	MCF/HDR, Université de Sfax	Rapporteur
Lina Soualmia	MCF/HDR, Université de Rouen	Examinatrice
Rim Faiz	Professeur, Université de Carthage	Examinatrice
Lynda Tamine	Professeur, Université de Toulouse 3	Directrice
Chiraz Latiri	MCF/HDR, Université de la Manouba	Co-directrice
Cécile Chouquet	MCF, Université de Toulouse 3	Invitée

École doctorale et spécialité :
MITT : Image, Information, Hypermedia
Unité de Recherche :
Institut de Recherche en Informatique de Toulouse (UMR 5505)
Directeur(s) de Thèse :
Lynda Tamine et *Chiraz Latiri*
Rapporteurs :
Ikram Amous et *Catherine Berrut*

Contribution à l'analyse et l'évaluation des requêtes expertes : cas du domaine médical

Eya ZNAIDI

8 juillet 2016

Dédicace

À cœur vaillant rien d'impossible
À conscience tranquille tout est accessible.
Quand il y a la soif d'apprendre
Tout vient à point à qui sait attendre.
Quand il y a le souci de réaliser un dessein
Tout devient facile pour arriver à nos fins.
Malgré les obstacles qui s'opposent
En dépit des difficultés qui s'interposent.
Les études sont avant tout
Notre unique et seul atout.
Ils représentent la lumière de notre existence
L'étoile brillante de notre réjouissance.
Comme un vol de gerfauts hors du charnier natal
Nous partons ivres d'un rêve héroïque et brutal.
Espérant des lendemains épiques...

Je dédie cette thèse
À mes parents Mounir et Dalila

Remerciements

Je souhaite avant tout exprimer mes plus vifs remerciements à Lynda Tamine-Lechani ma directrice de thèse , pour m'avoir fait confiance pour m'avoir guidé, encouragé, aidé, pour ses multiples conseils et pour toutes les heures qu'elle a consacré à diriger cette recherche. J'aimerais également lui dire à quel point j'ai apprécié sa grande disponibilité, sa patience et son soutien indéfectible. J'ai été extrêmement sensible à ses qualités humaines d'écoute et de compréhension tout au long de ce travail. Mes chaleureux remerciements pour son attention de tous les instants sur mes travaux, pour ses conseils avisés et son écoute qui ont été prépondérants pour la bonne réussite de cette thèse. Son énergie et sa confiance ont été des éléments moteurs pour moi.

Je souhaite exprimer ma sincère reconnaissance et mes remerciements à ma co-directrice de thèse Chiraz Latiri, pour la gentillesse et la patience qu'elle a manifesté à mon égard durant cette thèse, pour tous les conseils, son respect sans faille des délais serrés de relecture des documents que je lui ai adressés. Je la remercie également pour sa bienveillance et ses remarques précieuses. Je la remercierai pas assez pour la confiance, le soutien et la sympathie qu'elle m'a témoigné au cours de cette thèse.

Je remercie également Mme Cécile Chouquet pour son aide, sa collaboration et participation pour ce travail.

Je voudrais remercier les rapporteurs de cette thèse Mme. Catherine Berrut, Professeur des Universités de l'Université Joseph Fourier, et Mme Ikram Amous, MCF/HDR de l'Université de Sfax, our avoir accordé du temps à une lecture attentive et détaillée de mon manuscrit ainsi que pour leurs remarques encourageantes et constructives.

J'associe à ces remerciements Mme Lina Soualmia MCF/HDR de l'Université de Rouen, et Mme Rim Faiz, Professeur de l'Université de Carthage, pour avoir accepté d'examiner mon travail.

Je souhaite exprimer ma gratitude à M. Claude Chrisment, puis Mme Josiane Mothe, pour m'avoir accueillie au sein de l'équipe SIG. Je remercie l'ensemble des membres de l'équipe SIG pour leur gentillesse et les différents échanges que j'ai pu avoir avec chacun d'entre eux. Des remerciements tout particuliers aux membres de l'équipe IRIS, merci à Gilles de m'avoir accordé sa confiance en me laissant participer à ses enseignements.

Merci également à tous les doctorants amis et collègues : Ghada, Laure, Bilel, Amjed, Firas, Rafik, Manel, Chiraz, Faten, Mariem, Diep, Ismail, Ophélie, Imen, Thomas, Inès, Lynda Said L'hadj,.. pour leur sympathie, leur amitié. J'ai eu beaucoup de plaisir à travailler avec eux merci pour les bons moments partagés. Je tiens à remercier Thibaut pour ses nombreuses relectures, pour ses conseils, pour son aide écoute et support sans failles. Merci Ghada, j'ai trouvé en toi plus qu'une amie.

Ma reconnaissance va à ceux qui ont plus particulièrement assuré le soutien affectif pendant ma thèse : ma famille, merci pour votre irremplaçable et inconditionnel soutien. Je tiens à remercier tout particulièrement papa et maman pour leur soutien quotidien indéfectible et leur enthousiasme contagieux à l'égard de mes travaux comme de la vie en général. Ils sont mon moteur, ma source de motivation, ils n'ont jamais cessé de me soutenir, me supporter, m'encourager. Je n'aurais pas assez de mots pour vous remercier et vous dire ce que vous représentez pour moi. Vous m'avez toujours encouragé à aller de l'avant malgré la difficulté d'être loin de ses proches. Merci d'avoir été là pour écarter les doutes, soigner les blessures et partager les joies. Cette thèse est aussi la vôtre.
Enfin, je remercie mon futur époux Fares pour sa patience.

Encore un grand merci à tous pour m'avoir conduit à ce jour mémorable.

Résumé

La recherche d'information nécessite la mise en place de stratégies qui consistent à (1) cerner le besoin d'information ; (2) formuler le besoin d'information ; (3) repérer les sources pertinentes ; (4) identifier les outils à exploiter en fonction de ces sources ; (5) interroger les outils ; et (6) évaluer la qualité des résultats. Ce domaine n'a cessé d'évoluer pour présenter des techniques et des approches permettant de sélectionner à partir d'un corpus de documents l'information pertinente capable de satisfaire le besoin exprimé par l'utilisateur. De plus, dans le contexte applicatif du domaine de la RI biomédicale, les sources d'information hétérogènes sont en constante évolution, aussi bien du point de vue de la structure que du contenu. De même, les besoins en information peuvent être exprimés par des utilisateurs qui se caractérisent par différents profils, à savoir : les experts médicaux comme les praticiens, les cliniciens et les professionnels de santé, les utilisateurs néophytes (sans aucune expertise ou connaissance du domaine) comme les patients et leurs familles, etc.

Plusieurs défis sont liés à la tâche de la RI biomédicale, à savoir : (1) la variation et la diversité du besoin en information, (2) différents types de connaissances médicales, (3) différences de compétences linguistiques entre experts et néophytes, (4) la quantité importante de la littérature médicale ; et (5) la nature de la tâche de RI médicale. Cela implique une difficulté d'accéder à l'information pertinente spécifique au contexte de la recherche, spécialement pour les experts du domaine qui les aideraient dans leur prise de décision médicale.

Nos travaux de thèse s'inscrivent dans le domaine de la RI biomédicale et traitent les défis de la formulation du besoin en information experte et l'identification des sources pertinentes pour mieux répondre aux besoins cliniques.

Concernant le volet de la formulation et l'analyse de requêtes expertes, nous proposons des analyses exploratoires sur des attributs de requêtes, que nous avons définis, formalisés et calculés, à savoir : (1) deux attributs de longueur en nombre de termes et en nombre de concepts, (2) deux facettes de spécificité terme-document et hiérarchique, (3) clarté de la requête basée sur la pertinence et celle basée sur le sujet de la requête. Nous avons proposé des études et analyses statistiques sur des collections issues de différentes campagnes d'évaluation médicales CLEF et TREC, afin de prendre en compte les différentes tâches de RI. Après les analyses descriptives, nous avons étudié d'une part, les corrélations par paires d'attributs de requêtes et les analyses de corrélation multidimensionnelle. Nous avons étudié l'impact de ces corrélations sur les performances de recherche d'autre part. Nous avons pu ainsi comparer et caractériser les différentes requêtes selon la tâche médicale d'une manière

plus généralisable.

Concernant le volet lié à l'accès à l'information, nous proposons des techniques d'appariement et d'expansion sémantiques de requêtes dans le cadre de la RI basée sur les preuves cliniques. En effet, les médecins ont appris à formuler leur requêtes selon les facettes PICO (Population/Problem (P), Intervention (I), Comparaison (C) et Outcome (O)). Ces facettes sont rarement annotées ou identifiées dans le texte des documents et des requêtes.

Dans le but de mieux répondre aux questions cliniques PICO, nous avons proposé différentes approches pour évaluer les questions cliniques expertes : (i) une représentation sémantique sous forme de graphes conceptuels des questions PICO, (ii) une approche d'expansion de requêtes basée sur un algorithme de propagation de scores pour sélectionner les meilleurs concepts ; et (iii) un modèle d'ordonnancement requête-document, basé sur une méthode d'agrégation prioritaire de scores de pertinence. Ainsi, nous avons contextualisé les scores d'importance des facettes aux documents et aux requêtes dans l'objectif d'améliorer la performance de recherche et de retourner des réponses pertinentes. Nous avons mené des évaluation expérimentales pour évaluer nos contributions dans la cadre de la recherche PICO. Nous avons utilisé pour cela la collection de données CLIREC, construite dans le but d'évaluer la RI clinique (Boudin *et al.*, 2010c). Les résultats ont montré la performance des différentes approches proposées.

Abstract

The research topic of this document deals with a particular setting of medical information retrieval (IR), referred to as expert based information retrieval. We were interested in information needs expressed by medical domain experts like praticians, physicians, etc. It is well known in information retrieval (IR) area that expressing queries that accurately reflect the information needs is a difficult task either in general domains or specialized ones and even for expert users. Thus, the identification of the users' intention hidden behind queries that they submit to a search engine is a challenging issue. Moreover, the increasing amount of health information available from various sources such as government agencies, non-profit and for-profit organizations, internet portals etc. presents opportunities and issues to improve health care information delivery for medical professionals, patients and general public. One critical issue is the understanding of users search strategies and tactics for bridging the gap between their intention and the delivered information.
In this thesis, we focus, more particularly, on two main aspects of medical information needs dealing with the expertise which consist of two parts, namely :

— Understanding the users intents behind the queries is critically important to gain a better insight of how to select relevant results. While many studies investigated how users in general carry out exploratory health searches in digital environments, a few focused on how are the queries formulated, specifically by domain expert users. We address more specifically domain expert health search through the analysis of query attributes namely length, specificity and clarity using appropriate proposed measures built according to different sources of evidence. In this respect, we undertake an in-depth statistical analysis of queries issued from IR evaluation compaigns namely Text REtrieval Conference (TREC) and Conference and Labs of the Evaluation Forum (CLEF) devoted for different medical tasks within controlled evaluation settings.

— We address the issue of answering PICO (Population, Intervention, Comparison and Outcome) clinical queries formulated within the Evidence Based Medicine framework. The contributions of this part include (1) a new algorithm for query elicitation based on the semantic mapping of each facet of the query to a reference terminology, and (2) a new document ranking model based on a prioritized aggregation operator. we tackle the issue related to the retrieval of the best evidence that fits with a PICO question, which is an underexplored research area. We propose a new document ranking algorithm that relies on semantic based query expansion leveraged by each question facet. The expansion is moreover bounded by the local search context to better discard irrelevant documents. The experimental evaluation carried out on the CLIREC dataset shows the benefit of our approaches.

Table des matières

Dédicace **iii**

Remerciements **v**

Résumé **vii**

Abstract **ix**

Contexte de recherche et contributions de la thèse **9**
1 Contexte et problématique 9
2 Contributions 11
3 Organisation de la thèse 12
4 Publications dans le cadre de la thèse 15

I Recherche d'information et domaine médical : Synthèse des travaux de l'état de l'art **17**

1 Recherche d'information : Concepts et modèles **19**
1 Introduction 20
2 Concepts de base de la RI 20
2.1 Fondements 20
2.2 Mise en œuvre d'un SRI 22
2.2.1 Indexation 22
2.2.2 Interrogation 23
3 Taxonomie des modèles de RI 24
3.1 Modèle booléen 25
3.2 Modèle vectoriel 26
3.3 Modèles probabilistes 27
4 Reformulation de la requête 29
4.1 Reformulation par réinjection de la pertinence 31
4.2 Reformulation par pseudo-réinjection de la pertinence 32
5 Évaluation des performances en RI 33
5.1 Collections de test 33
5.2 Campagnes d'évaluation 34
5.2.1 Protocole d'évaluation TREC 34

5.2.2 Protocole d'évaluation CLEF . . . 35
5.3 Mesures d'évaluation en RI . . . 37
6 Conclusion . . . 39

2 Recherche d'information médicale : principes de base 41
1 Introduction . . . 43
2 Aperçu du domaine de la RI médicale . . . 44
2.1 Enjeux et défis du domaine . . . 44
2.2 Typologie de l'information médicale . . . 45
2.2.1 La littérature biomédicale . . . 46
2.2.2 Les dossiers médicaux des patients . . . 48
2.3 Quelques ressources termino-ontologiques du domaine biomédical . . . 48
2.3.1 Thésaurus MeSH . . . 49
2.3.2 Méta-thésaurus UMLS . . . 50
2.3.3 Nomenclature SNOMED . . . 51
2.3.4 Gene Ontology GO . . . 51
3 Étude des besoins en information médicale . . . 52
3.1 Catégories de requêtes médicales . . . 53
3.2 Caractéristiques des requêtes médicales . . . 56
3.2.1 Caractéristiques selon les utilisateurs . . . 57
3.2.2 Caractéristiques linguistiques . . . 59
3.2.3 Sujet de la requête . . . 61
3.3 Stratégies et comportements de recherche . . . 62
4 Techniques et modèles de RI médicale . . . 65
4.1 Expansion de requêtes . . . 65
4.2 Expansion de documents . . . 67
5 Modèles de RI basées sur le paradigme PICO . . . 68
5.1 EBM et paradigme PICO . . . 68
5.2 Identification des éléments PICO . . . 72
5.3 Modèles de RI basés sur les facettes PICO . . . 75
6 Campagnes d'évaluation pour la RI biomédicale . . . 78
6.1 Campagne d'évaluation TREC . . . 78
6.2 Campagne d'évaluation ImageCLEF . . . 81
7 Aperçu de systèmes de RI médicaux . . . 83
7.1 PubMed . . . 83
7.2 EMERSE . . . 84
7.3 CisMef . . . 85
7.4 Doctissimo . . . 86
7.5 iMed . . . 86
7.6 MedSearch . . . 86
8 Conclusion . . . 87

II Analyse et évaluation de requêtes médicales 89

3 Analyses statistiques exploratoires des facteurs caractéristiques des requêtes expertes 95
1 Introduction . . . 96

2 Motivations et questions de recherche . . . 96
2.1 Motivations . . . 96
2.2 Objectifs et questions de recherche . . . 97
3 Définition et formalisation des attributs de requêtes . . . 99
3.1 Longueur de la requête . . . 99
3.2 Spécificité de la requête . . . 100
3.3 Clarté de la requête . . . 101
4 Caractérisation du besoin en information des experts du domaine médical . . . 102
4.1 Description des données d'analyses . . . 103
4.1.1 Collections de la campagne d'évaluation TREC . . . 103
4.1.2 Collections de la campagne d'évaluation CLEF . . . 104
4.2 Analyses et résultats . . . 106
4.2.1 Identification des Caractéristiques de requêtes . . . 106
4.2.2 Analyses de corrélation des attributs de requêtes . . . 108
4.3 Synthèse et discussion . . . 109
4.3.1 Analyses multidimensionnelles des corrélations entre les attributs . . . 110
4.3.2 Impact des attributs des requêtes sur les performances de recherche 112
5 Caractérisation des requêtes PICO . . . 114
5.1 Analyses descriptives . . . 114
5.2 Analyses de corrélations entre attributs de requêtes PICO . . . 115
6 Bilan et conclusion . . . 118

4 Évaluation des requêtes cliniques : Modèles sémantiques pour mieux répondre aux questions PICO 121
1 Introduction . . . 123
2 Motivations et hypothèses . . . 125
3 Aperçu général de nos contributions . . . 127
4 Processus d'élicitation sémantiques des requêtes PICO . . . 128
4.1 Méthode de génération de graphes . . . 128
4.2 Identification de concepts les plus pertinents . . . 130
5 Évaluation des requêtes cliniques . . . 133
5.1 Expansion sémantique des requêtes cliniques . . . 133
5.2 Modèle d'ordonnancement basé sur un opérateur d'agrégation prioritaire des scores . . . 135
5.2.1 Cadre général . . . 135
5.2.2 Calcul de scores de pertinence des documents . . . 136
6 Évaluation expérimentale . . . 137
6.1 Données expérimentales . . . 137
6.2 Évaluation de l'approche d'expansion de requêtes . . . 137
6.2.1 Métriques d'évaluation et modèles de référence . . . 137
6.2.2 Ajustement des paramètres . . . 138
6.2.3 Analyse de l'efficacité de la recherche : évaluation comparative . . . 140
6.2.4 Discussion . . . 140
6.3 Évaluation du modèle d'ordonnancement de documents . . . 141
6.3.1 Objectifs . . . 141
6.3.2 Métriques et protocole d'évaluation . . . 141
6.3.3 Efficacité du modèle de pertinence basé sur l'agrégation des scores . 142
6.3.4 Analyse de la robustesse du modèle d'agrégation de pertinence . . . 143

6.3.4.1 Analyse globale 144
6.3.4.2 Analyse au niveau requête 146
7 Bilan et conclusion 149

Conclusion générale **151**
Synthèse des contributions 151
Perspectives 152

Bibliographie **172**

Table des figures

1.1 Processus en U de la RI. 24
1.2 Taxonomie des modèles en RI (Baeza-Yates et Ribeiro-Neto, 2011). 25
1.3 Représentation algébrique des documents et des requêtes dans l'espace des termes à deux dimensions. 27
1.4 Aperçu du processus de la reformulation de requête. 30
1.5 Description graphique des éléments nécessaires pour calculer le rappel et la précision. 38

2.1 Exemple de résultats utilisant GO. 52
2.2 Distribution de requêtes médicales selon le type (White et Horvitz, 2012). 55
2.3 Stratégie de recherche globale de la MFF. 69
2.4 Approche de l'évidence basée sur les faits. 69
2.5 Aperçu de domaines du système PubMed. 84
2.6 Capture d'écran de EMERSE montrant les termes d'expansion du mot "cigarette". . 85
2.7 Aperçu du système CisMef 85
2.8 Aperçu du site web Doctissimo. 86

3.1 Distribution des facettes d'attributs de requêtes par collection 107
3.2 Nuage de points résultats de l'ACP, 1 désigne la collection *TRECMedical* ; 2 désigne la collection ImageCLEF ; 3 désigne la collection *TRECGenomics03* ; 4 désigne la collection *TRECGenom04* et 5 désigne la collection *TRECGenom06*. 111
3.3 Scores moyens des performances de recherche par tâche médicale. 113
3.4 Distribution des attributs de requêtes par collection 116
3.5 Analyses de corrélations des attributs de requêtes PICO 117
3.6 Distribution des deux facettes de la longueur pour les sous-graphes PICO 118

4.1 Aperçu du processus d'élicitation sémantique de requêtes PICO. 128
4.2 Exemple de graphe sémantique de requête PICO 129
4.3 Exemple d'un appariement sémantique requête-document 130
4.4 Exemple de graphes sémantiques conceptuels 133
4.5 Aperçu de notre modèle de traitement des questions PICO. 134
4.6 Architecture de notre modèle de traitement des questions PICO. 135
4.7 Variation du paramètre pour le modèle PRF 139
4.8 Variation du paramètre pour le modèle BQE 139
4.9 Variation du paramètre pour l'expansion de requête 139
4.10 Paramétrage du modèle GQE 140
4.11 Variation du paramètre α. 142

4.12 Statistiques sur l'amélioration/dégradation en terme de MAP comparée au modèle *PLM* 144
4.13 Statistiques sur l'amélioration/dégradation en terme de MAP comparée au modèle *AGM* 144

Liste des tableaux

1.1 Fonctions principales de classement de termes basés sur l'analyse des distributions des termes dans les documents de pseudo-réinjection de pertinence (Carpineto et Romano, 2012). . . . 33
1.2 Notations des mesures d'évaluation en RI. . . . 37

2.1 Classification de la littérature médicale . . . 47
2.2 Classification de l'information textuelle médicale . . . 47
2.3 Le processus de réalisation d'un examen systématique et du méta-analyse . . . 48
2.4 Les différentes catégories ou domaines du MeSH . . . 50
2.5 Les onze axes de la nomenclature SNOMED . . . 51
2.6 Catégories de requêtes médicales selon le modèle de (Pratt et Wasserman, 2000) . . . 54
2.7 Questions les plus posées, poursuivies et satisfaites selon (Ely *et al.*, 2002). . . . 55
2.8 Types d'information médicale selon l'utilisateur . . . 56
2.9 Tableau récapitulatif des études et observations sur les caractéristiques des requêtes médicales (Tamine *et al.*, 2015) . . . 61
2.10 Tableau récapitulatif des études et observations sur le comportement de recherche dans le domaine médical (Partie 1) (Tamine *et al.*, 2015) . . . 63
2.11 Tableau récapitulatif des études et observations sur le comportement de recherche dans le domaine médical (Partie 2) (Tamine *et al.*, 2015) . . . 64
2.12 Exemple de requête clinique PICOTT . . . 70
2.13 Exemple de requête PICO . . . 71
2.14 Classification de travaux selon l'utilisation des éléments cliniques . . . 78
2.15 Statistiques des collections TREC Medical . . . 79
2.16 Statistiques des collections TREC Genomics . . . 79
2.17 Tâches de la campagne TREC Genomics (Hersh, 2009) . . . 80
2.18 Statistiques sur la piste TREC de prise de décision médicale . . . 80
2.19 Questions cliniques utilisées dans les campagnes 2014 et 2015 . . . 81
2.20 Statistiques des collections *ImageCLEF* (*Case-based IR*) . . . 82
2.21 Statistiques des collections CLEF eHealth. . . . 83

3.1 Formalisation des facettes des attributs de requêtes . . . 99
3.2 Tableau des notations . . . 100
3.3 Exemples de requêtes avec les facettes de la longueur . . . 101
3.4 Exemples de requêtes avec les facettes de la spécificité . . . 102
3.5 Exemples de requêtes avec leurs différents scores de clarté . . . 103

3.6 Descriptions des collections utilisées pour l'analyse 103
3.7 Exemples de requêtes de la collection de test de la tâche TRECMedical 2011. 104
3.8 Exemples de requêtes de différentes collections de test de la tâche TRECGenomics. . 104
3.9 Récapitulatif des critères selon la tâche médicale. 105
3.10 Résultats des corrélations entre les facettes d'attributs de requêtes des différentes collections 109
3.11 Scores moyens de performance par collection (et écart-type) avec la *p-value* du test de comparaison des scores entre collections († † † : *p-value* $< 0,001$)) 112
3.12 Résultats de la modélisation du score de performance $P@100$ en fonction des attributs des requêtes (par une MANCOVA) : estimation des paramètres associés aux attributs significatifs (et erreur standard, *s.e.*) et *p-value* (*ns :* †† *:* $0,001 <$ *p-value* $< 0,01$ *;*) 113
3.13 Résultats de corrélations deux à deux entre les attributs de requêtes relatives à la collection CLIREC ($N = 423$) 117

4.1 Notations utilisées pour les différents algorithmes. 126
4.2 Exemple de requête textuelle avec les annotations PICO correspondantes. 132
4.3 Top 5 documents retournés pour la requête A3.1 par le modèle BM25. 132
4.4 Exemple de sous requêtes PICO de la requête A3.1. 132
4.5 Exemple de concepts médicaux actifs 132
4.6 Exemple de concepts de documents avec les scores associés. 133
4.7 Statistiques de la collection de test CLIREC. 137
4.8 Comparaison de l'impact sur la performance de recherche de l'expansion de requêtes basée sur le graphe sémantique. $\%Acc$: Test de significativité de Student sur la métrique MAP † : $0,01 < t \leq 0,05$; †† : $0,001 < t \leq 0,01$; † † † : $t \leq 0,001$. 141
4.9 Comparaison de l'impact de notre modèle sur l'efficacité de recherche. $\%Acc$ indique les taux d'amélioration du modèle PSM en terme de MAP. Les Symboles †, †† et † † † indiquent le test de significativité de Student : † : $0,01 < t \leq 0,05$; †† : $0,001 < t \leq 0,01$; † † † : $t \leq 0,001$. 143
4.10 Statistiques pour comparer la performance entre le modèle PSM et le modèle PLM. 145
4.11 Statistiques pour comparer la performance entre le modèle PSM et le modèle AGM. 145
4.12 Analyse comparative de cas de requêtes types PSM vs. PLM 147
4.13 Analyses de l'échec et de l'amélioration de la performance des requêtes avec le modèle PLM. 148
4.14 Description des requêtes de l'analyse (*Cf.* Tableau 4.13). 148
4.15 Analyse de l'échec et de l'amélioration de la performance des requêtes avec le modèle AGM. 148
4.16 Description des requêtes de l'analyse (*Cf.* Tableau 4.15). 149

Contexte de recherche et introduction générale

1 Contexte et problématique

La recherche d'information (RI) est l'ensemble des techniques permettant de sélectionner à partir d'une collection de documents ceux qui sont susceptibles de répondre au besoin de l'utilisateur exprimé via une requête (Salton, 1970). À travers cette définition, nous retenons trois concepts clés : le document, le besoin et la pertinence. La recherche d'information nécessite ainsi la mise en place d'une stratégie qui consiste à (1) cerner le besoin en information ; (2) formuler le besoin en information ; (3) repérer les sources pertinentes ; (4) identifier les outils à exploiter en fonction de ces sources ; (5) interroger ces outils ; et (6) évaluer la qualité des résultats. Durant la dernière décennie, le domaine de la RI n'a cessé d'évoluer du point de vue de la définition de modèles d'indexation, de modèles d'appariement requête-document ainsi que de la spécification de cadres d'évaluation. Il a été décliné en plusieurs domaines d'application tels que la RI multilingue, la RI contextuelle ou encore la RI médicale.

Nos travaux s'inscrivent spécifiquement dans le cadre de la RI biomédicale où des utilisateurs experts ou néophytes utilisent divers moteurs de recherche pour effectuer leurs tâches de recherche d'information. Ils peuvent s'orienter vers des sites spécialisés dans le domaine de la santé, vers des moteurs de recherche généraux ou encore vers des systèmes de RI généraux ou médicaux. En effet, les SRI médicaux ont connu une grande évolution depuis ces deux dernières décennies tant du point de vue de leur architecture que de la qualité et de la diversité des services autour du stockage de l'information, de l'accès à l'information pertinente pour une médecine basée sur des niveaux de preuve ainsi que de l'aide à la décision pour l'amélioration de la qualité des soins (Hersh, 2006).

Dans ce cadre général, l'information biomédicale utilisée comme support pour les tâches de recherche, d'extraction d'information et de connaissances concerne principalement la littérature médicale et les dossiers médicaux des patients. Grâce aux immenses progrès réalisés dans le domaine biomédical, la littérature scientifique biomédicale publiée augmente de manière exponentielle, nous citons à titre d'exemple la ressource MEDLINE.

Toutefois, la diversification au niveau du volume de données, de l'hétérogénéité de la structure et du contexte ainsi que l'exigence de qualité et de la sécurité des informations sont à l'origine de défis dans le domaine de RI biomédicale. Ces défis se traduisent par les difficultés rencontrées par

des utilisateurs qui cherchent l'information médicale pertinente en vue de prendre une décision. Ces utilisateurs se caractérisent souvent par différents niveaux d'expertise, des connaissances hétérogènes et des comportements de recherche multiples ainsi qu'une manière spécifique à chacun selon son contexte d'exploiter les ressources à disposition pour une prise de décision pertinente.

Autour de ce sujet, la revue de la littérature liée à la RI biomédicale révèle plusieurs limites, à savoir : la difficulté de formulation des requêtes, la difficulté de l'interprétation des besoins en information induits par les SRI et plus spécifiquement, la non-exploitation des éléments PICO (Population/problème, Intervention, Comparaison, Output) de la médecine basée sur les niveaux de preuves, qui représentent une source d'évidence pour améliorer la réponse aux questions cliniques, etc.
Plusieurs verrous restent donc à lever, notamment l'étude d'une manière précise et approfondie du besoin en information exprimé par les experts du domaine médical en vue d'améliorer la prise de décision ainsi que l'optimisation des modèles d'appariement requête-document en RI.

Dans ce cadre général, nous nous positionnons dans le cadre spécifique de l'analyse et de l'évaluation des requêtes d'experts du domaine médical en vue d'améliorer les systèmes d'aide à la décision en médecine, permettant de mieux répondre à des besoins en information spécifiques. Ce sujet aborde, plus spécifiquement, le problème de la formulation des requêtes expertes ainsi que la pertinence des résultats retournés pour mieux répondre aux questions des professionnels de santé étant donné que les requêtes médicales sont particulièrement complexes.

La problématique de recherche a été abordée dans un double objectif, à savoir :

1. Mener des analyses statistiques qui portent sur les spécificités des requêtes expertes afin d'étudier leurs caractéristiques et analyser les corrélations existantes. Le but est de révéler les facteurs caractéristiques des requêtes en analysant leurs points communs et différences ainsi que leur impact sur la performance de recherche en considérant différentes tâches du domaine médical.
2. Se focaliser sur les requêtes cliniques PICO [1] dans le cadre de la médecine basée sur les faits, en s'intéressant à leur identification et leur exploitation dans le cadre de la RI médicale. Nous proposons une approche sémantique d'expansion de requêtes ainsi qu'un algorithme de propagation de scores basé sur les relations entre les concepts médicaux dans les requêtes et les documents. Un modèle d'ordonnancement basé sur un opérateur prioritaire d'agrégation de scores exploitant ces facettes PICO est également proposé, en vue de mieux répondre aux questions cliniques.

1. Patient/Problem, Intervention, Comparaison, Output

2 Contributions

Les travaux présentés dans ce mémoire se situent dans le contexte précis de l'analyse et de l'extraction de connaissances ainsi que l'accès à l'information médicale. Plus précisément, nous nous sommes intéressés à un volet du domaine médical qui traite le besoin en information des experts.

Nos contributions sont scindées en deux parties, à savoir :

1. **Caractérisation et analyses des besoins en information des experts du domaine médical**. Dans ce contexte, nous nous sommes intéressés à : (a) Étudier les requêtes expertes des cliniciens. Nous avons mené pour cela des analyses statistiques sur 173 requêtes issues de 3 tâches de recherche d'information médicales issues de deux campagnes d'évaluation différentes TREC [2] et CLEF [3]. Nous avons défini différents attributs de requêtes comme éléments et critères d'analyse, à savoir : *la longueur des requêtes en nombre de termes et en nombre de concepts*, *la clarté de la requête avec la clarté basée sur le sujet de la requête et la clarté basée sur la pertinence*, *la spécificité de la requête avec la spécificité hiérarchique et la spécificité terme-document.*

   (b) Analyser les besoins en information exprimés par les experts médicaux dans l'objectif de les caractériser et mesurer l'impact de leur structure sur les résultats de recherche. Nous avons mené une étude exploratoire basée sur des analyses statistiques multidimensionnelles issues de campagnes d'évaluation internationales en l'occurrence TREC et CLEF. Nous nous sommes focalisés sur deux aspects ; un premier aspect consiste à identifier et analyser les corrélations entre attributs de requêtes (la longueur, la spécificité et la clarté), en utilisant des mesures appropriées construites selon différentes sources d'évidence. Un deuxième aspect porte sur l'étude de l'impact de ces attributs sur les performances de recherche des SRI liées à ces requêtes.

   Ces travaux d'analyse statistique exploratoire ont montré le besoin de contextualiser les modèles de RI médicale à la tâche. De plus, les résultats obtenus suggèrent la prise en compte de plusieurs caractéristiques et spécificités à inclure dans les nouveaux SRI médicaux comme l'utilisation de niveaux de concepts terminologiques appropriés afin d'améliorer la clarté de la requête, la personnalisation des résultats de recherche selon le niveau d'expertise, la prise en compte de la catégorie de la requête et de la nature de la tâche de RI.

2. **Évaluation des requêtes cliniques PICO en proposant une représentation sémantique des requêtes sous forme de graphes ainsi que des approches de calcul de scores de pertinence pour mieux répondre aux questions PICO**. Notre contribution est triple, à savoir :

   (a) Proposition d'une approche de génération de graphes sémantiques, permettant de mieux représenter chacune des facettes PICO de la requête. La particularité de notre approche réside dans la représentation des requêtes en se basant sur les concepts extraits à partir de MeSH, permettant de représenter le contexte spécifique de la requête, où chacune des facettes PICO est représentée par un sous-graphe conceptuel.

2. Text Retrieval Conference
3. Cross-language Forum

(b) Proposition d'une approche d'expansion de requêtes pour mieux répondre aux questions cliniques PICO. Nous avons adressé le problème lié à la recherche des preuves cliniques pertinentes, adaptées aux questions PICO. Nous proposons un nouvel algorithme d'ordonnancement de documents basé sur une approche d'expansion de requêtes délimitée par le contexte local de la recherche qui permet d'éliminer les documents non pertinents. Ensuite, nous avons proposé un algorithme de classement de documents qui permet de lier chacun des éléments de la requête à un document qui lui est pertinent afin de retourner des réponses qui correspondent au mieux au besoin exprimé par la requête ;

(c) Pour mieux répondre aux requêtes cliniques, nous avons proposé l'application d'une approche d'agrégation prioritaire de l'information sémantique identifiée au sein du besoin en information exprimé via les requêtes PICO. Nous avons proposé un nouveau modèle d'ordonnancement basé sur une approche sémantique de requête qui prend en compte le contexte de chacune des facettes PICO. La spécificité de cette méthode réside dans le nouvel algorithme de propagation de scores qui permet de sélectionner les meilleurs concepts qui correspondent à la fois aux requêtes et aux documents pertinents pour classer et identifier les meilleures preuves qui traduisent les documents pertinents.

Afin de prendre en compte l'importance de chacune des facettes P, I, C et O dans le calcul de score global de pertinence des documents, nous avons proposé : (*i*) une méthode de génération de graphes PICO, qui englobe une technique de calcul de score conceptuel correspondant à chaque facette PICO, (*ii*) une approche d'appariement sémantique requête-document permettant de sélectionner les meilleurs concepts de chaque graphe et (*iii*) une approche de calcul de scores de pertinence basée sur un opérateur prioritaire d'agrégation de scores.

3 Organisation de la thèse

Cette thèse est constituée d'un chapitre introductif ainsi que de deux parties principales. La première partie présente une synthèse des travaux de l'état de l'art et donne un aperçu sur le contexte de recherche et la problématique dans le cadre de la RI biomédicale, présentée en deux chapitres : *Recherche d'information : concepts et modèles* (Chapitre 1), *Recherche d'information médicale* (Chapitre 2). La deuxième partie présente nos contributions et est subdivisée en deux chapitres : *Analyses exploratoires des facteurs caractéristiques des requêtes médicales expertes* (Chapitre 3), *Évaluation des requêtes cliniques : Modèles sémantiques pour mieux répondre aux questions PICO.* (Chapitre 4). Enfin, nous concluons le manuscrit et discutons des perspectives de recherche. Nous présentons le contenu ci-après.

— Le Chapitre 1, **Recherche d'Information : concepts et modèles**, présente les principes et concepts de base de la RI classique. Nous présentons en particulier dans la section 2 les notions et concepts de base en RI de manière générale. Ensuite, nous décrivons le processus général de la RI dans la section 2.2. Nous passons par la suite en revue les modèles de RI les plus représentatifs de l'état de l'art dans la section 3. Les techniques de reformulation de requêtes sont présentées dans la section 4. Puis, nous abordons le protocole d'évaluation ainsi que les mesures d'évaluation des performances de la RI dans la section 5.

— Le Chapitre 2, **Recherche d'information médicale : principes de base**, présente d'une manière spécifique les fondements et les concepts de la RI médicale. Nous présentons en particulier dans la section 2 les principes et notions de base de la RI médicale en exposant

les enjeux du domaine ainsi qu'une description de la typologie de l'information médicale et de quelques ressources termino-ontologiques du domaine. Nous donnons dans la section 3 une classification du besoin du domaine médical, où nous détaillons la typologie des requêtes médicales et les caractéristiques de la recherche experte. Les différents techniques et modèles de la RI médicale sont présentés dans la section 4. Nous présentons les modèles de RI basés sur le paradigme PICO dans la section 5. Nous détaillons ainsi les deux volets de travaux concernant l'identification des éléments PICO et de leur utilisation en RI. Enfin, la section 6 présente les principales campagnes d'évaluation pour la RI médicale et la section 7 donne un aperçu sur les systèmes de RI médicaux.

— Le Chapitre 3, **Analyses statistiques exploratoires des requêtes d'experts médicaux**, présente nos contributions sur l'analyse des caractéristiques des besoins en information des experts du domaine médical. Le chapitre commence par les motivations et les questions de recherche posées dans le cadre de cette contribution. La section 3 détaille la formalisation des attributs de requêtes, utilisés pour nos analyses exploratoires. Nous décrivons dans la section 4 les données d'analyse. Les résultats des deux principales contributions sont détaillés dans les sections 5 et 6. La section 7 présente une analyse statistique complémentaire conduite sur les requêtes PICO.

— Le chapitre 4, **Évaluation des requêtes cliniques : Modèles sémantiques pour mieux répondre aux questions PICO**, nous commençons par présenter les motivations concernant la RI médicale basée sur le paradigme PICO ainsi que les hypothèses de recherche dans la section 2. La section 3 détaille notre approche de génération de graphes sémantiques. Nous présentons dans la section 4 nos contributions dans le cadre de l'évaluation des requêtes cliniques PICO, structurés en deux sections : la section 4.1 présente l'approche d'expansion de requêtes, la section 4.2 détaille le modèle d'ordonnancement basé sur une agrégation prioritaire des scores de pertinence que nous détaillons dans ce qui suit :

  — Nous présentons une approche sémantique d'expansion de requêtes médicales PICO. Nous donnons un aperçu général de notre approche sémantique basée sur un modèle de calcul de pertinence combiné, qui exploite les propriétés sémantiques de la requête et les documents pertinents. Nous présentons par la suite les expérimentations, avec une description du cadre expérimental. Les résultats expérimentaux obtenus sont présentés et discutés en utilisant la collection biomédicale dédiée à la médecine basée sur les faits, et les requêtes PICO, appelée CLIREC. Afin de montrer l'efficacité de notre approche de RI, nous comparons nos résultats expérimentaux aux meilleurs résultats obtenus avec un modèle testé sur cette même collection.

  — Notre modèle d'appariement sémantique dans le cadre de la RI basée sur le paradigme PICO, qui a pour but d'améliorer la représentation des requêtes PICO et de répondre au mieux à ces questions. Cela comprend un aperçu général de l'architecture du modèle proposé, suivi des différentes approches de calcul de scores proposées : d'une part, le calcul de scores de pertinence des documents utilisant une fonction d'agrégation de scores avec l'ordre de priorité des éléments PICO ; d'autre part, le calcul de scores de pertinence pour ordonner les documents basés sur l'approche combinée qui se base sur les termes et les concepts. Les résultats expérimentaux obtenus sont présentés et discutés. Nous présentons tout d'abord le cadre d'évaluation, puis les résultats des deux contributions majeures : l'évaluation de l'impact de la représentation sémantique des requêtes sur la RI et l'évaluation des performances du modèle proposé. Nous terminons par une évaluation comparative des différentes approches avec les modèles de l'état de l'art.

L'évaluation expérimentale conduite ainsi que les résultats des performances de recherche sont

présentés dans la section 5.

Ce manuscrit se termine par une conclusion générale dans la laquelle nous présentons la synthèse de nos contributions et nous énonçons quelques perspectives de recherche.

4 Publications dans le cadre de la thèse

Articles publiés dans des conférences internationales avec comités de lecture

1. Eya Znaidi, Lynda Tamine, Chiraz Latiri (2015). Aggregating Semantic Information Nuggets for Answering Clinical Queries. Dans : 31st ACM Symposium on Applied Computing **(SAC 2016)**, Pisa, Italy, April 4-8, 2016.

2. Eya Znaidi, Lynda Tamine, Chiraz Latiri (2015). Answering PICO Clinical Questions : A Semantic Graph-Based Approach. Dans : The 15th Conference on Artificial Intelligence in Medicine **(AIME 2015)**, Pavia, Italy, June 17-20, 2015 (Short paper).

3. Eya Znaidi, Lynda Tamine, Cécile Chouquet, Chiraz Latiri (2013). Characterizing health-related information needs of domain experts. Dans : The 14th Conference on Artificial Intelligence in Medicine **(AIME 2013)** May 29 - June 1, 2013, Murcia, Spain. Artificial Intelligence in Medicine, 48-57.

Articles publiés dans des conférences et ateliers nationaux avec comités de lecture

1. Eya Znaidi, Lynda Tamine, Chiraz Latiri (2016). Répondre à des requêtes cliniques PICO. COnférence en Recherche d'Information et Applications **(CORIA 2016)**, Toulouse 08/03/2016-11/03/2016.
2. Eya Znaidi, Lynda Tamine, Cécile Chouquet, Chiraz Latiri (2013). Analyse exploratoire des requêtes d'experts médicaux : cas des campagnes d'évaluation TREC et CLEF. Dans : 2e édition du Symposium sur l'Ingénierie de l'Information Médicale **(SIIM 2013)**, Lille, 1 Juillet 2013.

Première partie

Recherche d'information et domaine médical : Synthèse des travaux de l'état de l'art

Chapitre 1

Recherche d'information : Concepts et modèles

Sommaire

1 Introduction . . . 20
2 Concepts de base de la RI . . . 20
2.1 Fondements . . . 20
2.2 Mise en œuvre d'un SRI . . . 22
2.2.1 Indexation . . . 22
2.2.2 Interrogation . . . 23
3 Taxonomie des modèles de RI . . . 24
3.1 Modèle booléen . . . 25
3.2 Modèle vectoriel . . . 26
3.3 Modèles probabilistes . . . 27
4 Reformulation de la requête . . . 29
4.1 Reformulation par réinjection de la pertinence . . . 31
4.2 Reformulation par pseudo-réinjection de la pertinence . . . 32
5 Évaluation des performances en RI . . . 33
5.1 Collections de test . . . 33
5.2 Campagnes d'évaluation . . . 34
5.2.1 Protocole d'évaluation TREC . . . 34
5.2.2 Protocole d'évaluation CLEF . . . 35
5.3 Mesures d'évaluation en RI . . . 37
6 Conclusion . . . 39

1 Introduction

La recherche d'information (RI) traite de la représentation, du stockage, de l'organisation et de l'accès à l'information (Manning *et al.*, 2008). Le but principal de la RI est de permettre à l'utilisateur d'exprimer son besoin en information et de chercher l'information pertinente qui répond à son besoin spécifique. Ce besoin en information est souvent formulé en langage naturel par une requête décrite par un ensemble de mots clés. Satisfaire un besoin en information se traduit concrètement par la mise en correspondance d'une requête exprimée souvent sous forme de mots-clés, d'une part, et des informations disponibles dans les documents textuels d'une collection. Ce processus se déroule au sein d'un *système de recherche d'informations* (SRI).

L'essor du web avec la diversité des tâches et des utilisateurs, ainsi que l'évolution exponentielle des données, représente plusieurs défis pour l'accès à l'information, à savoir : retrouver l'information pertinente, qui répond au besoin spécifique de l'utilisateur, en tenant compte du cadre de recherche et du contexte de la tâche de RI. En effet, la difficulté majeure à laquelle sont soumis la plupart des moteurs de recherche provient de deux éléments : l'expression de requêtes non spécifiques qui ne couvrent pas la thématique de recherche et donc ne retournent pas de documents pertinents, et l'inadéquation de l'appariement requête-document, provenant essentiellement des modèles sur lesquels se basent les moteurs de RI classiques. Ces approches généralistes considèrent le besoin comme une requête représentée principalement par son sujet. Pour pallier ces lacunes, des réflexions ont été menées dans le but de mieux cerner la notion de pertinence du point de vue de l'utilisateur et d'identifier les différents facteurs ayant un impact sur cette notion (Borlund, 2003). Les études menées dans ce cadre ont montré que la pertinence n'est pas une relation isolée entre un document et une requête ; elle intègre différents facteurs liés à l'utilisateur et son environnement dans le cadre d'une tâche de recherche d'information, à savoir : la diversité, l'accessibilité et la fraîcheur des résultats de recherche, la crédibilité des auteurs, etc.

Ce premier chapitre est dédié à la représentation des concepts et fondements du domaine de la recherche d'information. En effet, il traite des concepts de base de la RI classique et il est organisé comme suit : la section 2 présente les notions et les modèles de base de la RI classique, ainsi que la démarche d'évaluation des systèmes de RI. Dans la section 3, nous présentons une taxonomie des différents modèles de RI. La section 4 détaille les approches de reformulation de requêtes. Nous détaillons dans la section 5 les mesures d'évaluation utilisées dans la RI suivis d'un aperçu des principales campagnes d'évaluation. La section 6 conclut le chapitre.

2 Concepts de base de la RI

2.1 Fondements

La recherche d'information (RI) est un domaine de recherche qui intègre des modèles et des techniques dont le but est de faciliter l'accès à l'information pertinente pour un utilisateur ayant un besoin en information. L'objectif principal de la RI est de trouver l'information pertinente qui répond et satisfait le besoin en information/requête via un SRI (Système de Recherche d'Informaion). Ce dernier permet de retrouver, à partir d'une collection de documents, les documents susceptibles d'être pertinents au besoin en information d'un utilisateur.

Plusieurs concepts clés s'articulent autour de la définition d'un système de RI, à savoir :

— **Besoin en information** : cette notion est souvent assimilée au besoin exprimé par l'utilisateur (Bigot, 2013). Afin de déterminer différentes classes de besoins en information selon un critère de difficulté, l'auteur a identifié des groupes de besoins qualifiés de faciles, moyens, difficiles ou encore très difficiles en se basant sur l'hypothèse que les systèmes de RI sont plus ou moins performants selon la difficulté des besoins traités. Par ailleurs, (Cutrell et Guan, 2007) ont classé les besoins utilisateurs selon la recherche navigationnelle ou exploratoire.

— **Une requête** : il s'agit de l'expression d'un besoin en information par l'utilisateur. C'est un ensemble de mots clés, termes ou expression, exprimés en langage naturel, que l'utilisateur soumet aux moteurs de recherche afin de récupérer les documents ciblés. Pour la recherche documentaire *ad hoc*, l'utilisateur spécifie son besoin en information par une requête qui déclenche la recherche (exécutée par un SRI) des documents pertinents.

— **Collection de documents** : la collection de documents (ou corpus) constitue l'ensemble des informations (des documents) exploitables et accessibles. Nous utiliserons dans la suite du manuscrit les termes : corpus ou collection.

— **Un document** : un document est souvent assimilé à une pièce écrite, servant d'information ou de preuve. C'est un support qui contient des informations qui peuvent être hétérogènes au niveau de la forme et du contenu. Le contenu peut être du texte, graphique, image, son, vidéo, références, liens hypertextes, etc. Quant à la forme, les documents peuvent être structurés, semi-structurés ou non structurés. Les méta-données peuvent apporter des informations supplémentaires aux documents décrivant son contexte, comme par exemple date de rédaction, support de publication, etc.

— **La pertinence** : les chercheurs et les concepteurs de SRI ont distingué deux types de pertinence, à savoir : la pertinence *système*, c'est-à-dire l'évaluation par un système de l'adéquation entre des documents et une requête, et la pertinence *utilisateur* qui se traduit par des jugements de pertinence sur les documents fournis en réponse à une requête. Dans (Mizzaro, 1997), l'auteur met en évidence la complexité et la diversité des types de pertinence. Il définit la pertinence de manière générale comme une relation entre deux entités, l'une étant liée à la collection de documents et l'autre à l'utilisateur. Il distingue ainsi trois dimensions : (1) **le contexte** qui comprend par exemple les documents déjà connus de l'utilisateur (et qui ne seront donc pas pertinents pour lui) ou le temps voire l'argent alloué pour la recherche, (2) **la tâche** qui représente l'activité que l'utilisateur va réaliser avec les documents retrouvés ; et (3) **le domaine du sujet**, le champ disciplinaire auquel l'utilisateur se rapporte.

  Par ailleurs, une définition de la pertinence donnée par (Boughanem et Savoy, 2008), consiste en la correspondance entre un document et une requête, ou encore la mesure d'informativité du document par rapport à la requête.

Nous abordons dans la suite de cette section le processus général d'un système de RI avec les différentes étapes de sa mise en œuvre. Nous passons ensuite en revue les principaux modèles de RI et nous présentons le processus de reformulation de requêtes. Nous décrivons enfin la démarche classique d'évaluation des systèmes de RI classique.

2.2 Mise en œuvre d'un SRI

Un système de recherche d'information (SRI) permet d'automatiser la tâche de RI. L'objectif d'un SRI est de retourner un ensemble de documents, en réponse au besoin en information de l'utilisateur, traduit par la requête qu'il a soumis. L'enjeu est de trouver et retourner l'information pertinente qui satisfait le besoin en information. Pour cela, un SRI emploie un processus bien défini pour identifier les documents pertinents, à partir d'une collection, en réponse à un besoin en information spécifique.

Ce processus est composé de trois fonctions principales, à savoir : (1) l'indexation des documents et des requêtes ; (2) l'appariement requête-document, qui permet de comparer la requête et le document ; (3) et la fonction de modification, qui intervient en réponse aux résultats obtenus. Les modifications éventuelles concernent les documents (ajout ou suppression éventuels de la collection documentaire) ou la requête sachant que les plus courantes concernent la requête seulement, appelée : *reformulation de la requête* ou encore *expansion de requête*.
Dans ce qui suit, nous introduisons dans un premier temps les éléments de base de la mise en oeuvre du système de RI, à savoir l'indexation et l'interrogation.

2.2.1 Indexation

L'objectif principal de cette étape est de fournir des représentations des documents et des requêtes facilement exploitables par le système dans la phase de recherche. Cette représentation est souvent une liste pondérée de mots-clés significatifs que l'on nomme descripteurs du document (ou de la requête). Dès lors, l'indexation consiste à détecter les termes les plus représentatifs, *i.e.*, discriminants, du contenu du document.

Le processus d'indexation est constitué de trois étapes principales, à savoir : (1) définir la source de données, (2) transformer le contenu du document pour générer un aperçu ou une approche logique ; et (3) construire l'index selon cette approche. La tâche d'indexation recouvre un ensemble de techniques visant à transformer les documents pour faciliter l'accès à l'information textuelle. Plus spécifiquement, le but principal de cette étape est de fournir des représentations des documents et des requêtes facilement exploitables par le système dans la phase de recherche. L'indexation peut être manuelle, semi-automatique ou automatique. Elle est définie comme suit :

— **Indexation manuelle** : C'est un spécialiste ou un documentaliste qui analyse le document et sélectionne par la suite les termes qu'il juge représentatifs. L'indexation manuelle fournit une terminologie spécifique pour indexer et rechercher les documents, garantissant ainsi une meilleure représentation des documents et une meilleure qualité des résultats. Ce type d'indexation permet d'avoir un vocabulaire d'index contrôlé, ce qui permet d'accroître la consistance et la qualité de la représentation obtenue.

— **Indexation automatique** : C'est la forme la plus répandue d'indexation utilisée par la plupart des SRI. Le processus consiste en la production automatique des descripteurs (termes d'index) d'un texte. Dans le cas des documents textuels, chaque terme est un élément potentiel de l'index du document qui le contient. Il est identifié selon un processus standard intégrant l'extraction, la suppression des mots vides, la normalisation et la pondératio (Pirkola et Järvelin, 2001).

— **Indexation semi-automatique** : Elle se base sur l'indexation automatique. Toutefois, une intervention humaine peut être réalisée afin d'effectuer des choix sur les termes

significatifs, dans le but de valider la représentation finale des descripteurs. Ces choix sont souvent réalisés en utilisant un thésaurus ou une base terminologique qui est une liste organisée de descripteurs (mots-clés) liés à des règles terminologiques propres et reliés entre eux par des relations sémantiques.

D'une manière générale, l'indexation comprend un certain nombre de traitements qui sont appliqués sur les documents et les requêtes. On distingue : l'extraction des mots, l'élimination des mots vides de sens, la lemmatisation et la pondération. À la fin de cette étape, les documents sont représentés dans des fichiers index qui stockent la cartographie des couples terme-document en y associant un poids. La formule de pondération la plus utilisée est celle basée sur la fréquence des termes dans les documents, appelée $tf - idf$ (Salton et McGill, 1986). Ces facteurs permettent de combiner les pondérations locales (dans le document) et globales (dans la collection) d'un terme.
La mesure $tf - idf$ est donnée par la multiplication des deux mesures TF et IDF comme suit :

$$\text{tf-idf} = log(1 + tf) * idf \tag{1.1}$$

Les mesures tf et idf sont définies comme suit :

1. tf (Term Frequency) : Cette mesure est proportionnelle au nombre d'occurrences d'un terme dans un document (pondération locale). Toutefois, il existe différentes variantes de cette mesure qui dépendent de la façon dont la pertinence est mesurée.

$$\begin{aligned} tf_{ij}^1 &= 1 + log(td_{ij}) \\ tf_{ij}^2 &= \frac{td_{ij}}{\sum_k td_{kj}} \end{aligned} \tag{1.2}$$

   où td_{ij} est le nombre d'occurrences du terme t_i dans le document d_j. Le dénominateur est la taille du document d_f en nombre de termes. La dernière déclinaison permet de normaliser la fréquence du terme pour éviter les biais liés à la longueur du document.

2. idf (Inverse Document Frequency) : ce facteur mesure l'inverse de la fréquence d'un terme dans toute la collection, définie comme la pondération globale. En effet, un terme fréquent dans la collection, a moins d'importance qu'un terme moins fréquent. Cette mesure est exprimée selon l'une des déclinaisons suivantes :

$$\begin{aligned} idf_t^1 &= log\frac{N}{n_t}, idf_t \\ idf_t^2 &= log(\frac{N - n_t}{n_t}) \end{aligned} \tag{1.3}$$

   où N est la taille (nombre de documents) de la collection et n_t le nombre de documents contenant le terme t_i.

2.2.2 Interrogation

L'interrogation est le processus qui consiste à soumettre une requête par l'utilisateur au SRI comme l'illustre la Figure 1.1. C'est l'interaction entre l'utilisateur et le SRI qui comprend trois phases, à savoir : (1) l'utilisateur formule son besoin en information par la requête qu'il

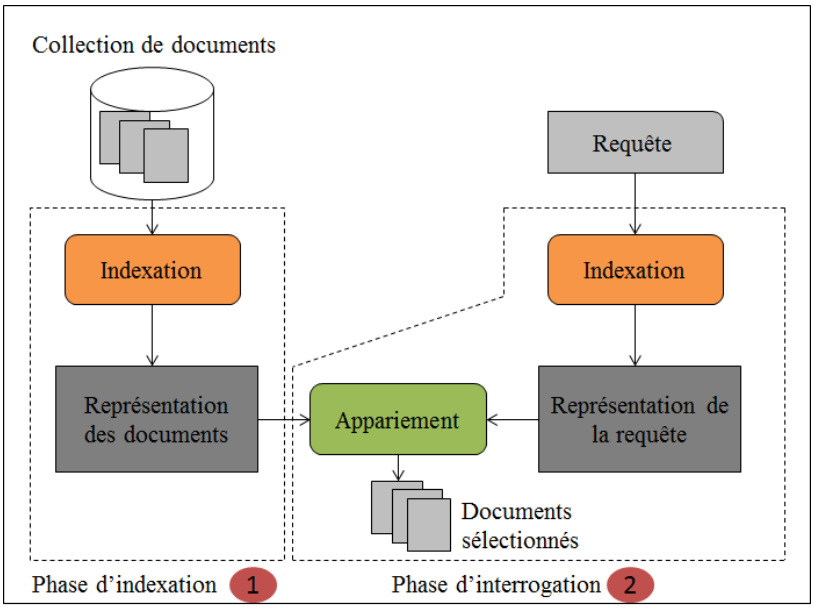


FIGURE 1.1 – Processus en U de la RI.

soumet au système ; (2) la requête est représentée selon un langage d'indexation défini ; et (3) la correspondance entre la requête et les documents est effectuée par exploitation de l'index. Plus précisément, l'interrogation décrit le scénario suivant : l'utilisateur exprime son besoin en information sous la forme d'une requête. Le système interprète la requête et crée son index qui sera compatible avec le modèle d'index des documents. Le système évalue ensuite la pertinence des documents par rapport à cette requête en utilisant une fonction de correspondance. C'est l'appariement requête-document qui permet d'associer à chaque document une valeur de pertinence vis à vis d'une requête. Plus concrètement, c'est un poids de pertinence calculé, reflétant le degré de similarité entre la requête et le document, noté $RSV(q, d)$, où q représente la requête de l'utilisateur et d le document considéré. Le résultat est une liste de documents (pondérés ou non), généralement triée par ordre de valeur de correspondance décroissante, du plus pertinent au moins pertinent. L'ensemble de documents renvoyés par le système est ensuite jugé par l'utilisateur selon son besoin, la satisfaction et le contexte de recherche.

Pour le mécanisme d'appariement document-requête, il est indispensable d'utiliser un modèle qui assure le calcul du degré de pertinence des documents pour les requêtes. Dans la section qui suit, nous décrivons les principaux modèles de RI.

3 Taxonomie des modèles de RI

Un modèle de recherche d'information est au cœur d'un SRI. Il est composé des représentations des documents, des représentations des requêtes d'un utilisateur, et d'une fonction de correspondance entre ces deux éléments. Les modèles de RI fournissent un cadre pour interpréter la notion de pertinence vis-à-vis d'un besoin en information. Appelés aussi modèles d'appariement document-requête, ils peuvent être classés en trois catégories principales, à savoir : modèles booléens, modèles vectoriels et modèles probabilistes. La Figure 1.2 présente

une classification des différents modèles de RI, proposée dans (Baeza-Yates *et al.*, 1999). Nous distinguons les modèles suivants :

— *Les modèles booléens* (Salton, 1969) sont fondés sur la théorie des ensembles et l'algèbre de Boole pour modéliser l'appariement document-requête. Son principe est simple, et repose sur trois éléments : (1) des requêtes exprimées en expressions booléennes, (2) une notion de pertinence ; et (3) de poids binaires (présence/absence). Il existe trois variations principales : le modèle booléen classique, le modèle booléen étendu et le modèle booléen flou.

— *Les modèles vectoriels* (Salton *et al.*, 1975) reposent sur une représentation vectorielle des documents et des requêtes. Il revient sur les limites du modèle booléen en proposant, à la différence des poids binaires, un poids positif de chaque terme pour les documents et les requêtes. Ils comprennent le modèle vectoriel généralisé, le modèle LSI (*Latent Semantic Indexing*) et le modèle connexionniste.

— *Les modèles probabilistes* (Maron et Kuhns, 1960; Robertson et Sparck Jones, 1988; Salton et McGill, 1986) ont été adoptés pour modéliser le degré de pertinence. D'une manière générale, ils calculent la probabilité qu'un document soit pertinent pour une requête donnée. Ils englobent le modèle de langue, le modèle probabiliste général et le modèle de réseau inférentiel (*Document Network*).

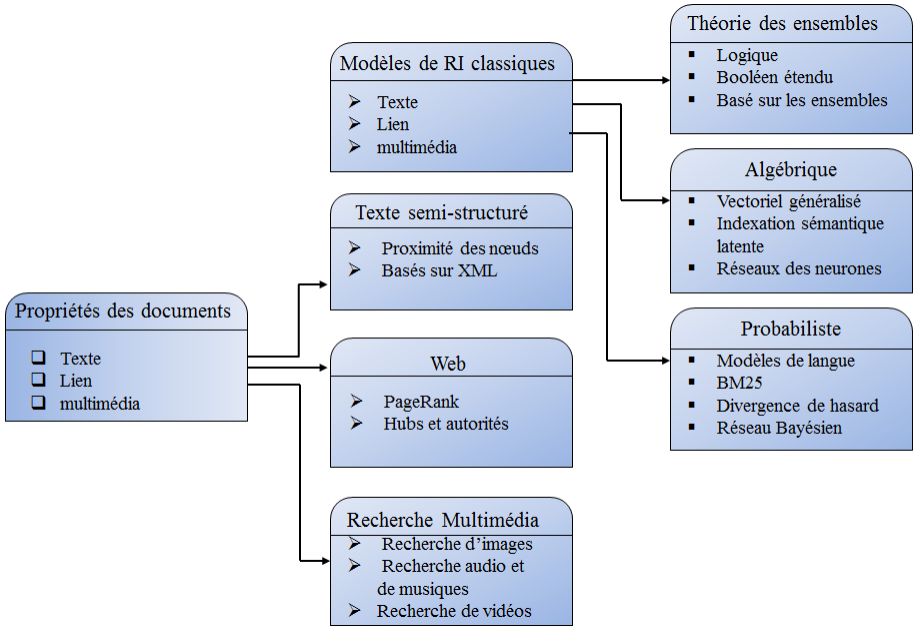


FIGURE 1.2 – Taxonomie des modèles en RI (Baeza-Yates et Ribeiro-Neto, 2011).

Nous détaillons dans ce qui suit ces différents modèles, en mettant l'accent sur les techniques utilisées par la communauté RI.

3.1 Modèle booléen

Le modèle booléen a été introduit par (Salton, 1969). C'est le premier modèle utilisé, et il est considéré le plus simple des modèles du domaine de la RI. Un modèle booléen est basé principalement sur la théorie des ensembles et l'algèbre de Boole, où les termes de requêtes

sont soit présents, soit absents (poids binaire des termes 0 ou 1), et un document est soit pertinent, soit non pertinent (pertinence binaire, jamais partielle).

La spécificité du modèle booléen est que la requête s'exprime avec des opérateurs logiques "AND", "OR", "NOT" et le document est pertinent si et seulement si son contenu respecte la formule logique demandée. Le score de chaque document sera ainsi représenté respectivement par 0 ou 1 selon la fonction booléenne RSV qui calcule la similarité entre la requête q et le document d. Un document d est représenté par un ensemble de termes comme suit : $d = t_1 \wedge t_2 \wedge t_3 ... \wedge t_n$; une requête q est représentée par une expression logique de termes comme suit : $q = (t1 \wedge t2) \vee (t3 \wedge t4)$. La fonction de correspondance est basée sur l'hypothèse de présence/absence des termes de la requête dans le document et vérifie si l'index de chaque document d implique l'expression logique de la requête q. Le résultat de cette fonction, décrite comme $RSV(q, d)$ est binaire (0 ou 1), soit :

— $RSV(d, ti) = 1$ si $ti \in d; 0$ sinon
— $RSV(d, q1 \wedge q2) = 1$ si $RSV(d, q1) = 1 \wedge RSV(d, q2) = 1; 0$ sinon
— $RSV(d, q1 \vee q2) = 1$ si $RSV(d, q1) = 1 \vee RSV(d, q2) = 1; 0$ sinon
— $RSV(d, \neg q) = 1$ si $RSV(d, q) = 0; 1$ sinon

3.2 Modèle vectoriel

Le modèle vectoriel a été introduit par (Salton *et al.*, 1975), représentant les modèles algébriques/statistiques, qui se basent sur un aspect quantitatif des termes et des documents avec un degré de similarité entre une requête et un document. Tout modèle vectoriel est basé sur l'hypothèse suivante : plus deux représentations contiennent les mêmes éléments, plus la probabilité qu'elles représentent la même information est élevée. Documents et requêtes sont représentés ainsi par un vecteur tel que :

— Les coordonnées du vecteur sont exprimées dans un espace euclidien à n dimensions (n est le nombre de termes) ;
— La longueur du vecteur (*i.e.* de sa projection sur chacun des axes/termes) est proportionnelle au poids des termes ;
— La pertinence du document correspond au degré de similarité entre le vecteur de la requête et celui du document.

Ce modèle représente les documents et les requêtes dans l'espace vectoriel des termes de l'index, comme illustré dans la Figure 1.3 qui considère un index de deux termes.
Le coefficient de similarité (*i.e.*, RSV) est calculé entre chaque document et chaque requête afin de trouver les documents dont le vecteur de représentation est le plus colinéaire avec le vecteur de la requête. La corrélation de deux vecteurs document-requête, qui représente la mesure de similarité entre les vecteurs de q et de d, peut être calculé par :

— *Produit scalaire* :

$$RSV(\overrightarrow{\mathrm{q}}, \overrightarrow{\mathrm{d_j}}) = cos(\overrightarrow{\mathrm{q}}, \overrightarrow{\mathrm{d_j}}) \tag{1.4}$$

— *Mesure Cosinus* :

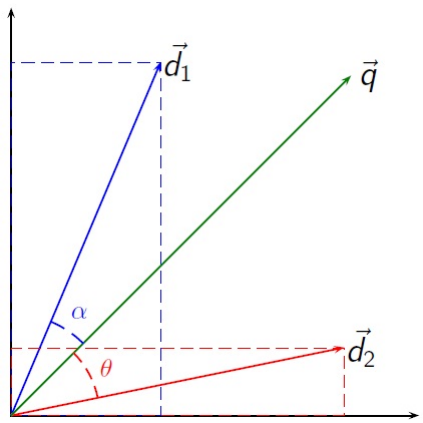


FIGURE 1.3 – Représentation algébrique des documents et des requêtes dans l'espace des termes à deux dimensions.

$$RSV(\overrightarrow{\mathrm{q}}, \overrightarrow{\mathrm{d_j}})) = \frac{\overrightarrow{\mathrm{q}} \cdot \overrightarrow{\mathrm{d_j}}}{\left|\overrightarrow{\mathrm{q}}\right| * \left|\overrightarrow{\mathrm{d_j}}\right|} \tag{1.5}$$

où $\left|\overrightarrow{\mathrm{x}}\right|$ représente la norme euclidienne du vecteur $\overrightarrow{\mathrm{x}}$.

— *La mesure de Jaccard* :

$$RSV(q, d_j) = \frac{|\mathrm{q} \cap \mathrm{d_j}|}{|\mathrm{q} \cup \mathrm{d_j}|} \tag{1.6}$$

où $|\mathrm{q} \cap \mathrm{d_j}|$ correspond au nombre de termes présents à la fois dans la requête q et le document d_j, tandis que $|\mathrm{q} \cup \mathrm{d_j}|$ représente le nombre de termes contenus dans la requête q ou le document d.

— *La mesure de Dice* :

$$RSV(q, d_j) = \frac{2 * |\mathrm{q} \cap \mathrm{d_j}|}{|\mathrm{q}| + |\mathrm{d_j}|} \tag{1.7}$$

où $|\mathrm{d_j}|$ et $|\mathrm{q}|$ désignent le nombre de termes dans le document d_j et dans la requête q, respectivement.

Les avantages principaux du modèle vectoriel sont les suivants : tout d'abord, la pondération non binaire des termes favorise une meilleure qualité des résultats. De plus, le modèle permet une correspondance partielle ou approximative entre les documents et les requêtes (*best match*). Les documents sont triés selon leur degré de similarité vis-à-vis de la requête. La longueur des documents est traitée naturellement dans l'appariement, car elle est considérée dans le calcul des poids des termes.

3.3 Modèles probabilistes

Les modèles probabilistes ont été proposés par (Robertson, 1997), basés principalement sur les probabilités. Ils permettent la modélisation de la notion de pertinence, en estimant la probabilité de pertinence d'un document par rapport à une requête. L'idée de base est de sélectionner les documents ayant à la fois une forte probabilité d'être pertinents et une faible probabilité

d'être non pertinents. Autres modèles de type probabiliste sont ceux fondés sur les réseaux bayésiens et le modèle de langue. La mesure de similarité document/requête est fondée sur une estimation de probabilité. Différentes approches de calcul de probabilités ont été introduites dans le domaine de la RI, afin d'optimiser les performances de recherche (Robertson, 1997; Song et Croft, 1999), nous citons les approches suivantes :

1. *Approche par modèle classique* : à partir d'un document et d'une requête, on détermine la probabilité d'avoir l'évènement "le document est pertinent pour la requête".
2. *Approche par réseau d'inférence* : à partir du contenu d'un document, on détermine la probabilité pour que la requête soit vraie.
3. *Approche par modèle de langue* : déterminer la probabilité pour qu'une requête soit générée à partir d'un document.

Ainsi, on modélise la pertinence comme un évènement probabiliste : pour une requête donnée q, il s'agit d'estimer $P(q|d)$ la probabilité qu'on obtienne une information pertinente par le document d. On peut estimer de la même façon $P(NR|d)$ la probabilité de non pertinence de d. On retourne le document d si $P(R|d) > P(NR|d)$. D'après le théorème de Bayes, la probabilité que le document d soit pertinent pour la requête q est donnée par :

$$P(R, d|q) \propto \prod_{t \in q} P(R, d|t) \tag{1.8}$$

ou $P(R, d|t)$ est la probabilité que le document d soit pertinent en observant le terme t. Plus en détail, cette formule peut être représentée comme suit :

$$P(R, d|t) = \frac{P(t, d|R) * P(R)}{P(t)} \tag{1.9}$$

Le modèle probabiliste est représenté par deux modèles phares, à savoir : modèle de langue (2poisson) et modèle $BM25$ (Okapi) détaillés ci dessous.

— **Modèle de pondération BM25** : le schéma de pondération est la fonction d'appariement document-requête du modèle BM25 [1] (Robertson *et al.*, 1996) est un des modèles les plus performants en RI. Le calcul du poids d'un terme dans un document intègre différents aspects relatifs à la fréquence locale des termes (tf_i), leur rareté et la longueur des documents. Il est calculé comme suit :

$$x_i = \frac{(k_1 + 1) * tf_i}{k_1 * (1 - b + b * \frac{dl}{avgdl}) + tf_i} \tag{1.10}$$

avec dl la taille du document d_j, $avgdl$ est la moyenne des tailles des documents dans la collection ; et k_1, b sont des paramètres qui déterminent l'importance de la fréquence du terme dans le document et dans la requête.

— **Modèle de langue** : le modèle de langue pour la RI repose sur l'idée que l'utilisateur lorsqu'il formule sa requête a une idée du document idéal qu'il souhaite trouver et que la requête est formulée pour trouver ce document idéal (Maron et Kuhns, 1960). Cette idée a été formulée depuis les années 60 dans les premiers travaux des RI probabilistes, et un des premiers modèles de langues en RI a été proposé par (Ponte et Croft, 1998). L'idée des modèles de langue est de capter les régularités linguistiques d'une langue,

1. Best Match

proposant un modèle probabiliste qui assigne une probabilité à toute séquence de mots. Plus précisément, ces modèles visent à déterminer la probabilité que la requête q soit générée par le modèle de langue M_d du document d. Le principe est d'ordonner chaque document d de la collection C suivant leur capacité à générer la requête q. Ainsi, il s'agit d'estimer la probabilité de génération $P(q|d)$ comme suit :

$$P(q|d) = \prod_{t \in q} P(t|\Theta_d)^{c(t;q)} \tag{1.11}$$

où $c(t;q)$ est la fréquence du terme t dans la requête q, et Θ_d est le modèle de langue du document.

Les modèles de langue se différencient par le type de distribution probabiliste où on observe la présence ou l'absence des termes dans les documents ou les requêtes.

4 Reformulation de la requête

La reformulation du besoin en information consiste à redéfinir le besoin de l'utilisateur au fur et à mesure de la session de recherche. Cette étape peut être effectuée de différentes manières :

— Manuellement, dans le cas où l'utilisateur soumet lui-même une nouvelle requête.

— De façon automatique, lorsque le système de RI s'appuie sur les termes importants dans les documents les plus pertinents ou visités par l'utilisateur, qui sont réutilisés.

L'approche automatique suppose que l'utilisateur soumet une requête au SRI afin de récupérer les documents pertinents, sinon cette requête devrait être ré-écrite ou reformulée afin de récupérer plus de documents pertinents. La Figure 1.4 donne un aperçu du principe de la reformulation de requête dans le cadre d'un SRI.

La reformulation de requête est un processus qui a pour objectif de générer une nouvelle requête plus adéquate afin d'obtenir un ensemble de résultats plus pertinents, à partir de connaissances du domaine cible, en utilisant les concepts clés contenus dans les documents. La requête initiale est formulée par l'utilisateur, sa modification peut se faire soit par **réinjection de pertinence** (*relevance feedback*) (Salton et Buckley, 1997), soit par **expansion de requêtes** (*query expansion*) (Efthimiadis, 1996).

La reformulation de la requête se fait en deux étapes principales, à savoir : (*i*) trouver des termes d'extension à la requête initiale, et (*ii*) re-pondérer les termes dans la nouvelle requête.

Le principe fondamental de la stratégie d'expansion de requête est de comparer simplement le contenu de la requête avec les documents de la collection. L'ensemble des documents pertinents restitué est alors très souvent incomplet. Des travaux de recherche ont proposé d'ajouter d'autres termes contenus dans les documents pertinents ou d'ajouter des termes sémantiquement proches ou encore d'ajouter des termes voisins en utilisant des calculs de poids de similarité entre termes. Différentes méthodes d'expansion de requêtes ont été proposées dans la littérature (Xu et Croft, 1996; Adriani et Rijsbergen, 1999; Baziz *et al.*, 2003; Latiri *et al.*, 2012; Carpineto et Romano, 2012; Nawab *et al.*, 2016). L'objectif de ces méthodes d'expansion est d'augmenter le nombre de documents pertinents retrouvés ainsi que d'améliorer le classement des documents les plus pertinents.

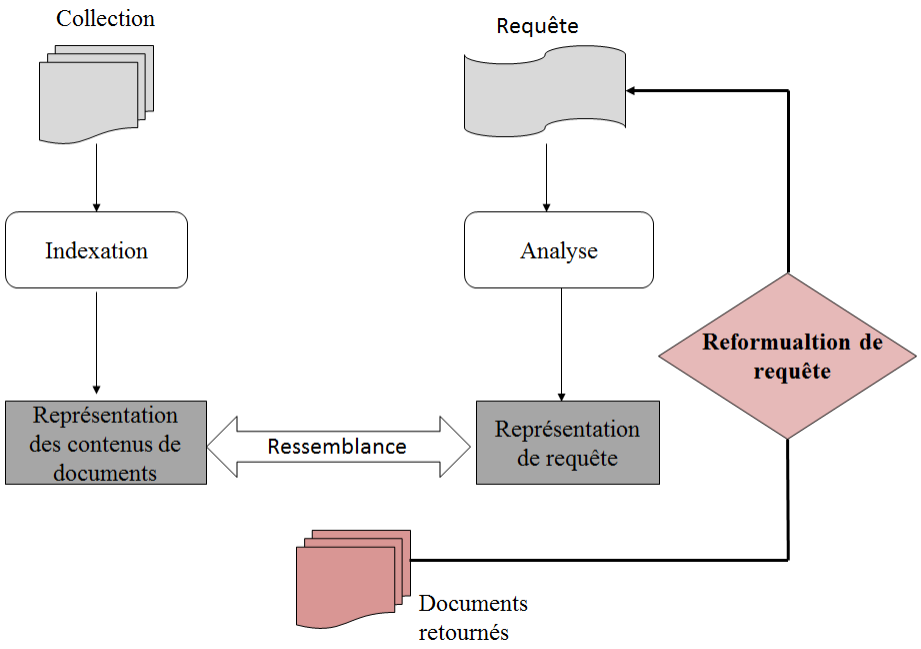


FIGURE 1.4 – Aperçu du processus de la reformulation de requête.

Dans le cadre de la RI, la barrière de la non compatibilité entre les termes des documents et termes des requêtes, connu aussi par "*term mismatch*" représente un obstacle pour retourner des résultats. Cela est connu comme étant le problème de vocabulaire (Furnas *et al.*, 1987), amplifié par les synonymes (mots différents avec le même sens comme "java"), la polysémie (différents termes avec le même sens, comme "tv" et "télévision"). Les synonymes avec les inflexions de mots (comme pour les formes au pluriel "télévision" et "télévisions"), peut entrainer un échec pour récupérer les documents pertinents, avec une diminution du rappel (la capacité du système à retourner tous les documents pertinents en réponse à la requête). Par ailleurs, la polysémie est à l'origine de la récupération de documents erronés et non pertinents, ce qui implique une baisse dans la précision des résultats (la capacité de retourner uniquement les documents pertinents).

Pour faire face au problème de vocabulaire, plusieurs approches on été proposées, y compris les raffinements interactifs de requêtes, la reformulation de requêtes par réinjection de la pertinence, la désambiguisation des sens de mots et le *clustering* des résultats de recherche. Une des approches les plus naturelles et ayant du succès, est la technique d'expansion/de reformulation des requêtes initiales avec d'autres termes qui représentent au mieux l'intention des utilisateurs, ou simplement produire une requête plus utile et plus susceptible de récupérer des documents pertinents.

Durant la dernière décennie, un grand nombre de techniques d'expansion automatique de requêtes a été présenté en utilisant une variété d'approches qui se basent sur plusieurs sources de données et utilisent des méthodes sophistiquées pour trouver de nouvelles fonctionnalités en corrélation avec les termes de la requête (Mitra *et al.*, 1998; Carpineto *et al.*, 2002; Liu *et al.*, 2004; Lee *et al.*, 2008; Latiri *et al.*, 2012). Ces contributions ont montré à travers les études expérimentales que les résultats de l'expansion automatique de la requête permettent

de donner des résultats plus pertinents avec des améliorations de 10% et plus.

Dans (Carpineto et Romano, 2012), les auteurs ont conduit une large étude comparative des approches qui concernent l'expansion automatique des requêtes. Ils ont montré que malgré la persistance du problème de vocabulaire dans certains travaux, l'expansion automatique de requêtes a le potentiel de surmonter le problème majeur des SRI, à savoir : la difficulté à fournir aux utilisateurs une description plus précise de leur besoin en information.

Dans (Huang et Efthimiadis, 2009), les auteurs ont analysé et évalué différentes stratégies de reformulation de requêtes à partir des fichiers logs du web. De cette analyse, ils ont conclu que les différentes stratégies de reformualtion ont différentes caractéristiques, et que ceux qui semblent les plus effectives sont ajout/suppression de mots, substitution de mots, expansion avec acronymes et la correction orthographique.

D'autres travaux se sont intéressés à l'expansion sémantique des requêtes, en vue de pallier le problème de dérive sémantique, *i.e.*, *semantic mismatch*, (Packer *et al.*, 2012; Curé *et al.*, 2013). En effet, ils considèrent que les termes d'expansion qui ont une liaison sémantique avec les termes initiaux de la requête rajoutent plus de spécificité au contexte de la recherche et améliore donc la pertinence des résultats. Différents domaines ont adopté l'expansion sémantique, comme la RI sociale, la RI biomédicale, etc. Ils utilisent les concepts issus des terminologies et des thésaurus ou encore les relations sémantiques entre les termes ou concepts afin de mieux représenter le thème, *i.e.*, le *topic*, de la requête.

Par ailleurs, la reformulation par retour de pertinence propose de formuler la requête initiale pour amorcer la recherche d'information, puis la modifier d'une manière itérative à partir des jugements de pertinence et/ou de non-pertinence de l'utilisateur afin d'ajuster la requête par expansion, re-pondération ou combinaison des deux procédures, jusqu'à ce que le résultat de la recherche soit satisfaisant.

Dans (Carpineto et Romano, 2012), les auteurs ont donné une classification des approches et techniques d'expansion automatique de requêtes en cinq groupes selon le paradigme conceptuel utilisé pour trouver les propriétés d'expansion, à savoir : méthodes linguistiques, approches statistiques spécifiques au contexte, approches statistiques spécifiques aux requêtes, analyse des fichiers de log et les données du web.

Nous détaillons dans ce qui suit deux principales techniques de reformulation de la requête, à savoir : (1) reformulation par réinjection de la pertinence (*relevance feedback*), (2) reformulation par pseudo-réinjection de la pertinence (*pseudo-relevance feedback* ou *blind query expansion*) (Rocchio, 1971).

4.1 Reformulation par réinjection de la pertinence

La réinjection de la pertinence est une technique utilisée pour améliorer la performance de la RI (Rocchio, 1971; Salton et Buckley, 1997). Au cours de ce processus, l'utilisateur utilise une requête initiale, puis fournit un retour sur la pertinence des documents. Les termes de ces documents (jugés pertinents) sont donc ajoutés à la requête initiale. La reformulation par réinjection de la pertinence est une technique qui vise à améliorer la qualité de recherche lorsque la seule évaluation de la similarité entre les requêtes et les documents n'est plus suffisante. Le principe de la reformulation par réinjection de pertinence se résume en quatre étapes principales, à savoir :

1. Les utilisateurs effectuent une première requête ;
2. Des documents sont retournés en fonction de cette première interrogation ;
3. Les utilisateurs doivent ensuite indiquer parmi les documents retournés, lesquels sont pertinents, et/ou lesquels ne le sont pas ;
4. La requête de départ est alors modifiée automatiquement pour tenir compte des jugements des utilisateurs.

La méthode de la réinjection de la pertinence a été utilisée dans différents domaines de recherche, integrée dans des SRI (Kwan *et al.*, 2015), utilisée dans le cadre de la RI d'image (Duan *et al.*, 2016) ou encore pour la recherche de vidéos (Fernandez-Beltran et Pla, 2016). Ces travaux ont montré une amélioration de performance en utilisant cette méthode par rapport aux techniques standards de recherche.

4.2 Reformulation par pseudo-réinjection de la pertinence

La reformulation par pseudo-réinjection de la pertinence (*Blind Feedback* ou encore *Pseudo Relevance Feedback*, notée PRF) utilise des techniques de réinjection automatique à l'aveugle pour construire la nouvelle requête. L'idée de base de la PRF est basée sur l'hypothèse que les premiers documents pertinents contiennent de nombreux termes utiles qui aident à distinguer les documents pertinents des non pertinents. En général, les termes d'expansion sont extraits soit selon leur distribution dans les documents retournés, ou selon la comparaison entre la distribution de termes dans les documents retournés et l'ensemble de documents de la collection. Plusieurs autres critères ont été proposés par exemple *idf* (Rocchio, 1971). De plus, la PRF est une technique courament utilisée pour faire face à l'explosion de l'information sur le web afin d'améliorer la performance de recherche (Buckley *et al.*, 1992; Yu *et al.*, 2003).

L'utilisation de Pseudo-Relevance Feedback a fait l'objet d'un grand nombre d'études depuis plusieurs décennies et beaucoup de modèles ont été proposés dans ce cadre (Thesprasith et Jaruskulchai, 2014; Min *et al.*, 2010; Hammache *et al.*, 2013). À titre d'exemple, dans (Hammache *et al.*, 2013), les auteurs additionnent les poids des relations d'un terme candidat avec chacun des termes de la requête pour déterminer les termes d'expansion. Les termes candidats sont choisis s'ils sont fortement en relation avec les termes de la requête. Ils ont intégré cette technique dans le cadre de modèle de langue. Dans le même cadre du modèle de langue, dans une étude plus récente (Hazimeh et Zhai, 2015), les auteurs procèdent à une analyse des méthodes de lissage dans les modèles de langue pour la PRF. Par ailleurs, (Li et Wang, 2012) a utilisé la technique de PRF pour estimer la difficulté des requêtes permettant d'estimer la performance de la recherche pour les requêtes de recherche d'images. Le Tableau 1.1 montre quelques fonctions de classement de termes basées sur la distribution des termes dans les documents de pseudo-réinjection de la pertinence.

La notation dans le Tableau 1.1 est comme suit :

— t est un terme ;

— $w(t, d)$ indique le poids du terme t dans le document de pseudo-réinjection de pertinence d ;

— $p(t|R)$ et $p(t|C)$ représentent respectivement la probabilité d'occurrence du terme t dans les documents de la pseudo-réinjection de pertinence R ainsi que dans toute la collection de documents C

Référence	Fonction	Forme mathématique
(Rocchio, 1971)	Poids de Rocchio	$\sum_{d \in R} w(t,d)$
(Robertson et Sparck Jones, 1988)	Modèle Indépendant Binaire	$log \frac{p(t\|R)[1-p(t\|C)]}{p(t\|C)[1-p(t\|R)]}$
(Doszkocs, 1979)	Chi-square	$\frac{[p(t\|R)-p(t\|C)]^2}{p(t\|C)}$
(Robertson, 1991)	Robertson selection value (RSV)	$\sum_d w(t,d).[p(t\|R) - p(t\|C)]$
(Carpineto *et al.*, 2001)	Kullback-Leibler distance (KLD)	$p(t\|R).log \frac{p(t\|R)}{p(t\|C)}$

Tableau 1.1 – Fonctions principales de classement de termes basés sur l'analyse des distributions des termes dans les documents de pseudo-réinjection de pertinence (Carpineto et Romano, 2012).

5 Évaluation des performances en RI

L'évaluation est un processus primordial pour toutes les tâches de recherche d'information afin d'estimer leur performance. C'est une stratégie qui permet d'étudier les systèmes de RI et d'identifier l'impact des méthodes et techniques employées dans les approches de recherche. Un but d'un SRI est de satisfaire au mieux le besoin en information des utilisateurs, en retournant les documents pertinents. De ce fait, la qualité des résultats par rapport au besoin exprimé par l'utilisateur ainsi que la capacité du système à retourner ces résultats d'une manière optimale en temps et avec facilité représentent les principaux facteurs pour évaluer un SRI (Heuwing et Mandl, 2007).

Dans cette section, nous présentons le cadre d'évaluation d'un système de RI ainsi que les mesures d'évaluation sous-jacentes.

5.1 Collections de test

Les collections de test permettent de comparer directement des résultats obtenus par des systèmes en utilisant des modèles différents. Nous détaillons ci-dessous différents éléments de ces collections.

— **Les requêtes** sont un ensemble de besoins en information utilisés pour le test. Cet ensemble est appelé également *topics* dans le jargon des campagnes d'évaluation telles que TREC, CLEF, etc. Le nombre de requêtes doit être important afin d'être le plus représentatif possible de la réalité et pour avoir une évaluation objective. Il faut au moins 25 requêtes pour garantir la qualité de l'évaluation au regard de la statistique (Buckley et Voorhees, 2000). Les requêtes sont souvent créées par les assesseurs des organismes qui organisent l'évaluation. Toutefois, elles peuvent être de vraies requêtes extraites à partir des logs des moteurs de recherche (Baeza-Yates *et al.*, 1999).

— **Le corpus de documents** est l'ensemble de documents pré-sélectionnés. Il existe plusieurs corpus disponibles. Ces corpus diffèrent selon plusieurs critères en fonction de la tâche de recherche que l'on veut évaluer, des documents plus ou moins généraux, plus ou moins spécialisés dans un domaine, exprimés dans une ou plusieurs langues, etc.

— **Les jugements de pertinence** identifient les documents pertinents pour une requête et représentent la vérité du terrain. Un score de pertinence graduel peut éventuellement être associé à chaque couple document/requête. La réalisation de ces jugements est loin d'être une tâche facile. Il s'agit d'un processus long et coûteux impliquant des humains. Pour de petites collections comme celle de Cranfield (Cleverdon, 1991), il existe des jugements de

pertinence exhaustifs pour chaque paire requête-documents. Cependant, pour les grandes collections modernes, les jugements ne se font généralement que pour un sous-ensemble des documents pour chaque requête.

Ces éléments constituent le contexte de l'évaluation, en d'autres termes, ils représentent les éléments principaux qui vont servir à évaluer un modèle de RI. Ce cadre d'évaluation correspond au paradigme de Cranfield (Cleverdon, 1997) qui a suscité la mise en place de plusieurs campagnes d'évaluation. Ces dernières permettent de cibler une tâche particulière et d'évaluer l'efficacité des SRI répondant à cette tâche.

5.2 Campagnes d'évaluation

Les campagnes d'évaluation les plus réputées sont :

1. La campagne TREC [2] (*Text REtrieval Conference*) est une campagne de référence en RI permettant d'évaluer des SRI au regard de la dimension thématique. C'est le protocole le plus utilisé pour évaluer les performances de la RI, fournissant plusieurs collections de test. Initiée en 1992 comme une partie du programme TIPSTER, la campagne TREC est un projet du NIST [3] et la DARPA [4].
2. La campagne CLEF [5](*Conference and Labs of the Evaluation Forum*) est une initiative permettant de promouvoir la recherche d'information multilingue basée sur les langues européennes. Nous trouvons plus de détails sur l'évaluation à base de différentes collections de test dans (Clough et Sanderson, 2013).
3. La campagne INEX (Initiative for the Evaluation of XML Retrieval) oriente ses tâches de recherche vers des collections de documents structurés.
4. Le projet NTCIR [6] a développé diverses collections de test de tailles similaires aux collections de TREC, en se concentrant sur les langues d'Asie de l'Est et la recherche d'information multilingue. Les requêtes sont faites dans une langue, toutefois, les collections de documents contiennent des documents dans une ou plusieurs autres langues.

Nous détaillons dans ce suit les deux campagnes les plus populaires TREC et CLEF ainsi que quelques tâches associées.

5.2.1 Protocole d'évaluation TREC

Les différentes méthodes et mesures d'évaluation sont le cœur des campagnes d'évaluation TREC (Text REtrieval Conference) (Voorhees, 2005). TREC vise à fournir une plate-forme commune pour évaluer les systèmes de recherche d'information en développant des collections de test. Comme présenté ci-dessus une collection de test est constituée d'un corpus de documents, un ensemble de requêtes et un ensemble de jugement de pertinence fournis par des utilisateurs experts. Le corpus de documents et les requêtes associées sont mises à la disposition des différentes équipes participant au TREC.

2. http ://trec.nist.gov
3. National Institute of Standards and Technology
4. Defence Advanced Research Projects Agency
5. http ://www.clef-initiative.eu/
6. http ://research.nii.ac.jp/ntcir/ntcir-12/tasks.html

Les équipes utilisent les méthodes de recherche qu'ils ont développé pour exécuter les requêtes et soumettre les résultats aux organisateurs, sous forme d'une liste ordonnée de documents. Ces derniers évaluent donc les soumissions de chacune des équipes selon les jugements de pertinence.

TREC traite plusieurs domaines et concerne les documents des collections de blogs (Ounis *et al.*, 2006), de vidéos (Smeaton *et al.*, 2006), de tweets (LaRock *et al.*, 2014) et la RI biomédicale dans le cadre de TREC Genomics (Hersh et Voorhees, 2009).

L'évaluation de performance des résultats renvoyés par les différentes approches de RI est principalement basée sur le modèle de Cranfield (Voorhees, 2002). Cette approche d'évaluation est basée sur l'hypothèse que chaque document de la collection est jugé pertinent ou non-pertinent pour chaque requête à évaluer (Cleverdon, 1991). Selon le protocole de TREC, la collection de test contient des documents originaux et elle est fournie pour chaque groupe participant. Une première phase est d'indexer la collection selon une technique choisie par les participants. Ensuite, le SRI retourne les 1000 premiers documents pour chaque requête. Les résultats finaux sont soumis à TREC pour être évalués d'une manière officielle.

Plusieurs tâches sont apparues, considérant diverses dimensions et domaines de la RI, telles que la RI orientée utilisateur, la RI médicale, la RI contextuelle, etc. Parmi ces tâches, nous citons :

1. La tâche *TREC Filtering* : c'est une tâche dans laquelle l'algorithme de recherche décide uniquement si un document est pertinent (cas ou il a été sélectionné) ou pas (cas ou il n'a pas été sélectionné).
2. La tâche *TREC Interactive* : c'est une tâche qui consiste en la résolution d'un besoin complexe. L'utilisateur interagit avec les systèmes de RI pour sélectionner les documents pertinents.
3. La tâche *TREC NLP* : le but de cette tâche est de vérifier si les algorithmes basés sur le traitement du langage naturel présentent des avantages comparés aux algorithmes traditionnels de RI basés sur les termes de l'index.
4. La tâche *Cross language* : c'est une tâche ad-hoc dans laquelle les documents paraissent dans une seule langue mais les requêtes sont de diverses langues.
5. La tâche *TREC Contextual Suggestion* : elle consiste en une suggestion de lieux d'activités à partir d'un ensemble de profils utilisateur et d'un contexte, traduisant respectivement les préférences et la localisation des utilisateurs.

5.2.2 Protocole d'évaluation CLEF

Cette campagne offre une infrastructure pour diverses tâches, à savoir : évaluer les systèmes multilingues et multimodaux ; possibilité de réglage de paramètres et d'évaluation ; possibilité d'utiliser et d'accéder aux données non structurées, semi-structurées ou très structurées ainsi que les données sémantiquement enrichies ; création des collections de test réutilisables pour les analyses comparatives ; exploration de nouvelles méthodes et approches d'expérimentations et d'évaluation innovantes ; et possibilité de discuter les résultats, comparer les approches et échanger les idées afin de partager les connaissances.

CLEF propose différentes pistes d'évaluation : *Ad-Hoc Track editions in CLEF* pour la recherche de documents multilingues (de 2000 à 2009), *Cross-Language Retrieval in Image*

Collections (ImageCLEF)(de 2003 à 2009), Multilingual Web Track (WebCLEF)(de 2005 à 2008) et beaucoup d'autres pistes comme GeoCLEF et videoCLEF. Plus de détails sur l'évaluation à base de collections de test sont donnés dans (Sanderson, 2010).

Différentes tâches apparaissent au fil du temps dans le cadre de plusieurs initiatives CLEF, telles que :

1. *CLEF eHealth* : cette tâche explore des scénarios dont le but est d'aider les patients à comprendre et accéder à l'information médicale en ligne "*eHealth*". Le but est de développer des méthodes et des ressources dans un cadre multilingue afin d'améliorer la compréhension des textes médicaux difficiles. Dans le cadre de CLEF eHealth, il existe deux tâches :
   - Tâche 1 - *Information Extraction from Clinical Data* : elle comprend deux parties, à savoir : (a) *Clinical speech recognition* et (b) *Named entity recognition from clinical narratives in European languages.*
   - Tâche 2 - *User-centered Health Information Retrieval* qui comprend deux parties, à savoir : (a) *Monolingual IR (English)*, et (b) *Multilingual IR (Chinese, Czech, French, German, Portuguese, Romanian).*
2. *ImageCLEF* : *ImageCLEF* offre quatre tâches dans un objectif global de l'évaluation automatique des annotations et d'indexation des images. Cette tâche s'adresse à différents aspects problématiques de l'annotation dans différents domaines, à savoir :
   - Tâche 1 - *Image Annotation* : cette tâche a pour but de développer des systèmes pour l'annotation multi-concept des images, de la localisation ainsi que la génération des descriptions médicales.
   - Tâche 2 - *Medical Classification* : propose de résoudre le problème de la séparation des figures composées de la littérature biomédicale.
   - Tâche 3 - *Medical Clustering* : résout les problèmes de l'agrégation des images représentant différentes parties du corps en radiologie médicale (*x-rays*).
   - Tâche 4 - *Liver CT Annotation* : permet d'étudier les rapports automatiques structurés en offrant un système qui aide à annoter automatiquement les scanners appelés aussi *computed tomography.*
3. *Question answering (QA)* : dans le cadre de cette tâche, un effort pour améliorer les requêtes ou les questions de utilisateurs. En effet, répondre à une question initiale écrite en langage naturel peut demander des données liées à cette requête, des inférences textuelles ou encore un requêtage en texte libre. Les tâches associées sont donc :
   - Tâche 1 – *QALD : Question Answering over Linked Data* ;
   - Tâche 2 – *Entrance Exams : Questions from reading tests* ;
   - Tâche 3 – *BioASQ : Large-Scale Biomedical Semantic Indexing* ;
   - Tâche 4 – *BioASQ : Biomedical Question answering.*

Afin d'accomplir la tâche d'évaluation expérimentale, des SRI expérimentaux ont été développés, les deux principaux dans le domaine de la RI, étant Terrier [7] et Lemur [8]. **Terrier** est un moteur de recherche à accès libre très flexible, efficace et facilement déployable sur les grandes collections de documents. Il offre des fonctionnalités pour différentes tâches de RI, comme

7. http ://terrier.org/
8. http ://www.lemurproject.org/

l'indexation, la recherche et l'évaluation expérimentale sous la dernière version (Terrier 4.0). Le projet **Lemur** développe et offre des moteurs de recherche, les barres d'outils, des outils d'analyse de texte, et les ressources de données qui appuient la recherche et le développement de la RI ainsi que des outils de fouille de textes. Ces deux moteurs de recherche expérimentaux sont les plus utilisés par la communauté de RI. Dans le cadre de notre travail, nous avons utilisé la dernière version de Terrier (Terrier 4.0).

5.3 Mesures d'évaluation en RI

En RI, l'évaluation des résultats de SRI est une phase essentielle et primordiale basée sur les composants (documents, requêtes, jugements) décrits précédemment. La phase expérimentale suppose l'utilisation des métriques d'évaluation qui visent à comparer les modèles utilisés et mettre au point leurs paramètres. En effet, le SRI sélectionne des documents de la collection en utilisant un modèle de RI spécifique et les renvoie à l'utilisateur. Pour mesurer la capacité d'un SRI à retrouver les documents pertinents et à rejeter ceux qui ne le sont pas, deux métriques de base sont utilisés pour évaluer l'efficacité de la RI, à savoir : la précision et le rappel.

Soit $|B|$ le nombre de documents renvoyés par le système pour la requête q, $|S|$ le nombre de documents pertinents dans la collection pour cette requête et $|S+|$ le nombre de documents pertinents renvoyés par le système (*cf.* Figure1.5).

Nous présentons dans la suite les mesures les plus classiques de "Rappel" et de "Précision", ainsi qu'un ensemble de mesures les plus courantes. Le Tableau 1.2 présente les notations utilisées pour les mesures d'évaluation.

Notation	**Decription**
q	la requête q
Q	l'ensemble de requêtes
$\|B\|$	le nombre de documents renvoyés par le système pour la requête q
$\|S\|$	le nombre de documents pertinents dans la collection pour la requête q
$\|S+\|$	Le nombre de documents pertinents renvoyés par le système pour la requête q

Tableau 1.2 – Notations des mesures d'évaluation en RI.

1. **Précision**
   La mesure de précision indique la capacité d'un SRI à retourner des documents pertinents par rapport aux résultats renvoyés par le système. Donnée par le rapport entre les documents sélectionnés comme pertinents et la totalité des documents sélectionnés, la précision est calculée comme suit :

$$Précision = \frac{|S+|}{|B|} \in [0,1] \tag{1.12}$$

2. **Rappel**
   La mesure de rappel calcule la capacité d'un SRI à restituer un maximum de documents pertinents pour une requête. Il mesure la proportion des documents pertinents sélectionnés relativement à l'ensemble des documents pertinents de la collection. Formellement :

$$Rappel = \frac{|S+|}{|S|} \in [0,1] \tag{1.13}$$

   La Figure 1.5 donne une description des mesures du rappel et de la précision.

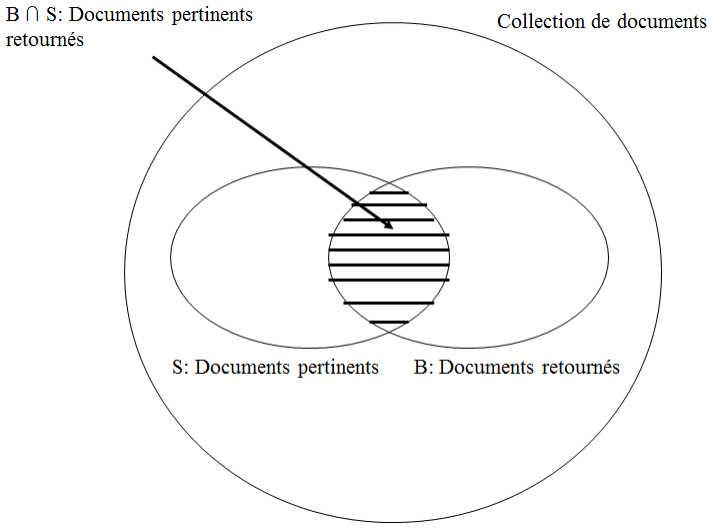


FIGURE 1.5 – Description graphique des éléments nécessaires pour calculer le rappel et la précision.

   La précision et le rappel sont calculés indépendamment de l'ordre avec lequel les résultats sont représentés. Pour pallier ce manque, d'autres mesures tenant compte de l'ordre de classement des documents ont été proposées. Ces mesures ne se basent pas uniquement sur le classement de documents pertinents retournés mais aussi sur le nombre de premiers documents pertinents retournés. L'ordre d'apparition des documents étant crucial, nous citons la *précision@n* documents et la précision moyenne (MAP).

3. **La précision exacte à n documents**
   La *précision@n* est la précision des résultats vis-à-vis de la requête à n documents, et permet de mesurer la capacité d'un SRI à retourner les documents pertinents par les n-premiers documents retournés en réponse à la requête. Par exemple, pour une requête q, une précision à 10 documents notée $P@10$, est calculée par la proportion de documents pertinents retournés parmi les dix premiers documents.

4. **Précision Moyenne (*Mean Average Precision*)**
   C'est la moyenne des valeurs de précision moyennes (*Average Precision-AP*) pour chaque document pertinent, appelée aussi la précision à 11 points de rappel. Elle permet de calculer la précision des résultats vis-à-vis d'un ensemble de requêtes et se focalise principalement sur les documents pertinents classés dans les premiers rangs, soit :

$$MAP = \frac{\sum_{q \in Q} AP_q}{|Q|} \tag{1.14}$$

avec AP_q est la précision moyenne d'une requête q, Q est l'ensemble des requêtes et $|Q|$ est le nombre de requêtes.

5. **F-mesure**
   La F-mesure est une mesure qui combine la précision et le rappel et leur pondération. Cette mesure permet d'évaluer la performance globale du SRI. Dans (Blair et Maron, 1985), l'auteur a proposé de combiner le rappel et la précision dans la F-mesure, qui se calculent comme suit :

$$F_\gamma = \frac{(1+\gamma^2) * (Précision * Rappel)}{(\gamma^2 * Précision + Rappel)} \tag{1.15}$$

   pour des valeurs réelles positives de γ traduisant l'importance relative du rappel et de la précision. Un cas particulier de la mesure générale F_γ est F_1 ($\gamma = 1$), dans ce cas particulier la F-mesure définit la moyenne harmonique du rappel et de la précision :

$$F\text{-}mesure = \frac{2 * Précision * Rappel}{(Précision + Rappel)} \tag{1.16}$$

6. **La mesure BPREF** : pour les collections volumineuses, la construction de jugements de pertinence complets est difficile voir impossible puisque elle est très coûteuse en terme de temps. Afin de pallier cet inconvénient, (Buckley et Voorhees, 2000) ont proposé la mesure BPREF (*Binary PREFerence-based measure*). Cette mesure ne considère que les documents jugés et elle prend en compte les documents pertinents et les documents non pertinents. Elle est donnée par la formule suivante :

$$bpref = \frac{1}{R} \sum_r 1 - \frac{\text{n classés avant r}}{R} \tag{1.17}$$

   Avec R le nombre de documents pertinents pour la requête, r est un document pertinent et n est le nombre de documents non pertinents classés avant le document pertinent r.

6 Conclusion

Ce chapitre a porté principalement sur les notions et concepts de base du domaine de la recherche d'information. Nous avons présenté et décrit les étapes principales du processus de RI, à savoir : l'indexation de la collection de documents et l'interrogation de l'information qui est basée sur les modèles d'appariement document-requête. Les principaux modèles de RI ont été présentés ainsi que les caractéristiques de chaque modèle. Nous avons présenté ensuite la notion de reformulation/expansion de requêtes, une stratégie permettant de pallier les problèmes rencontrés lors de la phase de l'appariement document-requête. Nous avons également présenté les fondements du processus d'évaluation à travers la présentation des campagnes d'évaluation de référence ainsi que les mesures d'évaluation. Le chapitre qui suit est consacré aux fondements et aux spécificités de la RI médicale.

Chapitre 2

Recherche d'information médicale : principes de base

Sommaire

1 Introduction . . . 43

2 Aperçu du domaine de la RI médicale . . . 44

2.1 Enjeux et défis du domaine . . . 44

2.2 Typologie de l'information médicale . . . 45

2.2.1 La littérature biomédicale . . . 46

2.2.2 Les dossiers médicaux des patients . . . 48

2.3 Quelques ressources termino-ontologiques du domaine biomédical . . . 48

2.3.1 Thésaurus MeSH . . . 49

2.3.2 Méta-thésaurus UMLS . . . 50

2.3.3 Nomenclature SNOMED . . . 51

2.3.4 Gene Ontology GO . . . 51

3 Étude des besoins en information médicale . . . 52

3.1 Catégories de requêtes médicales . . . 53

3.2 Caractéristiques des requêtes médicales . . . 56

3.2.1 Caractéristiques selon les utilisateurs . . . 57

3.2.2 Caractéristiques linguistiques . . . 59

3.2.3 Sujet de la requête . . . 61

3.3 Stratégies et comportements de recherche . . . 62

4 Techniques et modèles de RI médicale . . . 65

4.1 Expansion de requêtes . . . 65

4.2 Expansion de documents . . . 67

5 Modèles de RI basées sur le paradigme PICO . . . 68

5.1 EBM et paradigme PICO . . . 68

5.2 Identification des éléments PICO . . . 72

5.3 Modèles de RI basés sur les facettes PICO . . . 75

6 Campagnes d'évaluation pour la RI biomédicale . . . 78

6.1 Campagne d'évaluation TREC . . . 78

6.2 Campagne d'évaluation ImageCLEF . . . 81

7 Aperçu de systèmes de RI médicaux 83
7.1 PubMed 83
7.2 EMERSE 84
7.3 CisMef 85
7.4 Doctissimo 86
7.5 iMed 86
7.6 MedSearch 86
8 Conclusion 87

1 Introduction

Les progrès technologiques et le poids croissant des facteurs sociaux et économiques de production de l'information ont profondément transformé le fonctionnement et le comportement des utilisateurs dans le domaine de la RI médicale. En effet, la médecine est un champ d'information-intensive. Comme l'accès à l'information pertinente et en un temps limité est essentiel pour délivrer des services de santé efficaces, la médecine est ainsi dépendante de la technologie de l'information et plus spécifiquement des SRI médicaux. Une grande partie des données médicales disponibles aujourd'hui est sous forme non structurée, à savoir texte libre. La recherche et l'interprétation de ces données présente des défis spécifiques au domaine médical. Au cœur de ces défis est le problème du "fossé sémantique", défini comme la différence entre les informations brutes et la façon dont un être humain peut les interpréter.

Par ailleurs, le développement d'internet a transformé l'accès à l'information médicale. En effet, des études larges et exhaustives menées par *Pew Internet and American Life Project* révèlent que 80% des utilisateurs cherchent l'information de santé en ligne et que l'information retournée impacte leurs décisions quant à la prise en charge de leur propre santé ou celle de leurs proches (Fox et Jones, 2009). En outre, ces études montrent que la population d'utilisateurs est caractérisée par une variabilité significative à la fois sur le volet de l'âge que du niveau d'expertise ; les besoins en information liés à la santé sont également très variés couvrant des besoins liés à la médication, santé et bien-être, traitements et pathologies.

Malgré la diversité des SRI, qu'ils soient généraux ou dédiés tel que PubMed [1], ainsi que la multiplicité des sources d'information, les utilisateurs demeurent insatisfaits de la pertinence des résultats retournés par ces systèmes (Zhang, 2010). En effet, formuler des requêtes claires et qui reflètent le besoin en information spécifique est une tâche difficile pour les experts ainsi que les novices du domaine médical (Spink et Jansen, 2006; White et Moris, 2008). Plus spécifiquement, dans le cadre de la recherche médicale experte, médecins et cliniciens ont été éduqués à formuler des questions cliniques selon la structure PICO qui désigne *Population/-Problem*, *Intervention*, *Comparison* et *Outcome* (Boudin *et al.*, 2010b; Boudin *et al.*, 2010a). Cette structure a été introduite dans le cadre de la médecine basée sur les faits afin d'aider les praticiens à poser des questions bien structurées et à trouver les meilleures preuves leur permettant une prise de décision optimale. Néanmoins, malgré de nombreux travaux du domaine (Schardt *et al.*, 2007; Boudin *et al.*, 2010b; Chabou et Iglewski, 2015), accéder à l'information pertinente reste une tâche difficile dans le domaine médical.

Nous décrivons dans la section qui suit les fondements de la RI médicale, en passant en revue les enjeux du domaine, la typologie de l'information médicale ainsi que quelques ressources terminologiques. La section 3 présente les travaux sur les besoins en information médicale, répartis selon trois sous-sections, à savoir : (1) typologie des requêtes médicales, (2) caractéristiques de ces requêtes et les stratégies de recherche ; et (3) comportement de recherche d'information des utilisateurs. Nous détaillons plus précisément les besoins des experts médicaux. Différentes approches concernant l'expansion des requêtes et des documents sont présentées dans la section 4. Nous détaillons par la suite dans la section 5 les travaux portant sur la RI basée sur le paradigme PICO. Dans la section 6, nous passons en revue les principales campagnes d'évaluation du domaine médical. Enfin, la section 7 donne un aperçu sur les systèmes de RI médicaux.

1. http ://www.ncbi.nlm.nih.gov/pubmed

2 Aperçu du domaine de la RI médicale

2.1 Enjeux et défis du domaine

La RI médicale se réfère à des méthodologies et des technologies qui visent à améliorer l'accès aux archives d'informations médicales via un processus de RI. Ces informations sont potentiellement accessibles à partir de nombreuses sources, y compris le Web en général, les médias sociaux, des articles de journaux, et les dossiers hospitaliers. Le contenu médical, ainsi que l'information sur la santé est le sujet le plus abordé par les utilisateurs sur internet (Sadasivam *et al.*, 2013), ce qui rend ce domaine important pour les travaux de RI.

L'information médicale est d'intérêt à un large éventail d'utilisateurs, y compris les patients et leurs familles, les chercheurs, les médecins généralistes, les cliniciens et les praticiens ayant une expertise spécifique. De ce fait, il existe plusieurs services dédiés qui cherchent à rendre cette information plus facilement accessible, tels que les systèmes de recherche médicaux sur internet "*Health on the Net*", destinés au public général et aux praticiens [2].

Malgré la popularité du domaine médical pour les utilisateurs de moteurs de recherche, et de l'intérêt actuel qu'accorde la communauté en RI à ce sujet, le développement des technologies de recherche et d'accès demeure particulièrement difficile et sous exploré. Un des problèmes principaux dans le domaine de la RI médicale est la diversité des utilisateurs de ces services, correspondant à des différences dans les types et les champs de leurs besoins individuels. Autrement dit, la nature de la question médicale dépend des objectifs des utilisateurs, qui ont généralement des niveaux de connaissance médicale très variés. Ces compétences sont en étroite relation avec le comportement des utilisateurs ainsi que leur stratégie de recherche lors de la tâche de RI.
Ces défis peuvent être résumés comme suit :

1. **La diversité des besoins en information** (Clarke *et al.*, 2013; Dixon *et al.*, 2013) : Un patient avec une maladie récemment diagnostiquée bénéficie généralement d'une information introductrice à sa maladie et son traitement. Les patients vivants avec une maladie et qui essayent de gérer leur état de santé à long terme sont, quant à eux, à la recherche d'une information plus avancée, ou encore de groupes de discussion et de forums de santé.
   De la même façon, un médecin généraliste a besoin de l'information de base rapidement pendant une consultation pour conseiller le patient, mais d'une information plus détaillée lors de la prise de décision pendant la consultation. Cependant, un clinicien spécialiste pourrait chercher/être intéressé par une liste exhaustive de cas semblables ou de documents/articles scientifiques liés au cas des patients qui cherchent a être conseillés et guidés. Comprendre les différents types d'utilisateurs et leurs besoins en information est l'un des enjeux de la RI médicale. Ainsi, adapter le processus de RI pour mieux traiter ces besoins en vue de développer des systèmes efficaces, potentiellement personnalisés, est un des plus grands défis du domaine.
2. **Connaissances médicales différentes** (Zickuhr, 2006) : Les différentes catégories d'utilisateurs de SRI médicaux ont des niveaux très variés de connaissances médicales. De plus, la connaissance médicale des différents individus au sein de la catégorie d'utilisateurs peut varier considérablement. Cela affecte la façon avec laquelle les personnes

2. http ://www.hon.ch/

soumettent les requêtes au système et aussi le niveau de complexité de l'information qui doit être retournée, ou encore le type de support pour la compréhension et la désambiguisation qui s'avère nécessaire pour appréhender les résultats de recherche.

3. **Compétences linguistiques différentes** (Pang *et al.*, 2014) : Étant donné les différents profils d'utilisateurs, experts ou néophytes, la tâche de RI reste une tâche difficile. En effet, les experts médicaux comme les praticiens médecins et professionnels de santé se focalisent sur la structure PICO pour formuler leurs questions. De plus, ils se basent sur leurs connaissances du domaine en utilisant des ressources externes comme les ontologies du domaine. Cependant, les utilisateurs néophytes comme les patients et leurs familles, manifestent des difficultés dans la tâche de recherche dans les choix des termes.

Ainsi, plusieurs problèmes se posent, nous citons les plus abordés par la communauté de RI :

— **La quantité importante de littérature médicale** (Nourbakhsh *et al.*, 2012; Califf *et al.*, 2013). Prenant l'exemple de *Entrez*[3] qui publie des références de plus de 34000 journaux chaque année (plus d'un million d'articles de journaux au total). Cela rend la tâche de trouver l'information spécifique au besoin en information difficile.

— **Expression des besoins en information d'une manière inadéquate** (Soldaini *et al.*, 2015c; Hanbury, 2012). La plupart du temps, les médecins ignorent leur besoin en information, ce qui rend difficile d'exprimer leur besoin avec des requêtes appropriées. Par ailleurs, exprimer des requêtes claires et spécifiques, par les experts ou les non experts du domaine médical demeure une tâche compliquée.

— **La nature de la tâche de RI médicale avec la contrainte du temps "*Time-consuming*"** (Keenan *et al.*, 2013). En raison de la grande quantité de littérature médicale, le temps nécessaire pour trouver les documents pertinents dans la littérature liés aux patients s'est considérablement élargi. Dans la plupart des cas, le processus de collecte prend plus de temps que le médecin peut épargner. Selon (Merry et Korsmeyer, 1997), une tâche de recherche de plus de 30 secondes n'est pas acceptable pour les médecins.

2.2 Typologie de l'information médicale

Dans (Hersh, 2009), l'auteur a classé l'information médicale selon deux catégories, à savoir : (1) la littérature du domaine médical et (2) l'information qui concerne le patient.

La première catégorie représente la littérature primaire qui englobe les informations qui apparaissent dans les livres et les rapports. Ce sont les publications des travaux de recherche, produits par des chercheurs, comme les brevets, les rapports, les articles originaux de la presse spécialisée, conférences, livres, thèses et mémoires, nous citons :

— La littérature grise, qualifiée aussi de "non conventionnelle" ou de "souterraine", est constituée de la multitude de documents qui échappent aux circuits commerciaux traditionnels de l'édition (rapports de recherche, thèses, actes de congrès etc.). Ils ont la particularité d'être mal diffusés et mal répertoriés dans les bases bibliographiques.

— Les documents *EBM (Evidence Based Medicine)* constitués des consensus et recommandations faisant partie de cette catégorie. Il s'agit de documents émanant d'instances officielles comme l'ANAES ou l'AFFSAPS ou semi officielles comme les sociétés savantes.

3. http ://www.ncbi.nlm.nih.gov/gquery/

En plus de l'information primaire, la première catégorie englobe l'information secondaire. Cette dernière est la synthèse de la littérature primaire comme les articles de revue dans les journaux et les livres. Elle comprend les données bibliographiques et les commentaires des documents primaires. L'information secondaire concerne les informations spécifiques aux patients visant à informer ceux qui cherchent une information précise sur la maladie du patient, les praticiens et peuvent être présentes sous forme de comptes-rendus médicaux.

Nous détaillons dans ce qui suit les deux catégories principales d'information biomédicale susmentionnées.

2.2.1 La littérature biomédicale

La littérature biomédicale comprend l'information disponible dans les bases de données bibliographiques, faisant référence aux revues scientifiques et aux comptes rendus des conférences du milieu médical. MEDLINE est la base de données de référence dans le domaine. Créée et gérée par la National Library of Medicine [4] (NLM), MEDLINE contient des citations de journaux et des résumés pour la littérature médicale regroupant des informations et des références de millions d'articles scientifiques indexés en langue anglaise. Ces informations couvrent toute l'information du domaine de la médecine, de la médecine dentaire, des soins infirmiers, la médecine vétérinaire, des para-médicaux, etc. PubMed est un portail dédié à fournir un accès gratuit à l'information dans MEDLINE ainsi qu'aux liens des résumés d'articles ou d'articles en texte intégral. Couvrant plus de 24 millions de citations de la littérature biomédicale de MEDLINE, les citations et résumés de PubMed comprennent les champs du domaine biomédical et de santé, la science de la vie, la science du comportement, les sciences de la chimie, et la bio-ingénierie. De plus, PubMed fournit l'accès à des sites web supplémentaires pertinents et des liens pour d'autres ressources de la biologie moléculaire du National Center for Biotechnology Information (NCBI [5]).

La base MEDLINE est constituée d'un ensemble de documents, qui sont produits par un ensemble de spécialistes de l'information de la NLM, à savoir : (1) les bibliothécaires qui obtiennent les revues et vérifient les problèmes récurrents des journaux ; (2) les indexeurs humains, qui sont des spécialistes du domaine médical, analysent les contenus des articles et décrivent les concepts présentés en utilisant le vocabulaire contrôlé MeSH [6]. Plus précisément, l'indexation est soit manuelle ou semi-automatique en utilisant les termes MeSH ; et (3) les spécialistes de l'informatique et de l'information, qui développent et entretiennent le système ou le moteur de recherche.

Pour la langue française, un catalogue des sites médicaux en langue française a été crée : CiSMeF [7]. Ce portail fournit une sélection de sites, d'articles et de documents en libre accès. D'une manière générale, une classification de la littérature médicale est donnée par le Tableau 2.1.

Par ailleurs, une classification de l'information textuelle médicale est donnée dans le Tableau 2.2.

4. http ://www.nlm.nih.gov
5. http ://www.ncbi.nlm.nih.gov/
6. http ://www.nlm.nih.gov/mesh/MBrowser.html
7. http ://www.chu-rouen.fr/cismef/

Littérature Primaire	**Littérature Secondaire**	**Littérature Ternaire**
Recherche originale Résultats des journaux Dissertations Conférences Actes de conférence Correspondance	Services d'indexation et de résumés Articles de revues Revues systématiques Méta-analyses Guides de pratique	Manuels Encyclopédies Journaux Guides
Sources : NEJM, JAMA	Sources : PubMed, CINAHL, Cochrane Library, Web of Science	Sources : Goodman and Gilman's, Williams Obstetrics

Tableau 2.1 – Classification de la littérature médicale

1. Information spécifique au patient
   - **Structurée** : résultats du laboratoire, signes vitaux
   - **Narratif** : histoire et physique, notes de progrès, rapport de radiologie
2. Information basée sur les connaissances
   - **Primaire** : recherche originale
   - **Secondaire** : résumés de recherche (comptes rendus des articles, livres, directives cliniques)

Tableau 2.2 – Classification de l'information textuelle médicale

D'une manière plus spécifique, les revues internationales et les publications scientifiques dans le domaine de la recherche biologique, médicale et de la santé occupent une place prépondérante dans la littérature scientifique en langue française depuis 30 ans. Plus précisément, les revues systématiques permettent d'aider les praticiens à être informés de la littérature médicale, en offrant un résumé des grandes parties de preuves médicales et en aidant à expliquer les différences entre les études qui traitent les même questions (Cook *et al.*, 1997). Une méta-analyse est un type de revue systématique qui utilise des méthodes statistiques pour combiner et résumer les résultats de plusieurs études primaires (Lundahl *et al.*, 2013).

Les articles de revues traditionnelles et non structurées sont utiles pour obtenir une vue d'ensemble d'un état clinique, mais ne peuvent pas fournir une réponse fiable et non biaisée à une question clinique ciblée. Une revue systématique est un résumé de la recherche portant sur une question clinique ciblée d'une manière reproductible structurée. Elle est souvent accompagnée d'une méta-analyse, qui est une mise en commun des statistiques ou de l'agrégation des résultats des différentes études qui fournissent une estimation unique de l'effet. Le Tableau 2.3 résume le processus typique d'une revue systématique et méta-analyse incluant la garantie contre des résultats erronés (Murad *et al.*, 2014).

1. Formuler une question
2. Définir les critères d'éligibilité des études qui peuvent être inclus, comme Patient, Intervention, Comparaison, Outcome (PICO), et la conception de l'étude.
3. Développer les hypothèses pour expliquer les hétérogénéités
4. Conduire les recherches
5. Filtrage des titres et résumés pour leur incorporation
6. Voir les textes intégraux des études éligibles
7. Évaluer les risques de biais
8. Résumer les données
9. Après l'analyse des données :
   — Générer des sommaires d'estimation et intervalles de confiance
   — Chercher des explications sur l'hétérogénéité
   — Taux de confiance dans les estimations de l'effet

Tableau 2.3 – Le processus de réalisation d'un examen systématique et du méta-analyse

2.2.2 Les dossiers médicaux des patients

Le dossier médical d'un patient ou personnel (DMP) est un dossier médical informatisé. Il permet aux professionnels de santé qui prennent en charge le patient de partager les informations de santé. Il rassemble les informations médicales, strictement nécessaires à la coordination des soins : prescriptions, synthèses médicales, comptes rendus d'hospitalisation, résultats d'analyses, antécédents médicaux et chirurgicaux, etc. Il est organisé en huit espaces : synthèse médicale, traitements et soins, comptes rendus (hospitalisations, consultations), imagerie médicale (radios, IRM, etc.), analyses de laboratoire (sang, urine, etc.), prévention (rappels de vaccination) et certificats et déclarations. Le DMP peut être créé lors d'une consultation médicale ou lors d'une admission dans une structure de soins. Il constitue le noyau fondamental de la qualité des soins dans les hôpitaux et a deux objectifs principaux, à savoir :

— Fournir au médecin traitant l'information la plus complète pour qu'il puisse proposer le traitement ou les examens les plus adaptés et également d'éviter des redondances inutiles d'examens ou de prescriptions.

— Accélérer les capacités de l'ensemble des acteurs à produire et à partager des données de santé de manière sécurisée, dans le but de mieux coordonner les soins.

2.3 Quelques ressources termino-ontologiques du domaine biomédical

Les professionnels de santé ont recours à une terminologie spécifique au domaine de la Santé pour exprimer des besoins divers. De nombreuses terminologies médicales existent et chacune d'entre elles a été créée pour répondre à un besoin donné. Les besoins sont globalement de quatre natures : (1) coder l'information, en particulier l'information de soin de patient ou la Santé Publique ; (2) indexer des documents, en particulier la littérature biomédicale ; (3) représenter les entités dans les systèmes experts et les systèmes d'aide à la décision en général ; et (4) servir de support "en interface" pour le recueil d'informations.

Nous présentons dans ce qui suit les terminologies les plus utilisées dans le domaine médical : MeSH et UMLS, qui sont des ontologies médicales, la nomenclature SNOMED et Gene Ontology (GO).

2.3.1 Thésaurus MeSH

Créée par la NLM [8] en 1954, le thésaurus MeSH (Medical Subject Heading) permet d'indexer, classer et rechercher des documents (notamment ceux de MEDLINE). MeSH comprend essentiellement des termes qui désignent les concepts biomédicaux, des descripteurs, des relations et des qualificatifs. Concrètement, MeSH a un moteur de recherche en ligne qui évolue chaque année. Il est régulièrement mis à jour, et la traduction vers le Français a été assurée par l'INSERM [9]. La structure de MeSH est hiérarchique avec une arborescence composée de :

— **Terme** : Un terme est un mot ou un ensemble de mots ayant un sens particulier.

— **Concept** : Un concept est constitué de un ou plusieurs termes synonymes et il est désigné par le nom de l'un de ces termes appelé terme préféré (*preferred terms*).

— **Relation** : Elle représente les relations qui existent entre les concepts dans la hiérarchie MeSH. Il existe deux types de relations : les relations hiérarchiques et les relations associatives (associé à). La hiérarchie dans MeSH est représentée par un code reflétant l'arborescence à laquelle le concept appartient et peut véhiculer plusieurs sens, tels que à titre d'exemple :

1. relation "est une partie de" (méronymie), par exemple le concept "nez" (A01.456.505.733) est une partie de "visage" (A01.456.505).
2. relation "est sémantiquement proche de" (aboutness), par exemple le concept "sécurité" (G03.850.110.060.075) est sémantiquement proche de "accidents" (G03.850.110).
3. relation "est un type de" (hyponymie), par exemple le concept "prémolaire" (A14.549.167.860.200) est un type de "dent" (A14.549.167.860).

— **Descripteur** : connu sous le nom de *Main Headings* (MH), un descripteur est constitué d'un ou plusieurs concepts ayant des significations proches et porte le nom d'un de ses concepts dit préféré. Les autres concepts présentent des relations sémantiques avec le concept préféré, soit hiérarchiques soit associatives. Les descripteurs MeSH sont répartis en 16 catégories recouvrant différents domaines de la médecine.

Chaque catégorie est structurée en arborescence hiérarchique pouvant comprendre jusqu'à 11 niveaux de hiérarchie. Chaque descripteur est représenté par un code alphanumérique, la lettre indiquant la catégorie et la séquence numérique précisant la localisation dans la hiérarchie (*Cf.* Tableau 2.4). Un descripteur peut avoir plusieurs localisations, au sein d'une même catégorie ou de catégories différentes, et plusieurs codes alphanumériques représentant chacun une localisation. Par exemple, le descripteur "*Pain*" appartient à plusieurs hiérarchies, C10.597.617, C23.888.592.612, C23.888.646, F02.830.816.444, G11.561.600.810.444.

— **Qualificatif** : Les qualificatifs servent à décrire un aspect particulier sur les entrées MeSH, qui sont utilisés seuls ou associés à un descripteur. Notons que le sens du descripteur est plus précis quand il est accompagné du qualificatif.

8. *National Library of Medicine*
9. Institut National de la Santé et de la Recherche Médicale- http ://inserm.com/

[A]	Anatomie
[B]	Organismes
[C]	Maladies
[D]	Produits chimiques et pharmaceutiques
[E]	Techniques et équipements analytiques, diagnostics et thérapeutiques
[F]	Psychiatrie et psychologie
[G]	Phénomènes et processus
[H]	Disciplines et professions
[I]	Anthropologie, enseignement, sociologie et phénomènes sociaux
[J]	Technologie, industrie et agriculture
[K]	Sciences humaines
[L]	Sciences de l'information
[M]	Groupes individualisés
[N]	Soins de santé
[V]	Caractéristiques d'une publication
[Z]	Lieux géographiques

Tableau 2.4 – Les différentes catégories ou domaines du MeSH

2.3.2 Méta-thésaurus UMLS

UMLS (Unified Medical Language System) est un ensemble de fichiers et de logiciels qui réunit de nombreux vocabulaires biomédicaux et de santé, ainsi que des standards et des normes pour permettre l'interopérabilité entre les systèmes informatiques. L'UMLS permet d'améliorer ou développer des applications, telles que les dossiers de santé électroniques, les outils de classification, les dictionnaires et traducteurs automatiques.

Dans la pratique, un des points forts de UMLS est de relier l'information de santé avec les termes médicaux, les noms des médicaments, et les codes de facturation à travers les différents systèmes informatiques. Par exemple lier les termes et codes entre le médecin, le pharmacien et la compagnie d'assurance du patient. Il offre plusieurs fonctionnalités supplémentaires y compris un moteur de recherche, l'exploration de données, les rapports sur les statistiques de la santé publique et la recherche terminologique.
Il dispose de trois outils principaux, à savoir :

— *Méta-thésaurus* : il constitue la base unifiée des concepts médicaux. Il comprend des synonymes, des variations lexicales et des concepts associés, afin de dresser la liste de tout le vocabulaire des expressions médicales disponibles. Chaque concept, dans le Méta-thésaurus, permet de regrouper différentes expressions des différentes sources qui expriment le même concept.

— *Le réseau sémantique* : il spécifie les types sémantiques qui permettent de catégoriser tous les concepts médicaux définis dans le Métathésaurus. Il contient également les relations sémantiques existantes entre chaque type sémantique. Il existe 135 types sémantiques dans l'UMLS avec 54 relations sémantiques.

— *Le SPECIALIST Lexicon* : offre des outils nécessaires (informations syntaxiques, morphologiques et orthographiques) de traitement automatique du langage naturel.

2.3.3 Nomenclature SNOMED

La SNOMED [10] est une nomenclature pluri-axiale couvrant tous les champs de la médecine et de la dentisterie humaines, ainsi que la médecine animale (*Cf.* Tableau 2.5). Il s'agit d'un système de classification permettant de normaliser l'ensemble des termes médicaux utilisés par les praticiens de santé. La SNOMED a pour fonction d'attribuer un code à chaque concept permettant un grand nombre de combinaisons entre eux. Elle comprend également une liste des diagnostics interfacée avec la CIM 10 [11]. La SNOMED permet ainsi de stocker des informations médicales individuelles dans des entrepôts de données afin d'établir des outils d'analyse décisionnelle, de faciliter des décisions thérapeutiques, de contribuer aux études épidémiologiques et à l'enseignement. L'utilisation de SNOMED garantit l'universalité du vocabulaire médical.
La première traduction française de la SNOMED a été réalisée en 1998. Des mises à jour de cette version SNOMED 3.5 VF ont été effectuées depuis 1998 indépendamment des évolutions et mises à jour de la SNOMED International, actuellement utilisée dans sa version SNOMED CT [12]. Il existe donc des différences entre la SNOMED 3.5 VF et la SNOMED International CT.

Axe	Nom de l'axe
A	Agents physiques
C	Produits chimiques
D	Diagnostics
F	Fonctions
G	Qualificatifs
J	Métiers
L	Organismes vivants
M	Morphologie
P	Procédure
S	Contexte social
T	Topographie

Tableau 2.5 – Les onze axes de la nomenclature SNOMED

Chaque axe représente une hiérarchie de concepts qui peuvent représenter une combinaison de concepts. Ces concepts sont représentés par une série de termes au sein de laquelle on peut distinguer une formulation préférée et des synonymes de diverses natures syntaxiques.

2.3.4 Gene Ontology GO

Le projet *Gene Ontology* a été mis en place pour fournir un langage commun pour décrire les aspects d'une biologie du produit du gène (Ashburner *et al.*, 2000). L'utilisation d'un vocabulaire cohérent permet de comparer les gènes provenant d'espèces différentes en fonction de leurs annotations GO. L'objectif de GO est de fournir des vocabulaires contrôlés pour la description du processus biologique, fonction moléculaire, et composant cellulaire de produits de gènes. Ces termes sont utilisés en tant qu'attributs de produits génétiques de banques de

10. Systematized Nomenclature of Medicine
11. Classification Internationale des Maladies
12. SNOMED Clinical Terms

données de l'organisme, ce qui facilite les requêtes uniformes à travers eux. La Figure 2.1 présente un exemple de résultats de recherche avec le terme cancer utilisant le moteur *Gene Ontology Consortium*.

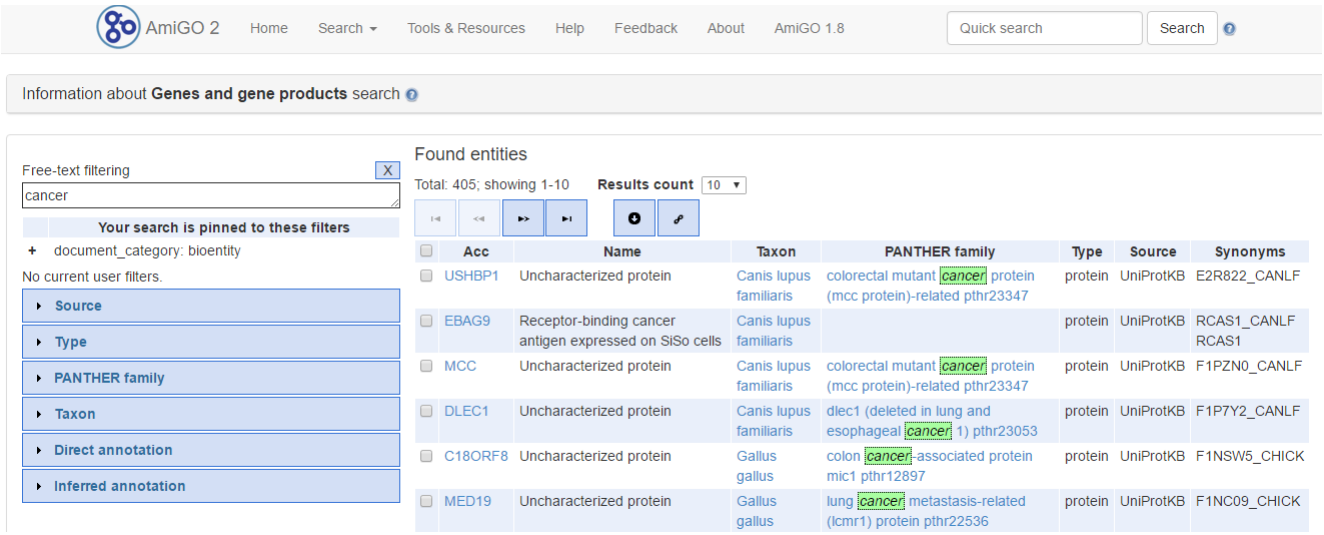


FIGURE 2.1 – Exemple de résultats utilisant GO.

Les vocabulaires contrôlés de termes sont structurés de manière à permettre l'annotation de produits de gènes à GO à différents niveaux de détail et d'interroger des produits de gènes qui sont impliqués dans des processus similaires.

Dans ce qui suit, nous allons donner un aperçu des travaux sur le besoin en information dans le domaine médical selon deux volets. Le premier volet concerne les travaux sur le besoin en information d'une manière générale. Plus spécifiquement, nous introduisons les typologies de requêtes médicales selon trois axes, à savoir : la nature du besoin en information, la nature des questions posées et la nature des utilisateurs. Le deuxième volet détaille d'une manière plus spécifique le besoin en information des experts. Autrement dit, les requêtes expertes dans le cadre de la RI médicale du point de vue de leur formulation, caractéristiques et utilisation.

3 Étude des besoins en information médicale

La RI médicale se caractérise par une diversité des différentes méthodes et techniques utilisées afin d'aboutir à des résultats pertinents. Concrètement, cette diversité se manifeste via l'analyse des besoins en information médicale sur plusieurs volets, tenant compte de la diversité des profils des personnes qui cherchent l'information par exemple, experts et novices, de l'utilisation des SRI médicaux vs. l'utilisation de moteurs de recherche du web ainsi que les différentes stratégies utilisées pour aboutir à des résultats pertinents.
Il est communément admis dans le domaine de la RI biomédicale, que l'expression des requêtes traduisant fidèlement les besoins en information est une tâche difficile aussi bien pour des novices que pour des experts (Spink *et al.*, 2004; White et Moris, 2008). En conséquent, l'analyse et la représentation des besoins en information "cachés" derrière les requêtes, sont devenues un réel défi dans le domaine.

Nous présentons dans ce qui suit une synthèse des travaux de l'état de l'art sur l'analyse du besoin en information médicale et biomédicale selon trois volets : la catégorisation des requêtes médicales en exposant les catégories des requêtes et les types de besoins en information, les caractéristiques des requêtes médicales en détaillant les caractéristiques linguistiques ; et les stratégies et comportements de recherche.

3.1 Catégories de requêtes médicales

La tâche de recherche d'information est motivée par les besoins en information, qui est la perception de l'insuffisance des connaissances (Griffin *et al.*, 2004). Les besoins en information médicaux incluent les *besoins cognitifs* qui comprennent l'information sur la maladie, le traitement et les *besoins affectifs* qui comprennent des informations pour faire face à la maladie émotionnellement (Wilson, 1999). L'information médicale peut être utilisée soit par des personnes malades (patients) dans le but de comprendre le diagnostic et les modalités de traitement ou par des gens en bonne santé (familles du patient ou autres) pour évaluer les risques et la prévention de la maladie (Brashers *et al.*, 2000).
La classification de questions a pour objectif d'assigner pour une question donnée, une ou plusieurs catégories. Plusieurs classifications des questions médicales ont été proposées dans la littérature. En effet, dans le domaine médical, les médecins posent des questions en se basant sur leurs connaissances spécifiques au domaine (Graber *et al.*, 1999). Selon (Ely *et al.*, 2000) les questions cliniques sur les soins primaires peuvent être classées en un nombre limité de types génériques.

Nous allons dans ce qui suit caractériser les requêtes selon le but de la recherche.
Les utilisateurs de l'information basée sur les connaissances ont des différences basées sur la nature de leur besoin en information et les ressources disponibles. Le besoin en information et la recherche d'information dépendent de l'utilisateur. Pour les médecins, dans (Gorman et Helfand, 1995; Dawes et Sampson, 2003), les auteurs ont défini quatre types de besoin en information dans le contexte clinique :

— *besoin méconnu* : clinicien ignorant le besoin en information ;
— *besoin reconnu* : clinicien conscient du besoin mais il peut ne pas le poursuivre ;
— *besoin poursuivi* : la recherche d'information se produit mais peut être sans succès ou peut ne pas aboutir ;
— *besoin satisfait* : la recherche d'information a abouti avec succès.

Plusieurs autres études avaient pour objectif d'identifier la nature du besoin en information observé (Timpka et Arborelius, 1990; Ely *et al.*, 1999). Tous ont montré que les questions médicales sont spécifiques au problème du patient.
Plus récement, (Andualem *et al.*, 2013) ont analysé le besoin en information des professionnels de santé dans les hôpitaux d'Éthiopie. Les premiers ont trouvé que le besoin en information le plus couramment perçu concerne trois sujets, à savoir : (1) soins cliniques, (2) se tenir à jour et (3) fournir des informations pour le patient. Les ressources les plus utilisées sont les collections personnelles des médecins suivies des ressources électroniques.

Les auteurs de (Pratt et Wasserman, 2000) ont proposé une catégorisation des requêtes médicales, présentée dans le Tableau 2.6.
Les obstacles pour la formulation de la bonne question clinique sont nombreux (Ely *et al.*, 2002). Ces derniers ont identifié 59 obstacles, parmi lesquels les suivants : reconnaître

Catégorie	Définition
Prévention	appartenant à une prévention d'un problème médical, maladie, syndrome ou blessure
Risque ou étiologie	facteurs de risque, corrélations ou causes de maladies
Tests de diagnostics	test de diagnostics ou bilans d'un problème médical spécifique
Diagnostics	concernant un diagnostic ou des symptômes spécifiques
Symptômes	symptômes d'une maladie ou un syndrome spécifiques
Traitement	traitement ou thérapies pour un problème spécifique
Effets	concernant des effets positifs ou négatifs, ou des complications d'un traitement ou une thérapie spécifique
Pronostic	psychophysiologie ou progression naturelle d'une maladie sur la durée, incluant des questions pronostiques comme le taux de mortalité
Synthèse	demande de synthèse sur quelques sujets médicaux
Autre	autre catégorie

Tableau 2.6 – Catégories de requêtes médicales selon le modèle de (Pratt et Wasserman, 2000)

une lacune dans les connaissances, formuler une question, rechercher des informations pertinentes, formuler une réponse, et utiliser la réponse aux soins directs aux patients. Six obstacles ont été considérés comme particulièrement marquants par les médecins : le temps excessif nécessaire pour trouver l'information ; la difficulté de modifier la question initiale, qui était souvent vague ; la difficulté à choisir une stratégie optimale pour rechercher l'information ; ressources non appropriées pour couvrir le sujet ; l'incertitude sur la façon de savoir quand toute la preuve pertinente a été trouvée afin d'arrêter la recherche ; et la synthèse inadéquate d'un grand volume de données dans une déclaration cliniquement utile.

Par ailleurs, dans (Ely *et al.*, 2002), les auteurs ont développé une taxonomie de questions génériques, trouvant 69 types différents, les 1 à 10 meilleurs de chacun d'eux est représenté dans le Tableau 2.7.

Dans (White et Horvitz, 2012), les auteurs ont récupéré des fichiers de log du web, et ont analysé le comportement de recherche des utilisateurs via une analyse de la composition des requêtes. La Figure 2.2 montre la répartition de requêtes médicales selon le type de la préoccupation des utilisateurs.

Selon (Marchionini, 1995), la plupart des activités intellectuelles, comme le processus de RI, implique une planification (sélection de termes de la requête), un suivi de l'évolution (nombre de documents retournés), une prise de décision (continuer ou arrêter la recherche) et une réflexion sur l'historique de recherche (modification de requête pour de meilleurs résultats). Selon lui, la personne avec la plus haute aptitude cognitive devrait mieux faire dans la tâche

Question générique	Combien posées (%)	Combien posées et poursuivies (%)	Combien poursuivies et satisfaites (%)
Quelle est la cause du symptôme X ?	9	9	50
Quelle est la dose du médicament X ?	8	85	97
Comment gérer la maladie ou les résultats ?	7	29	83
Comment je m'y prends avec l'observation ou la maladie ?	7	33	72
Quelle est la cause de ce signe physique ?	7	18	46
Quelle est la cause du contrôle de l'observation X ?	4	40	72
Est ce que le patient a cette maladie ou condition X ?	4	14	67
Est ce que le test X est indiqué dans la situation Y ?	4	29	83
Quel médicament X choisir pour l'état Y ?	3	47	76
Est ce que le médicament X est indiqué dans la situation Y ?	3	25	78

Tableau 2.7 – Questions les plus posées, poursuivies et satisfaites selon (Ely *et al.*, 2002).

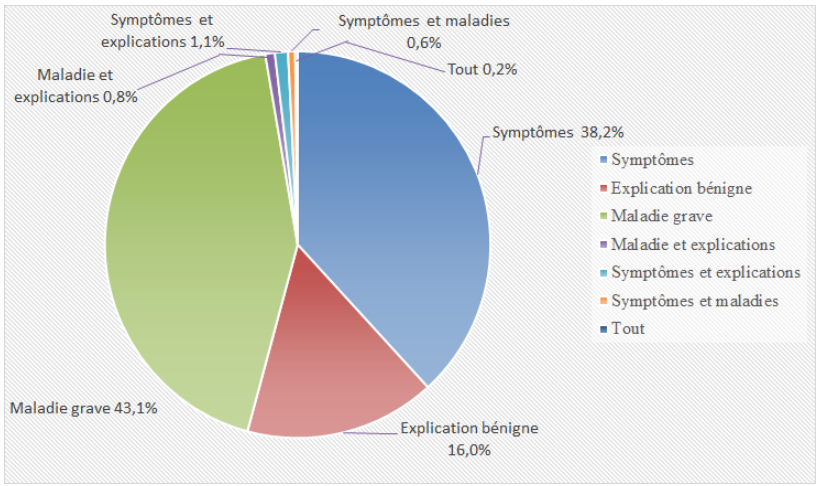


FIGURE 2.2 – Distribution de requêtes médicales selon le type (White et Horvitz, 2012).

de RI qu'une personne avec moins d'aptitudes. Dans (Hersh *et al.*, 2002), les auteurs ont évalué trois facteurs cognitifs, à savoir : la visualisation spatiale, le raisonnement logique et le raisonnement verbal, qui affectent la performance de la RI. Selon la revue de la littérature, rechercher l'information de santé s'est avéré lié aux caractéristiques individuelles de l'utilisateur ainsi qu'aux facteurs situationnels (Lambert et Loiselle, 2007). Dans (Wilson, 2006), l'auteur a donné une définition de la RI comme étant une activité centrée sur la personne, et elle est conduite par le rôle dans le but de réduire les incertitudes dans la prise de décision médicale. Ainsi, pour répondre aux besoins en information, les professionnels de santé peuvent sélectionner l'information à partir des différentes sources d'information soit dans leur lieu de travail (libraires des hopitaux) ou à partir des sources externes (ressources médicales du web) (Clarke *et al.*, 2013).
De plus, du point de vue de l'utilisateur, le besoin en information dépend de l'objectif spécifique de sa recherche. Le Tableau 2.8 donne une classification du type de l'information selon les consommateurs de l'information médicale en ligne.

Une autre catégorisation des requêtes des utilisateurs de PubMed est présentée comme suit : ***Requêtes informationnelles*** conçues pour accomplir le besoin en information (exemple

Type d'information	Définition
Information pour connaitre	L'information sur les maladies ; les traitements, le système de soins, partage d'expérience : pour mieux comprendre la maladie, se prendre en charge, décider, bénéficier d'un accompagnement psychologique, suivre des actions de prévention, le *self care*
Information pour choisir	L'information qui aide les patients à choisir un médecin, une assurance, etc.
Information pour superviser	L'information portant sur l'action des producteurs et des assurances : elle est mobilisée par les patients individuellement, des associations, des représentants de patients
Information pour produire	Données médicales personnelles : énoncé des symptômes, résultats d'examens, carnet de santé, etc.
Information pour se coordonner	L'information de coordination administrative et financière : feuilles de soins, prise de rendez-vous, etc.

Tableau 2.8 – Types d'information médicale selon l'utilisateur

"diabetes mellitus"[MeSH]), les **requêtes de navigation** conçues pour trouver et récupérer les documents pertinents (Yoo et Mosa, 2015), les **requêtes mixtes** qui ont pour but de rechercher un sujet spécifique dans un journal spécifique.
Dans ce contexte, (Broder, 2002) a donné trois classes de requêtes selon le contexte de l'intention derrière le besoin en information exprimé : *navigationnelle* avec l'intention d'atteindre un site particulier, *informationnel* avec l'intention d'acquérir des informations supposées être présentes dans une ou plusieurs pages web, *transactionnel* ou l'intention est de réaliser une activité intermédiaire sur le web. Dans (Nabarette, 2002), les auteurs ont donné une classification des facteurs explicatifs des différents types d'information des patients (Cf. Tableau 2.8).

3.2 Caractéristiques des requêtes médicales

Plusieurs travaux ont étudié les particularités des besoins en information dans le domaine médical afin de comprendre les différents aspects qui caractérisent les requêtes de la RI biomédicale comme l'analyse du besoin en information (Spink *et al.*, 2004; Zhang et Fu, 2011; Lykke *et al.*, 2012), l'analyse de la difficulté de la requête (Hersh *et al.*, 2002; Boudin *et al.*, 2012; Lykke *et al.*, 2012), l'impact du contexte de la recherche (Freund *et al.*, 2005; White et Moris, 2008; Cartright *et al.*, 2011; Lykke *et al.*, 2012), la précision de la recherche ainsi que la qualité de l'information médicale (Pandolfini et Bonati, 2002; Moturu *et al.*, 2008).

En général, ces études s'appuient sur des méthodes et techniques empiriques conduites sur un échantillon d'utilisateurs afin d'examiner les particularités et les caractéristiques du besoin en information des utilisateurs (Spink *et al.*, 2004; Zhang et Fu, 2011; Lykke *et al.*, 2012), la difficulté de la requête (Hersh *et al.*, 2002; Lykke *et al.*, 2012; Boudin *et al.*, 2012), le comportement de l'utilisateur (Dogan *et al.*, 2009; Ely *et al.*, 1999), l'effet du contexte sur la recherche (Freund *et al.*, 2005; White *et al.*, 2008; Cartright *et al.*, 2011; Lykke *et al.*, 2012) , et l'efficacité de la recherche ; la qualité et la fiabilité de l'information médicale (Pandolfini et Bonati, 2002; Moturu *et al.*, 2008). Les conclusions relatives à ces travaux donnent un aperçu sur l'activité de recherche d'information médicale et l'implication pour améliorer les systèmes de RI médicaux.

Les critères caractérisant les requêtes d'une manière générale du web, ou d'une manière spécifique concernant le domaine médical, ont été définis et mesurés dans la littérature et sont classés selon :

— le but de la recherche

— la structure et le vocabulaire de la requête

— la longueur de la requête

— la difficulté de la requête

De nombreux travaux (Cartright *et al.*, 2011) ont montré l'impact d'attributs comme la présence de catégories médicales, d'acronymes, spécificité et longueur des requêtes sur les résultats des performances. Par ailleurs, d'autres travaux ont montré l'impact du niveau d'expertise des utilisateurs sur la formulation des requêtes ainsi que sur les résultats de recherche (White et Moris, 2008; Lykke *et al.*, 2012). Ces travaux ont également montré l'importance de l'environnement professionnel pour l'interprétation des besoins en information.

Nous présentons dans ce qui suit les caractéristiques des requêtes médicales selon trois volets, à savoir : les caractéristiques liées à l'utilisateur, les caractéristiques linguistiques et les caractéristiques selon le sujet.

3.2.1 Caractéristiques selon les utilisateurs

Plusieurs études ont montré que les méthodes utilisées pour accomplir la tâche de RI sont différentes entre les experts du domaine et les utilisateurs novices (Downing *et al.*, 2005; Vibert *et al.*, 2007). De plus, la performance globale des experts du domaine est meilleure que celle des novices pour différents SRI comme le web et la recherche hypertexte (Marchionini *et al.*, 1993; Patel, 1994; Hölscher et Strube, 2000), ainsi que la recherche en ligne dans les bases bibliographiques (Hsieh-Yee, 1993; Wildemuth *et al.*, 1995).

Dans (White *et al.*, 2009), les auteurs ont montré que les experts au sein de leur domaine d'expertise effectuent la tâche de la recherche autrement que les non experts, du point de vue des sites visités, du vocabulaire utilisé dans la requête, de leur comportement et de la réussite de la recherche. Par ailleurs, dans (Liu *et al.*, 2012), les auteurs ont étudié l'impact de la difficulté de la tâche sur le comportement de recherche. Ils ont conclu que les utilisateurs passent moins de temps à voir le contenu des pages lorsque la tâche est difficile que quand il s'agit de tâches faciles. Par ailleurs, leur étude a montré une forte interaction entre la difficulté de la tâche et les connaissances du domaine.

En effet, les professionnels de santé considèrent PubMed comme l'une des sources les plus importantes et fiables pour la recherche de preuves médicales pertinentes et elle est mise à

jour régulièrement. Les auteurs dans (Yoo et Mosa, 2015) ont analysé des fichiers de log des utilisateurs PubMed, et ont comparé les utilisateurs expérimentés et non expérimentés de PubMed en terme de comportement de recherche ainsi que le nombre de requêtes par session de recherche. Ainsi, l'analyse des requêtes des utilisateurs de PubMed a permis de les classer en deux catégories :

— *Les utilisateurs PubMed expérimentés* : ils utilisent les fonctionnalités de recherche d'information du système comme les termes MeSH et les différents tags de recherche pour formuler leur requêtes, permettant d'avoir des documents pertinents d'une façon rapide.

— *Les utilisateurs PubMed non expérimentés* : ils se caractérisent par des sessions de recherche plus longues que celles des utilisateurs expérimentés, qui identifient leurs besoins en information à travers les requêtes ultérieurs.

Cette analyse confirme que les utilisateurs PubMed expérimentés obtiennent rapidement des documents pertinents en terme de longueur de session tandis que les non expérimentés se caractérisent par des sessions de recherche plus longues.

Selon *Pew Survey 2013* (Fox et Duggan, 2013), un adulte Américain sur trois fait des recherches en ligne sur une condition médicale spécifiques. En outre, ces études montrent que la population d'utilisateurs est caractérisée par une variabilité significative à la fois sur le volet de l'âge que des niveaux d'expertise (Zickuhr, 2006) ; les besoins en information liés à la santé sont également très variés couvrant des besoins liés à la médication, la santé et le bien-être, les traitements et pathologies.

Cependant, force est de constater que malgré la diversité des Systèmes de RI (SRI), qu'ils soient généraux ou dédiés tels que PubMed, ainsi que la multiplicité des sources d'information, les utilisateurs demeurent insatisfaits de la pertinence des résultats retournés par ces systèmes (Zhang, 2010). L'une des causes fondamentales à l'origine de ce constat est la difficulté de formulation des requêtes de la part des utilisateurs, d'une part, et l'interprétation des besoins en informations induits du côté des SRI, d'autre part. Ceci a encouragé une branche de recherches liées à l'analyse des besoins en information médicale. Plus précisément, les investigations de recherche ont porté sur deux volets, le premier est lié à l'analyse du comportement des utilisateurs de SRI médicaux et le deuxième à l'analyse des requêtes exprimées par leurs utilisateurs.

Dans (Bhavnani, 2001; Bhavnani, 2002), les auteurs ont observé des différences significatives dans la façon avec laquelle les utilisateurs explorent la recherche du point de vue des ressources clés utilisées, où les experts se basent sur leur expertise du domaine tandis que les novices se référent aux ressources générales du web. Les résultats dans (Wildemuth, 2004) suggèrent que le comportement de recherche change selon le gain de l'expertise du domaine médical.

Les études sur le domaine d'expertise ont montré plusieurs différences entre les experts et non experts, du point de vue de la sélection de sites (Bhavnani, 2002), du temps pour accomplir la tâche de recherche (Bhavnani, 2001), du vocabulaire et les expressions utilisés (Allen, 1991), du nombre et de la longueur de requêtes ainsi que la performance de recherche (Zhang *et al.*, 2005).

Les experts d'un domaine accomplissent la tâche de recherche différemment des gens avec peu de connaissances du domaine. Les recherches antérieures ont montré que les experts utilisent des stratégies différentes et plus performantes pour trouver ce qu'il veulent par rapport aux non experts. Ceux qui cherchent l'information sur le web sont différents les uns des autres sur plusieurs plans qui influencent considérablement les résultats de la recherche (HOBU *et al.*,

1987). Une différence peut être au niveau de leurs connaissances acquises sur un sujet ou un domaine précis. Dans (White et Morris, 2007), les auteurs différencient domaine d'expertise de la recherche experte, du fait que cette dernière concerne plus le sujet ou le thème exprimé par le besoin en information que les connaissances utilisées pour le processus de recherche.
En effet, lors d'une tâche difficile, les utilisateurs avec une bonne connaissance du domaine mettent moins de temps à voir le contenu des pages. L'expertise concerne aussi les caractéristiques des utilisateurs au niveau de leur expérience à faire des recherches et leurs différentes compétences cognitives. Dans ce contexte, dans (Al-Maskari et Sanderson, 2011), les auteurs ont montré que les utilisateurs avec plus de compétences et d'expérience ont tendance à trouver plus de documents pertinents et à accomplir la tâche de RI avec succès.
Partant de l'hypothèse qu'il existe une différence entre les experts et les novices dans la recherche, les auteurs de (Hembrooke *et al.*, 2005) ont conduit une étude comparative des comportements de recherche pour explorer ces différences et ont constaté les principaux points suivants : niveau de la complexité du choix de termes de la requête, différences dans le nombre de termes utilisés, et une différence au niveau de la définition de la stratégie de recherche. Ils ont conclu que les experts du domaine soumettent des requêtes plus complexes et plus longues que les novices.

3.2.2 Caractéristiques linguistiques

Des observations issues de travaux sur le sujet "*topic*" de la requête médicale (Zeng *et al.*, 2006; Dogan *et al.*, 2009; Zhang, 2010) ont permis de définir la notion de sujet de requête en utilisant des propriétés linguistiques ou des articles médicaux. Cela veut dire que les utilisateurs n'utilisent pas en général les terminologies et taxonomies du domaine, par contre une utilisation accentuée de leurs propres termes a été observée entrainant beaucoup de fautes d'orthographe et utilisation excessive d'abréviations. Ces études concernent des recherches sur le web et incluent des participants de différents milieux, diverses tâches et une variété de compétences. Plus spécifique au domaine médical, des études ont été conduites pour, d'une part, analyser les différences de stratégies de recherche utilisées entre les experts du domaine médical et les novices (Bhavnani, 2001; Bhavnani, 2002), et se sont focalisés sur le comportement de recherche d'information des experts (Ely *et al.*, 2002; Wildemuth, 2004) d'autre part.
Par ailleurs, plusieurs caractéristiques du besoin en information concernent les caractéristiques linguistiques. En effet, les requêtes ou les questions représentent un des principaux moyens par lequel les utilisateurs interagissent avec un système d'information. Par conséquent, les caractéristiques des expressions des utilisateurs impactent la qualité des résultats dans l'interaction utilisateur-système. Dans (Zhang et Fu, 2011), les auteurs ont analysé les caractéristiques linguistiques des besoins en information de santé sous le moteur de recherche *Yahoo* selon deux niveaux :

— *Niveau Terme* : les erreurs les plus courantes sont les fautes d'orthographe, les utilisateurs ont des difficultés dans l'écriture des termes/concepts médicaux (niveau orthographique). Sur un autre niveau, les utilisateurs ne trouvent pas les termes appropriés pour décrire leurs conditions.

— *Niveau Question* : le nombre de questions posées dans un message sont au nombre d'une (75%), deux (18%) ou trois questions (7%).

Un ensemble de critères et de propriétés caractérisent le besoin en information de la requête médicale, résumés dans le Tableau 2.9, et donne un aperçu sur les critères de requêtes médicales

et les principales conclusions des travaux de l'état de l'art. Nous avons étudié et analysé les caractéristiques des besoins en information, en identifiant les critères examinés, formellement définis et mesurés dans la littérature. Ces critères sont décris ci dessous.

— **Longueur de la requête** : La longueur de la requête est un critère important, qui permet de comprendre le but de la recherche de l'utilisateur et de retourner des documents susceptibles de répondre au besoin spécifique de l'utilisateur. En effet, dans un contexte général de recherche (recherche sur le web) et d'une manière spécifique (médical, social, etc.), les travaux ont montré la fiabilité de la longueur dans la mesure où elle a un rôle important dans la phase de RI. Dans le domaine médical, les requêtes exprimées par les utilisateurs sont généralement courtes et la plupart ne dépassent pas les 4 termes, et cela concerne les requêtes soumises aux moteurs de recherche et aux SRI médicaux.

— **Difficulté de la requête** : La difficulté de la requête est liée au succès ou l'échec du processus de recherche lancé par la requête. Plusieurs stratégies de recherche accomplies au niveau de la requête influencent la difficulté de la requête comme par exemple la représentation entre le vocabulaire de la requête et les entrées des terminologies médicales, l'utilisation des facettes sémantiques, l'utilisation de termes additionnels, etc.

— **Structure et vocabulaire** : Ce critère fournit des indices connus sur lesquels se basent tout les modèles de RI connus (Baeza-Yates *et al.*, 1999). En effet, le vocabulaire et la structure de la requête sont des propriétés caractérisant les modèles de RI qui se basent sur l'appariement requête-document ou les requêtes et les documents sont considérés comme des sacs de mots. De ce fait, et dans le contexte de RI médicale, ces caractéristiques ont été étudiées. Il s'est avéré que le vocabulaire de la requête médicale contient beaucoup de fautes d'orthographe et se caractérise par l'utilisation d'abréviations et d'acronymes ce qui augmente l'écart d'appariement document-requête basé sur les mots. Notons que, plusieurs études ont montré que les termes de la requête ne correspondaient pas forcément aux termes des terminologies médicales, ce qui crée le problème de l'écart sémantique.

— **Objectif de recherche** : L'objectif de la recherche d'information médicale représente le but derrière l'expression de besoin en information. Comprendre donc le besoin en information des utilisateurs a une grande importance pour identifier les stratégies de recherche d'information optimales et efficaces. L'analyse des sessions de recherche médicales via les questionnaires et les analyses sémantiques des résultats a montré que les utilisateurs effectuent leur recherche sur un grand nombre d'informations médicales et qui couvrent différents aspects : la gestion des maladies, le diagnostic, des conseils, le dosage de médicaments et l'actualisation des savoirs et des connaissances, etc.

Concernant les formes des requêtes, trois principales conclusions émergent d'études antérieures, énumérées ci dessous.

1. **Les requêtes médicales sont courtes** : Plusieurs études (Hong *et al.*, 2002; Zeng *et al.*, 2002b; Natarajan *et al.*, 2010) ont conclu que les requêtes sont généralement courtes, contenant moins de 3 termes avec un nombre moyen de termes égal à 2. Dans (Zeng *et al.*, 2002b), les auteurs ont analysé les requêtes de *MedlinePlus* et les sessions de recherche d'information en santé dans les hôpitaux, et ont trouvé que le nombre de termes des requêtes est compris ente 1 et 3. Les mêmes résultats ont été trouvés dans (Hong *et al.*, 2002) qui ont analysé les requêtes de *Healthlink* sur la base de 377000 requêtes issues des fichiers de transaction.
2. **Les termes de la requête ne sont pas fortement liés aux vocabulaires médicaux** : Des études (McCray et Tse, 2003; Yang *et al.*, 2011) ont identifié les sujets des

requêtes en utilisant des critères linguistiques. Ces études montrent que les utilisateurs n'utilisent pas forcément les terminologies médicales, en revanche, ils utilisent leur propre vocabulaire avec des fautes typographiques et des abréviations. À titre d'exemple, (Yang *et al.*, 2011) ont étudié un moteur de recherche spécialisé dans les dossiers de patients appelé *EMERSE*. L'étude expérimentale a montré que 18.9% des requêtes contiennent au moins un acronyme. Une autre étude développée dans (McCray et Tse, 2003) montre, suite à l'analyse de 4700 requêtes issues de *ClinicalTrials.gov* et *MedlinePlus*, que l'échec des requêtes a été principalement lié aux fautes typographiques et à l'utilisation d'abréviations.

3. **Les sujets des requêtes sont peu précis** : De nombreuses études (Song *et al.*, 2010; Liu et Huang, 2011) ont montré que les sujets des requêtes sont vagues. Les auteurs ont alors proposé des systèmes de recommandation de requêtes en s'appuyant sur le résultat qui indique que les requêtes contenant des termes les mieux corrélés aux catégories médicales retournent plus de documents pertinents.

Critère	Observations et constatations
Longueur de la requête	Les requêtes sont généralement courtes de 1.5 à 4 termes (Magrabi *et al.*, 2005; Zhang, 2014) ; de 1.5 à 2 termes selon (Lykke *et al.*, 2012), de 1.79 à 5 termes avec une moyenne de 2.81 (Zhang, 2014), une moyenne de 3 termes selon (Herskovic *et al.*, 2007b) et moins de 3 termes dans le moteur de recherche Caseof (Spink *et al.*, 2004).
Structure et vocabulaire	Utilisation fréquente des associations de catégories sémantiques (Dogan *et al.*, 2009) ; les termes ne correspondent pas forcément aux vocabulaires standards (Keselman *et al.*, 2008; Zhang et Fu, 2011) ; l'utilisation fréquente des termes inappropriés et des erreurs d'orthographe et de syntaxe ainsi que l'utilisation fréquente des abréviations (Boden, 2009; Zhang et Fu, 2011), une requête peut contenir deux ou trois sous-requêtes qui couvrent différentes facettes du besoin en information (Zhang et Fu, 2011).

Tableau 2.9 – Tableau récapitulatif des études et observations sur les caractéristiques des requêtes médicales (Tamine *et al.*, 2015)

3.2.3 Sujet de la requête

Le sujet de la requête reflète généralement le thème de la recherche. Dans (Zhang et Fu, 2011), les auteurs ont examiné les différents types d'information des utilisateurs ainsi que les caractéristiques des requêtes formulées par les utilisateurs en analysant des questions postées sur *Yahoo*. Cela a révélé un ensemble de typologies du besoin en information médicale, à savoir :

— *Information sur une maladie/problème particuliers* : ces questions concernent un aspect spécifique d'une maladie, principalement des symptômes des causes diagnostic, traitements et pronostiques.

— *Information sur les médicaments ou suppléments* : des questions sur les médicaments ou

des suppléments selon différents thèmes, à savoir : si un médicament est efficace, quel médicament prendre, combien et quand, s'il y a des effets secondaires, etc.

— *Information sur le mode de vie, notamment l'alimentation et l'exercice* : Les questions concernant le mode de vie sont souvent demandées pour des recommandations pour une alimentation saine ou un exercice de routine, étant donné le poids, la taille, l'âge ou l'état de santé spécifique ou encore pour vérifier avec d'autres si leur programme d'entraînement actuel est raisonnable.

— *Les personnes ayant des conditions similaires* : Certaines personnes utilisent le moteur de recherche pour chercher des cas similaires de personnes ayant la même maladie ou les même symptômes, avoir des informations sur un traitement ou un médicament.

— *Les sources d'information* : Certaines questions posées concernent les sources d'information, comme les sites Web fiables pour acheter des médicaments, pour poser des questions précises, ou encore voir ou consulter des statistiques.

— *Autres* : Certaines questions concernent des professions médicales, des mécanismes de fonctionnement de l'organisme, des lectures des tests médicaux, des informations pour faire face à des situations difficiles de la vie et de la gestion du stress ou encore l'assurance de santé.

Les travaux de (Nicholas *et al.*, 2003; Andreassen *et al.*, 2007; Zhang et Fu, 2011) ont rapporté que les recherches médicales conduites par les utilisateurs sur le web couvrent un large spectre de besoins en information sur différents thèmes, à savoir : les causes et les traitements des maladies, des informations sur la nutrition et l'alimentation et des conseils pour une mode de vie sain.

Les auteurs dans (Herskovic *et al.*, 2007a) ont analysé des fichiers de log de PubMed et ont identifié une catégorisation des sujets de requêtes en utilisant la hiérarchie MeSH. Ils ont rapporté que les thèmes des requêtes sont variés, sans qu'il n'y ait de thèmes et de termes dominants.

3.3 Stratégies et comportements de recherche

De manière générale, la stratégie et le comportement de recherche sont des facteurs corrélés. Le comportement de recherche durant la tâche de RI médicale fait partie de la stratégie de recherche. En effet, le comportement des utilisateurs influence la stratégie de recherche et concerne 4 critères importants, à savoir : les ressources utilisées, la stratégie de recherche, le modèle du processus de recherche et la difficulté de la recherche. Les Tableaux 2.10 et 2.11 présentent un récapitulatif des études conduites dans ce cadre ainsi que les conclusions et observations notées. Nous retenons :

— **Stratégie de recherche** : Les stratégies de recherche des utilisateurs dans le domaine médical ne se caractérisent pas par des particularités spécifiques comparées au processus de recherche général. Les stratégies principales sont en effet les suivantes : affiner la requête, consulter, rechercher et examiner les résultats retournés, plus spécifiquement les *top-ranked*.

— **Difficulté de la recherche** : À la différence de la difficulté de la requête, la difficulté de la recherche est traitée du côté de l'utilisateur. Plusieurs facteurs sont liés à l'utilisateur, comme la catégorie ou le type de l'utilisateur (étudiant, infirmier, médecin, etc.), l'expérience de l'utilisateur et la tâche visée par l'utilisateur, représentant tous des indicateurs importants de la difficulté de recherche.

— **Ressources utilisées** : Plusieurs ressources d'information sont exploitées par les utilisateurs (experts *vs* non experts) pour mener leur tâche de recherche d'information.

— **Processus de recherche** : Ce critère, lié au "mouvement" cognitif de l'utilisateur est spécifique aux tâches de RI médicale comme par exemple les diagnostics. Les conclusions sur les préoccupations majeures sur le processus de recherche sont les suivantes : (1) les approches de raisonnementd par hypothèse/déduction sont largement adoptées par les médecins et comportent deux phases d'analyse (phase dirigée par les preuves et phase dirigée par les hypothèses) ; (2) le raisonnement basé sur les preuves ayant pour but de lier les symptômes aux maladies correspondantes ; (3) intensification liée au modèle mental de l'utilisateur qui vise à construire un lien entre les symptômes et maladies graves d'un côté, et les symptômes graves et les troubles bénins d'un autre côté.

Critère	Observations et constatations
Stratégie de recherche	La recherche fréquente utilisant les mots clés de MEDLINE, les filtres des méta-données (langues, date de publication, auteurs, etc.), les opérateurs booléens et les profils afin de limiter la durée de la recherche ; amélioration de la requête, reformulation de requêtes (Lykke *et al.*, 2012; Zhang, 2014), utilisation de composantes sémantiques pour la recherche par facettes (Mu *et al.*, 2011; Lykke *et al.*, 2012), commencer la recherche par des moteurs de recherche généraux (Spink et Jansen, 2006), examiner généralement les top pages web retournées (les 5 premiers) ou la première page (Toms et Latter, 2007; Zhang, 2012) ; les stratégies de recherche évoluent avec les connaissances du domaine (Wildemuth, 2004) ; le classement des savoirs et des connaissances commençant par les ressources spécifiques du domaine (Bhavnani, 2001) ; la transition entre recherche et navigation pour les tâches complexes (Kules et Xie, 2011).
Difficulté de la recherche	La formation et l'expérience acquise avec les moteurs de recherche médicaux induisent l'amélioration des résultats ; le type de l'utilisateur (Pao *et al.*, 1994; Hersh *et al.*, 2000; Hersh *et al.*, 2002) et la tâche de l'utilisateur (Hersh *et al.*, 2002; Inthiran *et al.*, 2012; Zhang, 2014).

Tableau 2.10 – Tableau récapitulatif des études et observations sur le comportement de recherche dans le domaine médical (Partie 1) (Tamine *et al.*, 2015)

La stratégie de recherche dépend aussi de la tâche et de la situation surtout dans le domaine médical (Kules et Xie, 2011). Selon ces derniers, les utilisateurs peuvent regarder et utiliser plus les facettes de recherche quand les conditions médicales sont perçues comme étant graves, la transition entre la recherche et la navigation pour les tâches complexes. Les étudiants en médecine ont tendance à regarder l'ensemble des résultats mais moins de références, ils sont plus capables que les étudiants en soins infirmiers et les praticiens de convertir les réponses incorrectes en réponses correctes (Hersh *et al.*, 2002).

De nombreux travaux ont révélé des stratégies typiques spécifiques aux utilisateurs de SRI médicaux, dépendant de leur niveau d'expertise (Bhavnani, 2001). De manière générale, les travaux ont mis en évidence trois principales facettes du comportement, à savoir : (1)

Critère	Observations et constatations
Ressources utilisées	Utilisation de ressources non informatiques (livres, journaux..), ressources personnelles (parents ou amis médecins), ressources électroniques comme MEDLINE, bases de données médicales électroniques de plus en plus utilisées par les docteurs parfois quand les patients sont en train d'attendre ou au cours d'une consultations, utilisation de différentes ressources du web pour une même tâche de recherche : les moteurs de recherche comme Google pour la santé, les agents de santé, les sources du Web 2.0, l'utilisation d'une combinaison de bases de données médicales disponibles améliore l'efficacité de la recherche (Alper *et al.*, 2001; Cullen, 2002; Magrabi *et al.*, 2005; Andrews *et al.*, 2005; Zhang, 2012).
Processus de recherche	Les recherches basées sur les hypothèses (vérifier les hypothèses, recherche proche d'une hypothèse, rechercher sans hypothèse) (White et Horvitz, 2009a), les recherches basées sur les preuves (construire des modèle mentaux sur les signes et les symptômes en relation avec les troubles) (White et Horvitz, 2009a), raisonnement déductif basé sur les hypothèses en deux phases : dirigé par les preuves puis dirigé par les hypothèses (Eastin et Guinsler, 2006), partir des symptômes connus aux maladies graves vs. symptômes rares aux interprétations bénignes (White *et al.*, 2009), un processus d'essais et d'erreurs (Toms et Latter, 2007).

Tableau 2.11 – Tableau récapitulatif des études et observations sur le comportement de recherche dans le domaine médical (Partie 2) (Tamine *et al.*, 2015)

approche globale de recherche (Zhang, 2010) : les études montrent que la recherche est basée sur un processus itératif essai-erreur caractérisé par des transitions entre recherche et navigation ; (2) accès aux résultats (Toms et Latter, 2007) : de manière analogue aux autres utilisateurs de SRI, une préférence est clairement exprimée pour la haute précision ; (3) intentions, buts et motivations (Oh, 2012) : les résultats d'études empiriques montrent que la motivation est le principal facteur qui détermine l'échec ou le succès des sessions de recherche, plus particulièrement dans un cadre de recherche sociale.

Plus spécifiquement, concernant les experts médicaux, les auteurs dans (Ely *et al.*, 2002), ont montré que les experts n'ont pas de stratégie optimale de recherche, qu'ils sont démotivés en se doutant de trouver des réponses crédibles à leurs requêtes et enfin, qu'ils ne choisissent pas des ressources fiables. En complément, les auteurs dans (Lykke *et al.*, 2012) ont analysé la différence entre les pratiques de recherche d'information initiée par des médecins en considérant le lieu d'émission : dans l'environnement professionnel ou en dehors. Ils ont conclu que les requêtes en milieu professionnel sont très ciblées visant une haute précision alors que les requêtes émises en dehors du milieu professionnel sont exprimées à l'aide d'un vocabulaire généraliste.

De plus, les études sur les stratégies de recherche ont montré qu'il n'y a pas de particularités

spécifiques qui caractérisent les utilisateurs qui cherchent l'information médicale comparé à d'autres types de recherche. Les stratégies principales se résument en trois catégories, à savoir : (1) raffinement de la requête, (2) la navigation et la recherche ; et (3) l'analyse des résultats. D'après (Lykke *et al.*, 2012; Zhang, 2014), la recherche fréquente dans MEDLINE avec les mots clés, utilisant les filtres de méta-données (la langue utilisée, la date de publication, etc), les opérateurs booléens et les profiles permet de limiter le nombre de résultats retournés, l'amélioration de requêtes et la reformulation de requêtes. Dans (Mu *et al.*, 2011; Lykke *et al.*, 2012), les auteurs ont observé l'utilisation de composantes sémantiques pour la recherche basée sur les facettes. Par ailleurs, les auteurs de (Spink *et al.*, 2004) ont remarqué que les utilisateurs commencent la recherche avec des moteurs de recherche généraux, ils examinent les premières pages retournées par le moteur de recherche en générale ; les 5 premières, ou uniquement la première page (Toms et Latter, 2007; Zhang, 2012). L'importance d'élaborer une stratégie de recherche a été montrée par (Wildemuth, 2004).

Concernant le comportement des utilisateurs, les travaux se sont globalement intéressés à la caractérisation des sessions de recherche, le principe de reformulation de requêtes, les clics de souris, et l'influence du comportement sur la qualité des résultats (Richesson *et al.*, 2010). D'autres études se sont focalisées sur les pratiques de recherche comme le jugement de pertinence, les sites visités, les types d'information utilisés et ce, dans le but d'identifier les facteurs qui contribuent au succès ou à l'échec de la recherche (Dogan *et al.*, 2009).

4 Techniques et modèles de RI médicale

Dans le domaine médical, la plupart des techniques et modèles d'appariement reposent sur des modèles de RI sémantiques ou conceptuels qui se basent sur les terminologies biomédicales comme MeSH, UMLS ainsi que différentes ressources médicales pour améliorer les performances des systèmes de RI (Mao *et al.*, 2015). Une partie des travaux se sont intéressés à proposer des approches de reformulation de requêtes afin d'améliorer leur représentation (Zhou *et al.*, 2007; Stokes *et al.*, 2009; Trieschnigg, 2010). D'autres travaux se sont basés sur l'extraction des concepts à partir des documents et des requêtes pour améliorer leur représentation (Diem *et al.*, 2007; Gobeill *et al.*, 2009). Nous présentons dans ce qui suit deux approches fondamentales dans le domaine de la RI vu leur influence sur le processus de recherche, à savoir : l'expansion de requêtes et l'expansion de documents.

4.1 Expansion de requêtes

L'expansion de requêtes est une technique très répandue pour traiter le problème d'inadéquation du vocabulaire entre les requêtes et les documents. L'approche générale consiste à ajouter ou re-pondérer les termes de la requête initiale de l'utilisateur pour améliorer les résultats de recherche (Voorhees, 1994). Plusieurs techniques ont été développées pour améliorer la performance de la recherche d'information, une des techniques est l'expansion de requêtes, *i.e.*, ajouter des termes supplémentaires à la requête initiale (Baeza-Yates *et al.*, 1999). Les sources typiques de ces termes sont les thésaurus ou les documents retournés par les SRI. Le thésaurus propose des synonymes, antonymes, descendants, ou autres termes connexes. La fréquence des termes dans la requête et d'autres termes dans certaines collections, des données du log qui enregistrent le comportement de recherche des utilisateurs, sont devenus

aussi des sources pour les termes d'expansion (Cui *et al.*, 2003). Dans la littérature récente, plusieurs variations des techniques d'expansion de requêtes ont été proposées. Quelques techniques combinent différentes méthodes d'expansion, par exemple, combiner le feedback de recherche avec la cooccurrence de l'information (Xu et Croft, 1996) ou combiner différents thésaurus (Mandala *et al.*, 1999). Certains ont exploré la nature floue de la liaison entre les termes ou les concepts (Akrivas *et al.*, 2002).

Un des problèmes de la complexité des requêtes cliniques est le "fossé" sémantique entre le vocabulaire de la requête et celui du document. Pour traiter cette question, plusieurs travaux ont proposé des techniques de reconnaissance et identification des éléments de requêtes, basées sur des vocabulaires spécifiques au domaine. Cela consiste à lier le texte des requêtes et les documents aux concepts appartenant aux entrées des terminologies médicales telles que *UMLS*, *MeSH* et *ICD*. Plus particulièrement, l'expansion de requêtes basée sur les concepts est une méthodologie efficace pour traiter les problèmes d'inadéquation du vocabulaire (Hersh *et al.*, 2000; Plovnick et Zeng, 2004; Stokes *et al.*, 2009). La plupart de ces approches ont exploré l'utilisation des synonymes (Plovnick et Zeng, 2004), ou encore les relations hypernymes et hyponymes définies dans UMLS (Hersh *et al.*, 2000; Stokes *et al.*, 2009) afin d'étendre les requêtes originales avec les concepts préférés qui ont le sens le plus proche des termes originaux de la requête. Les analyses basées sur les règles spécifiques se sont montrées efficaces dans (King *et al.*, 2011). Ces règles ont permis de détecter l'incertitude permettant de réduire le nombre de concepts candidats utilisés pour l'expansion de la requête. La reformulation conceptuelle de la requête basée sur les terminologies médicales a été le sujet de plusieurs travaux (Srinivasan, 1996; Zhou *et al.*, 2007; Stokes *et al.*, 2009). Ils ont montré une amélioration de la performance de RI avec ces méthodes d'expansion.

Dans (Zhu *et al.*, 2014), les auteurs ont utilisé une technique d'expansion de requêtes pour extraire les termes d'expansion à partir de quatre collections de documents, à savoir : documents du web, dossiers médicaux des patients et deux collections d'articles médicaux. Plus particulièrement, ils ont suggéré que le fait d'utiliser les termes d'expansion extraits de ces quatre collections pour étendre la requête est plus efficace qu'utiliser les termes d'expansion pour chaque collection individuellement.

D'autres travaux ont conclu qu'il n'y a aucune amélioration des performances de RI avec l'expansion conceptuelle de la requête. En effet, les performances de la RI dépendent notamment de la façon dont les concepts sont identifiés dans le texte et comment ils sont exploités pour modifier ou reformuler la requête. Par exemple, l'approche conceptuelle proposée par (Hersh *et al.*, 2000) est basée sur l'identification manuelle des concepts issus d'UMLS à partir de chaque requête de la collection OHSUMED. Les requêtes sont ensuite reformulées avec une expansion des termes désignant les concepts. Les termes qui sont ajoutés à la requête originale sont : des synonymes, les termes issus des relations hiérarchiques (parent, ascendant, descendant), les termes reliés aux concepts ou encore les termes qui représentent la définition du concept. La phase de recherche est basée sur un modèle vectoriel. Ils ont ainsi observé une dégradation des performances de la RI, au niveau du rappel et de la précision.

Dans le cadre de la campagne TREC Medical 2011, une tendance commune des participants était d'utiliser l'expansion de requêtes, en soulignant la nécessité d'aborder le niveau de richesse de la synonymie dans les textes cliniques. Il y avait des variations dans les ressources utilisées

pour effectuer l'expansion. Les ontologies spécifiques au domaine médical comme UMLS ou MeSH (Daoud *et al.*, 2011), ou les ressources du web comme MEDLINE (Córdoba *et al.*, 2011). D'autres approches ont utilisé des ressources non spécifiques au domaine médical, comme DBpedia (Karimi *et al.*, 2011) ou encore Wikipédia. D'autres techniques de pseudo-réinjection de pertinence (PRF) ont été proposées et utilisées afin de générer les requêtes étendues basées sur les concepts identifiés dans les documents les mieux classés (Trieschnigg, 2010). Ces travaux ont montré l'avantage de combiner les termes et les concepts pour étendre les requêtes (Stokes *et al.*, 2009). Plus spécifiquement, dans le cadre de la RI biomédicale, plusieurs travaux ont exploité la technique de reformulation de requêtes PRF pour améliorer les performances de la recherche d'information (Srinivasan, 1996; Jiang et Zhai, 2007; Zhou *et al.*, 2007).

Dans (Zhou *et al.*, 2007), les auteurs ont proposé une méthode pour étendre la requête originale par les termes désignant les concepts MeSH trouvés dans les 15 premiers paragraphes retournés par le système de RI. Ensuite, les k premiers concepts les plus significatifs pour la requête sont retenus pour la reformulation. Afin de sélectionner les concepts, ils se sont basés sur une mesure de similarité concept/requête.

De plus, (Voorhees et Ellis, 2014; Soldaini *et al.*, 2015b; Soldaini *et al.*, 2015a) montrent que les systèmes de prise de décision médicaux utilisent des techniques de reformulation de requêtes afin d'améliorer les résultats de recherche. Par exemple, (Soldaini *et al.*, 2015b) ont proposé une approche de reformulation de requêtes pour aider les praticiens du domaine médical à sélectionner des publications pertinentes utilisées pour leur rapports cliniques. En comparant différentes méthodes de reformulation et d'expansion d'une part, et de réduction de requêtes d'une autre part, leur étude a montré que la méthode de réduction de requête améliore la performance.

Depuis des années, les techniques d'expansion de requêtes sont utilisées d'une manière optimale dans la RI médicale. Hersh et al (Hersh *et al.*, 2000) ont étendu les requêtes avec des termes manuellement sélectionnés et identifiés à partir des relations du métathésaurus UMLS pour améliorer la performance de la recherche d'information. Les résultats expérimentaux ont montré que l'expansion de la requête basée sur le thésaurus n'améliore pas nécessairement les résultats de recherche.

Dans une étude récente, les auteurs dans (Limsopatham *et al.*, 2015) ont exploré deux approches pour modéliser la probabilité que les termes d'expansion extraits de la collection de documents soient efficaces pour améliorer la représentation de la requête. Une première approche adapte une technique de sélection de ressources pour mesurer la vraisemblance qu'un document de la collection peut fournir de bons termes d'expansion extraits de cette collection, puis utiliser la probabilité pour mesurer les poids des termes d'expansion dans la collection. La deuxième approche combine les scores de pertinence des requêtes étendues calculés pour chaque document de la collection utilisant une technique d'ordonnancement (*Learning To Rank*).

4.2 Expansion de documents

Plusieurs approches d'expansion de documents ont été proposées dans la littérature dans le contexte de la RI pour résoudre le problème de défaut d'appariement entre les documents et les requêtes. L'expansion des documents a été appliquée dans divers domaines de la recherche d'information, comme la recherche de textes courts (Efron *et al.*, 2012), de discours (Singhal

et Pereira, 1999), la recherche d'images (Min *et al.*, 2010) et la RI dans le domaine médical (Trieschnigg *et al.*, 2006; Thesprasith et Jaruskulchai, 2014; Diem *et al.*, 2007; Dinh et Tamine, 2011b; Ghezaiel *et al.*, 2012). Plusieurs méthodes et techniques ont été utilisées pour la manipulation des vecteurs de documents en ajoutant les termes qui sont les plus significatifs et proches par rapport au vecteur de la requête. D'autres se sont orientés vers l'expansion documentaire basée sur des ressources et terminologies comme Wordnet (Agirre *et al.*, 2010), ou encore dans le contexte biomédical avec l'extraction de concepts du thésaurus MeSH (Thesprasith et Jaruskulchai, 2014) ou encore méta-thésaurus UMLS (Trieschnigg *et al.*, 2006; Diem *et al.*, 2007). En effet, dans le domaine de la RI biomédicale, la technique d'expansion conceptuelle a été introduite, permettant l'exploitation des concepts et des ressources médicales à la fois dans les documents et dans les requêtes (Trieschnigg *et al.*, 2006; Diem *et al.*, 2007; Thesprasith et Jaruskulchai, 2014; Dinh et Tamine, 2011b; Dinh et Tamine, 2011c; Dinh et Tamine, 2010).

(Thesprasith et Jaruskulchai, 2014) ont proposé une approche d'indexation et de RI conceptuelle basées sur une expansion combinée de documents et de requêtes utilisant les concepts MeSH. Ils se basent sur la construction d'une structure d'index qui comprend chacun des concepts MeSH, sa description et ses synonymes indexés comme un document unique. Le texte des documents et requêtes sont associés aux meilleurs concepts issus du thésaurus MeSH en utilisant une approche d'extraction de concepts. Enfin, les termes identifiés dénotant les concepts MeSH sont utilisés pour étendre le document et la requête. Par ailleurs, (Trieschnigg *et al.*, 2006) ont proposé d'ajouter les concepts issus du méta-thésaurus UMLS aux documents et à la requête. La recherche d'information a été effectuée en exploitant les concepts identifiés dans les documents et dans la requête dans un modèle de langue unigramme avec la méthode de lissage de Jelinek-Mercer. Utilisant l'outil MetaMap, (Diem *et al.*, 2007) ont étendu le contenu des requêtes et documents en exploitant les concepts ainsi que leurs relations sémantiques extraites de UMLS. L'évaluation de leur approche sur la collection ImageCLEFMed a montré une amélioration significative.

5 Modèles de RI basées sur le paradigme PICO

5.1 EBM et paradigme PICO

Plusieurs facteurs jouent un rôle très important dans la prise de décision médicale. Une des approches, est basée sur les preuves scientifiques, permettant de retrouver et évaluer l'information pertinente et nécessaire à la prise de décision est la médecine basée sur les faits (EBM [13]), le concept en français appelé MFF [14]. La médecine basée sur les preuves est fondée principalement sur quatre étapes :

1. La formulation claire et précise d'une question clinique à partir d'un problème clinique donné ;
2. La recherche d'articles pertinents dans la littérature (quel article lire ?) ;
3. L'évaluation systématique de la validité et de l'intérêt des résultats, et l'extraction des preuves qui sont à la base des décisions cliniques (quel article croire ?) ;
4. L'intégration de ces preuves dans la pratique médicale courante afin de répondre à la question posée au départ.

13. Evidence-Based Medicine
14. Médecine fondée sur les faits

La Figure 2.3 présente les étapes fondamentales pour récupérer l'information nécessaire à la prise de décision clinique.

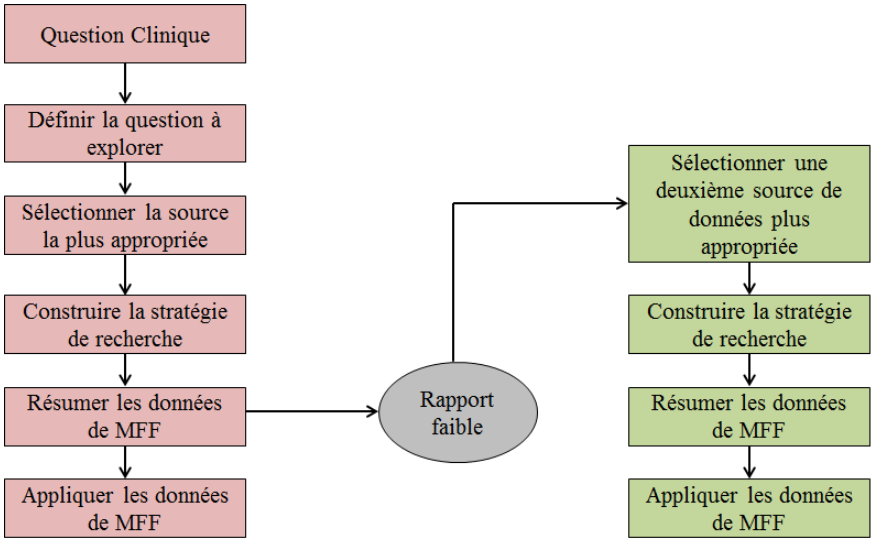


FIGURE 2.3 – Stratégie de recherche globale de la MFF.

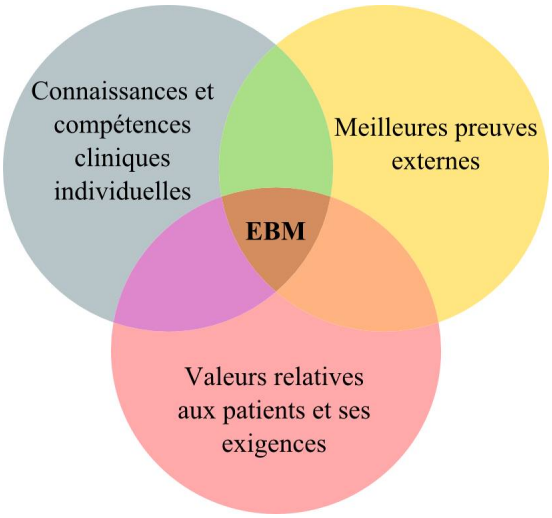


FIGURE 2.4 – Approche de l'évidence basée sur les faits.

L'Evidence-Based Medicine (EBM) ou médecine factuelle se définit donc comme l'utilisation consciencieuse et judicieuse des meilleures données (preuves) actuelles de la recherche clinique dans la prise en charge personnalisée de chaque patient (Sackett *et al.*, 1996). Ces preuves proviennent d'études cliniques systématiques, telles que des essais contrôlés randomisés, des méta-analyses, éventuellement des études transversales ou de suivi bien construites. La Figure 2.4 donne les composantes principales de l'approche EBM.

Les soins basés sur les preuves concernent trois volets :

— *Soin de santé basé sur les preuves* : utilisation des meilleures preuves pour la prise de décision médicale sur le soin des patients ou les services de santé. Les meilleures preuves

actuelles sont les preuves mises à jour, à partir des travaux pertinents et valides sur les effets de différentes forme de soins, les risques potentiels, la précision des tests de diagnostic et les facteurs prédictifs de pronostic (Greenhalgh, 2004).

— *La pratique clinique basée sur les preuves* : une approche de prise de décision, dans laquelle le clinicien utilise la meilleure preuve disponible, lors de la consultation avec le patient, pour prendre une bonne décision pour le patient (Paton, 1999).

— *La médecine basée sur les preuves* : c'est l'utilisation des meilleures preuves pour la prise de décision sur la santé des patients. La pratique de la médecine basée sur les faits signifie l'intégration de l'expertise clinique individuelle avec la meilleure preuve clinique externe issue des recherches systématiques (Sackett *et al.*, 1996).

Les requêtes cliniques se composent de quatre aspects fondamentaux : Patient/Problème (P), Intervention (I), Comparaison (C) et Outcome (O) connus comme éléments PICO. La structure PICO est utilisée par les experts du domaine médical pour exprimer leurs requêtes et ses éléments peuvent être identifiés à partir des documents de la littérature. Une extension de la question PICO est PICOTT [15]. Le Tableau 2.12 présente un exemple de requête clinique PICOTT. Ces questions sont formulées par les experts du domaine médical dans le cadre de la RI basée sur les preuves. Malgré l'expertise, formuler une question précise avec un contexte bien spécifique reste une tâche complexe.

P	Patient, Population, or Problem	How would I describe a group of patients similar to mine ?
I	Intervention, Prognostic Factor, or Exposure	Which main intervention, prognostic factor, or exposure am I considering ?
C	Comparison or Intervention	What is the main alternative to compare with the intervention ?
O	Outcome you would like to measure or achieve	What can I hope to accomplish, measure, improve, or affect ?
T	What type of question are you asking ?	Diagnosis, Etiology/Harm, Therapy, Prognosis, Prevention
T	Type of study you want to find	What would be the best study design/methodology ?

Tableau 2.12 – Exemple de requête clinique PICOTT

Nous avons distingué deux volets de travaux : (1) l'identification des éléments PICO, (2) l'utilisation de ces éléments dans la RI médicale, détaillés dans ce qui suit.

La théorie de la médecine fondée sur des preuves soulève des questions fondamentales sur la façon avec laquelle les médecins génèrent d'une manière fréquente des questions dans la pratique clinique, ainsi que les catégories de questions fréquemment posées (Ely *et al.*, 2000). Au cours des deux dernières décennies, un certain nombre d'études ont abordé directement ou indirectement, les besoins d'information des médecins. Une étude en particulier a décrit

15. Population, Intervention, Comparison, Type of study, Type of questions

Question Clinique : *In people with recurrent aggression having any antiepileptic drug in any dosage, whats is the length of time of placebo for observer reported aggression ?*	
P	*women with locally advanced cervical carcinoma*
I	*hyperthermia radiotherapy*
C	*radiotherapy*
O	*local tumor control*

Tableau 2.13 – Exemple de requête PICO

l'état du besoin en information clinique (Smith, 1996). Cette revue systématique met à jour cet examen et évalue les habitudes de recherche d'information des médecins en déterminant la nature de leurs préférences en matière de ressources d'information. En analysant le comportement des médecins, les auteurs de (Dawes et Sampson, 2003) ont conclu que les sources d'information les plus utilisées sont textuelles. De plus, les médecins ont tendance à demander l'information à des collègues et il est rare d'utiliser les différentes bases de données médicales. Par ailleurs, ils ont noté une grande variation dans le comportement de recherche d'information. Cela implique la nécessité de poursuivre la catégorisation des besoins en information et de sources d'information. De ce fait, une planification minutieuse de la prestation de renseignement aux médecins est nécessaire pour leur permettre de tenir à jour et d'améliorer le transfert des connaissances.
Plusieurs problèmes empêchent l'implémentation de la médecine basée sur les preuves au sein de la pratique clinique (Francke *et al.*, 2008). Ces problèmes proviennent de : (1) la complexité de la formulation d'une question précise, spécifique au contexte de la recherche qui permet de faciliter l'identification des preuves pertinentes, (2) un manque de niveau d'expertise suffisant, qui peut être utilisé pour réaliser une évaluation efficace de ces preuves. Selon (Richardson *et al.*, 1995), une question PICO est bien composée de 4 facettes clés : P (Patient/Problem), I (Exposure/Intervention), C (Comparison) et O (Outcome). Le Tableau 2.13 représente un exemple de question PICO. Peu de travaux ont traité le problème de répondre spécifiquement aux questions PICO (Demner-Fushman et Lin, 2007; Boudin *et al.*, 2010b; Boudin *et al.*, 2010c). Ces approches se basent sur deux étapes, à savoir : (1) l'identification des facettes PICO dans les documents et les requêtes, et (2) intégrer ces facettes PICO dans le processus de recherche, et plus spécifiquement dans les modèles d'appariement requête-document.

Les travaux sur l'évaluation automatique des requêtes PICO sont peu abondants et se scindent en deux volets. Dans la première catégorie de travaux (Boudin *et al.*, 2010a; Zhao *et al.*, 2010), le problème principal traité par les auteurs est la détection des facettes PICO, comme une étape en amont à la sélection de documents pertinents. La plupart des approches sont basées sur des techniques d'apprentissage supervisé afin d'identifier les éléments PICO à partir du texte. Par exemple, dans (Boudin *et al.*, 2010a), le processus d'identification des éléments PICO à partir du document a été conduit selon deux étapes : une première étape pour la segmentation du texte des documents en plusieurs phrases, puis dans une seconde étape, chaque phrase est transformée en un vecteur de propriétés utilisant les caractéristiques statistiques et linguistiques pour désigner les facettes P, IC et O. Les expérimentations sur un ensemble de 260000 résumés de PubMed ont montré que la combinaison linéaire de plusieurs classifieurs est l'approche la plus efficace pour la détection des éléments PICO.

La deuxième catégorie de travaux (Boudin *et al.*, 2010c; Boudin *et al.*, 2010b; Demner-Fushman et Lin, 2007) concerne la définition de modèles de recherche d'information qui exploitent les facettes PICO pour calculer les scores de pertinence des documents. Pour atteindre cet objectif, Boudin et al. (Boudin *et al.*, 2010c; Boudin *et al.*, 2010b) ont proposé une extension de la version originelle du modèle de langue (Song et Croft, 1999). Les auteurs ont modifié le modèle de pondération basé sur les termes des documents en tenant compte de la distribution des éléments PICO dans les différents passages de documents ainsi que la distribution des termes dans les différentes parties PICO. L'évaluation expérimentale conduite sur une collection de 1.5 millions de documents et 423 requêtes a montré que le modèle proposé a permis une amélioration de 28% de la MAP[16] sur l'ensemble des modèles de référence. Demner-Fushman et Lin (Demner-Fushman et Lin, 2007) ont également proposé un modèle unifié pour détecter et utiliser les éléments PICO dans une fonction de calcul de pertinence des documents S_{EBM}. Cette dernière est basée sur une combinaison linéaire des scores de pertinence partiels des documents, considérant trois éléments de l'EBM, à savoir, la structure PICO (S_{PICO}), la crédibilité de la preuve médicale (S_{SoE}) et le type de la tâche (S_{task}). Les expérimentations sur 24 questions cliniques ont montré que cette approche dépasse, en terme de performance, la recherche classique dans PubMed.

Nous détaillons ces travaux dans ce qui suit.

5.2 Identification des éléments PICO

Un premier volet de travaux concerne l'identification des éléments PICO (P, I, C et O) à partir des documents et des requêtes. Cette étape est cruciale, et représente la phase en amont pour récupérer les documents pertinents. Les travaux dans (Boudin *et al.*, 2010a; Zhao *et al.*, 2010) se sont basés sur des techniques d'apprentissage supervisés afin d'identifier les éléments PICO du texte. Les auteurs de (Boudin *et al.*, 2010a) ont transformé les résumés de PubMed en un ensemble de phrases, qui sont ensuite représentées par des descripteurs basés sur les propriétés tout en se basant sur des attributs structurels comme le nombre de "*cue-verbs*" et de "*cue-words*". De plus, ils ont exploité les propriétés statistiques et linguistiques pour reconnaitre les éléments via les types sémantiques fournis par UMLS. Le processus d'identification des éléments PICO à partir du document a été conduit selon plusieurs étapes : tout d'abord, le texte des documents est segmenté en plusieurs phrases, ensuite chaque phrase est transformée en un vecteur de propriétés utilisant les caractéristiques statistiques et linguistiques préalablement définies pour désigner P, I, C et O. Chaque vecteur est soumis à un ou plusieurs classifieurs. Plusieurs algorithmes de classification de l'état de l'art ont été utilisés tels que *Random forest (decision trees)*, SVM (radial kernel of degree 3), *multi-layer perception* (MLP) et *Naive Bayes* (NB). Les expérimentations ont montré que les éléments I et C sont plus difficiles à identifier que l'élément P. De plus, cela a montré que la détection des éléments PICO au niveau des phrases et non terme ou concept est possible si cela se restreint à définir un ensemble de propriétés utilisées avec les techniques d'apprentissage.

Pour pallier ces restrictions, (Zhao *et al.*, 2010) ont proposé une classification basée sur différentes classes PICO qui prend en compte les différences de format structure et présentation des documents. Pour cela, ils ont proposé une classification au niveau phrase et au niveau

16. Mean Average Precision

terme afin d'identifier les éléments PICO. Puis ils assignent les phrases aux 5 classes PICO utilisant un classifieur binaire en se basant sur Maximum Entropy avec les caractéristiques suivantes : tokens (N-gram de la phrase), phrases (longueur de la phrase et sa position dans le paragraphe ou dans l'article), entités nommées (extraites de OpenNLP Package), MeSH (si la phrase contient un terme Mesh qui appartient à l'une des 16 catégories), Lexica (si la phrase contient un mot qui désigne le sexe ou l'age). Au niveau terme, ils ont procédé à une classification basée sur les mots utilisant des caractéristiques différentes : token (les termes), syntagmes nominaux (position du mot dans le syntagme et le nom qui constitue le noyau du syntagme), entités nommées (extraites de OpenNLP Package), MeSH et Lexica. Dans (Chung, 2009), les auteurs proposent d'utiliser une méthode de traitement du langage naturel pour catégoriser les phrases et les associer aux différents types sémantiques.

Les auteurs de (Boudin *et al.*, 2010d) ont proposé une approche automatique d'identification des éléments PICO à partir des requêtes et des documents biomédicaux. Ensuite, ils ont testé plusieurs modèles en se basant sur le modèle de langue afin de comparer l'utilisation de ces éléments dans la RI. Les termes de la requête qui correspondent aux éléments P, I, O sont pondérés grâce à une extension du modèle de langue, soit :

$$p_1(t|M_Q) = \gamma * \frac{count(t,Q)}{Q} * \left(1 + \sum_{E \in P,I,O} w_{Q,E} * \delta(Q_E, t)\right) \tag{2.1}$$

où :

— $w_{Q,E}$ est le poids de l'élément E dans la requête Q, dénoté Q_E

— $\delta(Q_E, t)$ est une fonction binaire :

$$\delta(Q_E, t) = \begin{cases} 1 \text{si} t \in Q_E \\ 0 \text{sinon} \end{cases} \tag{2.2}$$

— γ est un facteur de normalisation

— $count(t, Q)$ est la fréquence du terme t dans Q et Q est la longueur de la requête.

Le score du document vis-à-vis de la requête Q est donné par une combinaison linéaire d'interpolation des poids comme suit :

$$Score(Q, D) = score(Q_{all}, D) + \sum_{E \in P,I,O} w_{Q,E} * score(Q_E, D) \tag{2.3}$$

où $score(Q_X, D)$ est le score correspondant à une partie de la requête, c-à-d P, I, O ou le reste, qui est calculé par la somme des poids des termes figurant dans la requête.

Afin de tenir compte de l'importance des termes dans chaque élément PICO dans le document, (Boudin *et al.*, 2010d) ont proposé d'étendre le modèle de représentation du document D comme suit :

$$p_2(t|M_D) = \gamma * \left(p(t|M_{D_{all}}) + \sum_{E \in P,I,O} w_{D,E} * p(t|M_{D_E})\right) \tag{2.4}$$

ou :

— γ est un facteur de normalisation

— $p(t|M_{D_X})$ est la probabilité que le terme soit généré par le modèle de langue correspondant aux éléments E du document en utilisant la fonction de lissage de Dirichlet.

Enfin, le score final du document D vis-à-vis de la requête Q est recalculé par le produit des probabilités p_1 et p_2 comme suit :

$$score(Q, D) = \sum_{t \in Q} p_1(t|M_Q) * p_2(t|M_D) \tag{2.5}$$

Ces travaux ont montré l'efficacité de l'utilisation des éléments PICO dans la tâche de la RI. Par ailleurs, le problème principal de l'utilisation du paradigme PICO est non pas l'identification correcte de ces éléments, mais leur bonne exploitation dans les modèles de RI. Des études antérieures ont cherché à résoudre les problèmes de détection des éléments PICO et de développer des algorithmes plus applicables et des méthodes de classification efficaces (McKnight et Srinivasan, 2003). Chung a apporté des rôles rhétoriques, à savoir : But, Méthode, Résultats et Conclusion, en utilisant un cadre séquentiel avec une méthode de CRF pour étiqueter PICO au niveau de la phrase (Chung, 2009). Sur cette base, Boudin et al. ont construit des classificateurs avec de plus grandes collections de données (Boudin *et al.*, 2010a), puis ils ont ajouté la pondération de termes basée sur la localisation des termes dans le texte (Boudin *et al.*, 2010b). La plupart de ces travaux ont récupéré l'information à partir d'une portion spécifique de données, mais peu d'entre eux ont pris en compte la partie restante des données pour la tâche de classification. Dans (Demner-Fushman et Lin, 2007; Kim *et al.*, 2011), les auteurs ont rajouté une catégorie appelée "autre", où ils ont recueilli toutes les informations non ciblées. Dans (Kim *et al.*, 2011), les auteurs ont conclu que parmi les différents critères, les titres des sections aident à identifier la nature de la tâche.

Comme mentionné précédemment, (Boudin *et al.*, 2010b) ont comparé différents classificateurs complexes mettant l'accent sur les premières phrases. Ils ont affirmé que la majorité des informations importantes est contenue dans la première phrase de chaque section et que le fait d'annoter toutes les phrases dans les sections P/I/O ne peut pas être plus pragmatique. Néanmoins, il n'y a eu aucune preuve de l'applicabilité des ensembles de données de première phrase. Pour cette raison, les auteurs dans (Huang *et al.*, 2013) ont mené une étude pour déterminer si un classifieur correspondant à la première phrase d'une section annotée explicitement est efficace pour détecter les éléments PICO au niveau de la phrase dans le texte. En effet, ils ont proposé une approche automatique pour reconnaître les éléments PICO à partir du texte des études cliniques avec moins d'implication d'experts humains. Pour cela, ils ont extrait $19,854$ résumés structurés à partir des essais cliniques étiquetés selon les éléments P/I ou C de PubMed, utilisées dans les classifieurs Bayésiens comme suit :

Pour un ensemble d'éléments $f = (f_1, ..., f_n)$, la probabilité que chaque classe c appartient à une classe positive c_p ou une classe négative c_n est calculée comme suit :

$$P(c|f) = P(c) * \frac{P(f|c)}{P(f)}, c \in \{c_p, c_n\} \tag{2.6}$$

Selon l'hypothèse d'indépendance, $P(f|c)$ peut être décomposée en :

$$P(f|c) = \prod_{i=1}^{n} P(f_i|c) \tag{2.7}$$

Donc, la probabilité d'une classe c pour un ensemble de facettes donné $f_1, .., f_n$ devient :

$$P(c|f) = P(c)\frac{\prod_{i=1}^{n} P(f_i|c)}{P(f)} \tag{2.8}$$

Dans une étude plus récente (Chabou et Iglewski, 2015), les auteurs ont proposé une approche hybride en combinant la robustesse des méthodes d'apprentissage (*machine learning methods*) et le plus haut niveau de précision des méthodes basées sur les règles (*rule-based methods*) pour améliorer le processus d'extraction des éléments PICO et faciliter la validité et la pertinence des réponses aux questions cliniques formulées dans le cadre de la recherche PICO.

Leur méthode d'extraction des éléments PICO repose sur deux étapes :

1. La première étape : extraction des éléments PICO au niveau des phrases en utilisant les méthodes d'apprentissage avec l'algorithme CRF (*Conditional Random Fields*). Cette méthode assigne la séquence des étiquettes Y les plus probables, à une séquence d'observations X utilisation une probabilité conditionnelle $P(Y|X)$ calculée comme suit :

$$P(Y|X) = \frac{1}{z_x} exp\left(\sum_{f=1}^{T}\sum_{k} \lambda_k f_k(Y_{f=1}, Y_f, x, t)\right) \tag{2.9}$$

   Pour une phrase donnée X, cette équation estime la probabilité que cette phrase soit classée comme étant un élément PICO Y ; $Y = \{P, I, CouP\}$; λ_k est le poids associé à chaque propriété (feature) $f_k(Y_{f=1}, Y_f, x, t)$. L'algorithme CRF utilise avec un ensemble de propriétés, à savoir : propriétés sémantiques, propriétés lexico-syntaxiques, propriétés selon la position.

2. La deuxième étape : reconnaissance des termes afin d'exploiter la sémantique des mots dans les phrases déjà marqués pour les éléments PICO et utiliser cette sémantique comme un moyen d'obtenir la pertinence. Le texte clinique est traité selon les étapes suivantes :
   - Les phrases déjà annotées avec les éléments PICO par la méthode d'apprentissage sont utilisées dans cTAKES (*Clinical Text Analysis and Knowledge Extraction System*) qui est un outil open source pour l'extraction d'informations à partir du texte naturel clinique ;
   - cTAKES traite le texte clinique et extrait les parties sémantiques au niveau des termes en utilisant différents dictionnaires dont UMLS. Les parties médicales sémantiques extraites sont enregistrées dans des fichiers.
   - Les annotateurs basés sur les règles traitent ce fichier, et extraient les instances des éléments PICO à partir des phrases, chaque implémentation d'annotateur est basée sur un ensemble de règles produites à la main ;
   - La création des phrases annotées avec les éléments PICO au plus haut niveau de précision.

5.3 Modèles de RI basés sur les facettes PICO

Le deuxième volet sur les questions PICO concerne l'utilisation des facettes PICO dans la mise en oeuvre de modèles de RI capables d'exploiter les facettes et de les incorporer dans le calcul de scores de pertinence des documents (Demner-Fushman et Lin, 2007; Boudin *et al.*, 2010c; Boudin *et al.*, 2010b).

Dans (Demner-Fushman et Lin, 2007), les auteurs ont proposé une méthode qui utilise soit un ensemble de règles de filtrage par motif définies manuellement ou une combinaison de classifieurs basiques afin de détecter les éléments PICO à partir des résumés médicaux. Au préalable, les auteurs ont utilisé MetaMap (Aronson, 2001) pour annoter les concepts biomédicaux dans les résumés et SemRep (Arnold et Rahm, 2015) pour extraire les relations entre ces concepts. Les deux outils sont basés sur UMLS (Unified Medical Language System). Le modèle propose de détecter les éléments PICO et de les utiliser dans une fonction de calcul de pertinence S_{EBM} calculée comme suit :

$$S_{EBM} = S_{PICO} + S_{SoE} + S_{task} \tag{2.10}$$

Ce score est basé sur une combinaison linéaire de scores de pertinence partiels calculés pour les documents vis-à-vis de chaque facette PICO S_{PICO}, la force des preuves fournies S_{SoE}, qui est un schéma de classification qui aide les médecins à évaluer les citations pour des objectifs cliniques et le type de la tâche S_{Task}.
Pour chacun des éléments P, I, C, et O, ils ont proposé un extracteur spécifique à chaque facette qui permet de les identifier à partir des résumés, soit :

$$S_{PICO} = S_{problem} + S_{population} + S_{intervention} + S_{outcome} \tag{2.11}$$

— Pour **Population** : elle représente en général les patients. Pour les identifier, les auteurs ont utilisé un ensemble de motifs pour les extraire du texte, en se basant sur un ensemble d'hypothèses, à savoir :
  — les concepts qui décrivent "population" appartiennent au type sémantique "Group" ou à l'un de ses descendants. En plus, certains noms sont utilisés pour décrire les participants des études dans les textes médicaux ;
  — le nombre de sujets qui participent à une étude sont souvent précédés ou suivis par un concept de type "Group" ;
  — une clause qui contient la population est influencée par sa position, selon les résumés structurés ou non structurés.

  Le score de confiance est donc attribué à chaque motif, dépendant de sa position dans le résumé et de sa position dans la clause de laquelle il a été extrait.

— Pour **Problem** : cette facette correspond aux concepts UMLS de type sémantique "*Disorder*". Afin d'attribuer les scores de confiance, ils se sont intéressés aux concepts dans les titres des résumés, avec les deux premières phrases et enfin dans tout le résumé.

— Pour **Intervention/Comparison** : les auteurs ont choisi de traiter l'intervention et la comparaison ensemble vu leur appartenance au même groupe sémantique "*Exposures*". L'algorithme d'extraction de ces deux éléments traite d'abord les entités qui peuvent participer aux relations du réseau sémantique UMLS associés à chaque tâche clinique.

Ensuite, selon les types sémantiques identifiés comme *diagnostic procedure*, *clinical drugs*, *health care activity*,etc. Les scores sont ajustés selon si les concepts apparaissent dans les sections titre, objectif ou encore méthodes. Pour ces concepts, un poids supplémentaire est ajouté vu leur importance. Pour les résumés non structurés, les concepts des résumés sont favorisés. D'une autre part, l'algorithme prend en compte certaines expressions qui décrivent l'étude comme "cette étude examine.." ou "Ce papier décrit..".

— Pour **Outcome** : les auteurs considèrent la tâche d'identification et d'extraction des résultats comme un problème de classification au niveau des phrases. Ils utilisent une combinaison de classifieurs et calculent un score d'interpolation linéaire des poids, soit :

$$S_{outcome} = \lambda_1 S_{cues} + \lambda_2 S_{unigram} + \lambda_3 S_{n-gram} + \lambda_4 S_{position} + \lambda_5 S_{length} + \lambda_6 S_{semantictype} \quad (2.12)$$

Le score de pertinence basé sur la force de la preuve est calculé comme suit :

$$S_{SoE} = S_{journal} + S_{study} + S_{date} \quad (2.13)$$

$S_{journal}$ est calculé selon le type du journal dans lequel la citation est publiée, elle a un score de 0.6 si le journal est important et 0 sinon. Pour S_{study} les essais cliniques reçoivent 0.5, les études d'observation comme par exemple des rapports d'études de cas 0.3, les publications non cliniques -1.5 et 0 sinon.

$$S_{date} = \frac{(year_{publication} - year_{current})}{100} \quad (2.14)$$

$$S_{task} = \sum_{t \in MeSH} \alpha(t) \quad (2.15)$$

où $\alpha(t)$ attribue au terme MeSH un score positif si le terme représente un indicateur positif pour le type de la tâche, ou à un score négatif si le terme représente un indicateur négatif pour le type de la tâche.

Cette méthode a obtenu des résultats intéressants avec un taux de justesse de 80% pour l'identification de population et intervention, 86% pour problème et entre 68% et 95% pour outcome. Les expérimentations ont été effectuées sur 24 questions cliniques avec un corpus de test modeste (143 résumés pour l'élément outcome et 100 résumés pour les autres éléments), et ont montré que la performance de cette approche dépasse les méthodes traditionnelles de recherche PubMed.

Dans (Boudin *et al.*, 2010c), les auteurs ont analysé la distribution des éléments PECO dans les documents pertinents, et ont proposé ensuite une approche de RI basée sur le modèle de langue. Le modèle de langue assume que les requêtes et les documents sont générés de la même distribution de probabilité du texte (Ponte et Croft, 1998).

L'approche proposée se base sur la distribution des éléments PECO, utilisée pour la phase de pondération. Dans le modèle proposé, les auteurs ont intégré la distribution des éléments PECO observés dans les documents afin d'identifier les éléments les plus informatifs des documents. L'idée est de résoudre en amont le problème de la détection des éléments PECO en utilisant un modèle de langue basé sur la position. Pour intégrer la position, les auteurs estiment une série de probabilités concernant le nombre de mots dans une partie spécifique

du document au lieu du document entier. Chaque document D est classé selon une interpolation linéaire pondérée. Pour un document donné D, divisé en 10 parties $p \in [P1, P2, ..., P10]$, $P(w|D)$ est estimée comme suit :

$$P(w|D) = \alpha * P(w|D) + \beta * P_{title}(w|D) + \gamma * \sum_{p_i \in D} \theta_e * P_{p_i}(w|D) \tag{2.16}$$

ou θ_e est le poids de chaque élément PECO, fixé d'une manière empirique selon les valeurs de la distribution des éléments observés dans les documents. Il proposent ensuite une fonction de calcul en intégrant les éléments PECO au modèle de langue de base. L'idée était d'utiliser la structure PECO pour identifier l'importance de chacun des éléments d'une manière équilibré. La fonction de calcul finale est définie comme suit :

$$Score_{final}(Q, D) = \sum_{e \in PECO} \delta_e * score(Q_e, D) \tag{2.17}$$

Les paramètres de pondération du modèle ont été déterminés par validation croisée, sur une collection de 1.5 millions de documents MEDLINE et 423 questions cliniques. Cette approche a montré une amélioration de performance de la RI de 28% de la MAP et 50% par rapport aux modèles de langues classiques.

Le Tableau 2.14 donne une classification de quelques travaux qui se basent sur des éléments PICO ainsi que d'autres propriétés comme la situation ou encore le type de l'étude.

	Patient/Population	Intervention	Comparaison	Outcome	TimeFrame	Context	Type of question	Type of study design	Professionals	Health Care Setting	Exposure	Duration	Results	Environment	Stakeholders	Situation
(Richardson *et al.*, 1995; Boudin *et al.*, 2010a; Boudin *et al.*, 2010d)	■	■	■	■												
(Huang *et al.*, 2013)	■	■		■												
(Fineout-Overholt et Johnston, 2005)	■	■	■	■	■											
(Petticrew et Roberts, 2006)	■	■	■	■		■										
(Schardt *et al.*, 2007)	■	■	■	■			■	■								
(Attia, 2013)	■	■		■					■	■						
(Dawes *et al.*, 2007)	■		■	■							■	■	■			
(Schlosser et O'Neil-Pirozzi, 2006)	■	■	■	■										■	■	

Tableau 2.14 – Classification de travaux selon l'utilisation des éléments cliniques

6 Campagnes d'évaluation pour la RI biomédicale

Il existe à ce jour deux campagnes d'évaluation en RI proposant des tâches dédiées à l'évaluation de la RI biomédicale : CLEF et TREC. Les sections suivantes présentent des éléments descriptifs de ces tâches.

6.1 Campagne d'évaluation TREC

Dans le cadre du domaine médical, on s'intéresse à deux pistes : (1) TRECMed pour la RI des dossiers médicaux des patients et (2) TRECGenomics pour la RI de la littérature biomédicale.

— *TRECMed* : TRECMed est destinée pour la RI biomédicale des comptes-rendus médicaux de patients. La piste TRECMed a pour but de favoriser la recherche et la récupération de documents liés aux dossiers médicaux de patients en fournissant l'accès basé sur le contenu des champs de dossiers médicaux électroniques. Introduite pour la première fois dans TREC en 2011, TRECMed représente un grand effort de la communauté de la RI biomédicale, fournissant une large collection de documents de test dans le cadre de l'évaluation de dossiers médicaux de patients.
La tâche principale de TRECMed est liée à la recherche ad-hoc des patients qui correspondent au besoin de l'utilisateur exprimé par une requête. La collection de test est caractérisée par un ensemble de dossiers médicaux, de rapports de comptes rendus qui sont associés à une ou plusieurs visites. La collection est fournie aux participants via l'accord avec l'université de Pittsburgh. Le Tableau 2.15 donne quelques statistiques sur les collections TRECMed2011 et 2012.

Année	Nb. de requêtes	Nb. de documents	Nb. documents pertinents
2011	35	95,701	1765
2012	50	95,701	58,640

Tableau 2.15 – Statistiques des collections TREC Medical

— *TRECGenomics* : La piste TREC Genomics destinée à la RI de la littérature biomédicale a duré de 2003 à 2007 et représente une des pistes de recherche les plus importantes dans le domaine médical. Lancé pour la première fois en 2003, pour répondre aux défis de gestion et de récupération de la littérature médicale afin d'identifier les interactions potentielles entre les gènes, les maladies et d'autres entités biologiques (Radhouani *et al.*, 2009). Cette tâche modélise un cadre où un chercheur en génomique entrant dans un nouveau secteur exprime une requête qu'il soumet à un moteur de recherche qui gère la littérature scientifique biomédicale, qui est la collection MEDLINE. Le Tableau 2.16 donne quelques statistiques sur les collections de 2003 à 2007. Les requêtes *TREC Genomics* ont évolué à travers les années : en 2003 les noms de gènes (eg. "*arginine vasopressin*"), le besoin en information exprimé par les nouveaux chercheurs en génomique en 2005 (eg. "*provide information about the role of the gene PRNP in the disease Mad Cow Disease*"), question-réponse dans le domaine biomédical en 2007 (*eg.* "*What is the role of gene gamma-aminobutyric acid receptors (GABABRs) in the process of inhibitory synaptic transmission ?*").

Année	Nb. de requêtes	Nb. de documents	Nb. documents pertinents
2003	50	525,938	566
2005	50	4,591,008	4584
2007	36	162,259	2001

Tableau 2.16 – Statistiques des collections TREC Genomics

La piste de TREC Genomics (http ://ir.ohsu.edu/genomics/) est l'une des campagnes les plus importantes et longues dans la biomédecine, dirigée par (Hersh *et al.*, 2006a).

Les différentes tâches sont données dans le Tableau 2.17.

Travaux	Tâche	Collection de documents	Topics
(Hersh et Bhupatiraju, 2003)	recherche adhoc	525,938 dossiers MEDLINE	noms de gènes, avec l'objectif de trouver toutes les références MEDLINE qui concernent la biologie des gènes ou leur production de protéine à partir de l'organisme concerné
(Hersh et Bhupatiraju, 2003)	GeneRRIF (ou *Gene Reference Into Function utilisé dans la banque de données LocusLink*[17]) annotation des titres d'articles et résumés	139 articles assignés à GeneRIFs	GeneRIFs concernés
(Hersh *et al.*, 2004)	recherche adhoc	4,591,008 documents MEDLINE	50 besoins en information avec les titres, le besoin et le contexte
(Hersh *et al.*, 2004)	catégorisation des documents contenant des données sur les gènes, assigner des annotations des code Gene Ontology pour la base Mouse Genome Informatics	11,880 articles de journaux complets	50 besoins en information avec les titres, le besoin et le contexte
(Hersh *et al.*, 2005)	recherche adhoc	4,591,008 documents MEDLINE	50 besoins en information similaire à celle de 2004 mais classés selon 5 Generic Topic Types (GTTs)
(Hersh *et al.*, 2005)	catégorisation des documents contenant des données sur les gènes, assigner des codes Gene onto pour identifier les tumeurs biologiques, expressions sur les gènes, allèles mutant pour la base Mouse Genome Informatics	11,880 articles de journaux complets	50 besoins en information similaire à celle de 2004 mais classés selon 5 Generic Topic Types (GTTs)
(Hersh *et al.*, 2006b)	recherche des passages (des parties de phrases et des paragraphes) en lien avec cinq entités (exemple : gènes, protéines) et la source de l'article	une collection de 162,259 documents HTML issus de 49 journal	28 questions basées sur les GTTs
(Hersh *et al.*, 2007)	question-réponse basée sur les entités pour rechercher les passages liés au 14 entités et à la source de l'article	162,259 documents HTML issus de 49 journal	36 questions basées sur 14 entités

Tableau 2.17 – Tâches de la campagne TREC Genomics (Hersh, 2009)

— *TREC Clinical Decision Support Track* : La piste de la prise de décision clinique, proposée par (Simpson *et al.*, 2014), étudie les techniques pour lier les cas médicaux aux informations pertinentes concernant les soins aux patients.

Le but de cette piste d'aide à la décision clinique est de simuler les exigences de ces systèmes et d'encourager la création d'outils et les ressources nécessaires pour leur mise en œuvre. Une description des collections de 2014 et 2015 est donnée par le Tableau 2.18. Les "topics" de cette piste sont crées par des experts médicaux et représentent des cas de patients. Ces cas décrivent des informations telles que les antécédents médicaux d'un patient, les symptômes actuels du patient, les tests effectués par un médecin pour le diagnostic de l'état du patient, le diagnostic final du patient, et enfin, les mesures prises par un médecin pour traiter le patient. Afin de simuler l'information réelle au besoin

Année	Nb. de requêtes	Nb. de documents
2014	30	733,138 articles PubMed
2015	30	733,138 articles PubMed

Tableau 2.18 – Statistiques sur la piste TREC de prise de décision médicale

des médecins, les sujets sont annotés selon les trois types de questions cliniques les plus génériques selon (Ely *et al.*, 2000) figurant dans le Tableau 2.19 ci-dessous. Les participants seront chargés de la récupération des articles biomédicaux utiles pour répondre aux questions génériques du type spécifié sur chaque rapport de cas.

Type	*Questions cliniques génériques*	Nombre de requêtes
Diagnostic	*What is the patient's diagnosis ?*	10
Test	*What tests should the patient receive ?*	10
Traitement	*How should the patient be treated ?*	10

Tableau 2.19 – Questions cliniques utilisées dans les campagnes 2014 et 2015

6.2 Campagne d'évaluation ImageCLEF

CLEF est en évolution permanente et couvre différentes tâches de recherche d'images. Parmi plusieurs pistes d'évaluation, la piste *ImageCLEF* apparait pour la première fois en 2003, et a pour principal objectif d'encourager l'évolution dans différents domaines comme l'analyse des médias visuels, l'indexation, la classification et la RI. Elle fournit pour cela une infrastructure pour l'évaluation des SRI que ce soit par le contexte ou par le contenu. Cela a impliqué la définition de diverses sous-tâches qui varient d'une année à une autre. Par exemple, en 2012 les quatre sous-tâches définies dans *ImageCLEF 2012* sont : recherche de cas de patients, recherche d'images et classification de modalités.

La tâche de recherche de cas de patients a été introduite pour la première fois en 2009, et a pour but de rechercher les cas de patients incluant des images qui répondent potentiellement à un cas clinique. Contrairement à la tâche ad-hoc, l'unité de recherche est le cas de patient et non l'image. Pour cet objectif, un cas est un identifiant PubMed "ID" qui correspond à un article de journal. Les documents sont constitués du texte intégral y compris les légendes des photos biomédicales. Les requêtes sont constituées d'une description de cas de patients, avec la démographie, les symptômes et les résultats de tests y compris les études des imageries. Quelques statistiques sur les collections *ImagesCLEF* (*case-based retrieval*) sont présentées dans le Tableau 2.20.

La tâche Image-CLEF médicale lancée en 2004 par (Müller *et al.*, 2007) se base sur l'utilisation de l'information multimodale (images et texte) dans le domaine médical. Depuis 2009, la tâche de la recherche des cas de patients (*case-based retrieval*) a été introduite, avec pour objectif de promouvoir la recherche des diagnostics. Le but principal de cette tâche est de récupérer les cas de patients incluant des images que les médecins jugent pertinentes ou pas par rapport aux différents diagnostics. En gros, l'objectif de cette tâche est de fournir au clinicien qui a formulé la requête une aide afin de lui permettre de prendre une décision pertinente sur un diagnostic ou un traitement considérant un cas difficile. Les requêtes sont créées à partir

d'une base médicale de cas incluant des descriptions en langage naturel des cas médicaux de patients, avec la démographie des patients, les symptômes, les résultats des tests et les études des images. Par exemple une requête peut être : "*A 49-year-old woman with a prolapsed mass in the opening of her urethra. Pelvic CT shows a heterogeneously enhanced mass on the female urethra. Pathology shows ramifying papillae, high nuclear/cytoplasmic ratio, and brisk mitotic activity*".

Année	Nb. de requêtes	Nb. de documents	Nb. documents pertinents
2009	5	5706	95
2010	14	77,506	95
2011	10	55,635	521
2012	26	74,654	247
2013	35	74,654	709

Tableau 2.20 – Statistiques des collections *ImageCLEF* (*Case-based IR*)

Les tâches de la campagne ImageCLEF sont différentes d'une année à l'autre. Par exemple, pour l'année 2011, plusieurs tâches ont été proposées, nous citons :

— La recherche médicale : elle comprend des tâches de classification d'images et de textes ainsi que la recherche adhoc des images et des cas de patients (Kalpathy-Cramer *et al.*, 2011).

— L'annotation d'images : c'est une tâche d'annotation de de détection visuelle des concepts. Elle de divise en deux sous-tâches : (1) l'annotation, et (2) la recherche basée sur les concepts.

— L'identification de plantes : l'objectif principal de cette tâche est d'associer les espèces de plantes à l'image correspondante.

— La recherche Wikipédia : l'objectif est d'étudier les approches de RI dans le cadre d'une grande collection hétérogène d'images et de textes issus des besoins en information exprimés par les utilisateurs.

Un des défis les plus connus dans le domaine de la RI médicale est de répondre aux questions des patients (Burstein *et al.*, 2005) dans un format compréhensible par les utilisateurs néophytes. Les prescriptions médicales et les résumés de décharge sont écrits dans la terminologie médicale professionnelle qui n'a aucun sens pour l'utilisateur final (patient). Pour cette raison, la communauté ShARe/CLEF a proposé la tâche eHealth en 2013 (Suominen *et al.*, 2013) avec l'objectif de développer un tel système en attirant les jeunes chercheurs de diverses organisations et universités du monde de l'informatique et du domaine biologique ; et présente une plate-forme commune pour mener la recherche. De plus, le but est d'évaluer les systèmes qui assistent les utilisateurs néophytes dans leur recherche et la compréhension de l'information médicale.

La campagne ShARE/CLEF eHealth 2014 (Kelly *et al.*, 2014) comprend trois tâches principales, à savoir :

— *Tâche 1 - Visualisation de l'information* : la première tâche de CLEF eHealth 2014 (Kelly *et al.*, 2014) vise à fournir une visualisation de l'information médicale extraite des résumés de décharge de manière convenable et compréhensible pour les néophytes du domaine médical ;

— *Tâche 2 - Extraction de l'information* : cette tâche nécessite la normalisation et l'expansion des abréviations et des acronymes présents dans les résumés de décharge. Le cas d'utilisation suppose que, compte tenu des résumés de décharge et les troubles diagnostiqués, les patients ont souvent des questions au sujet de leur état de santé (Zhu *et al.*, 2013) ;
— *Tâche 3 - Recherche d'information* : l'objectif de la troisième tâche (Goeuriot *et al.*, 2014; Goeuriot *et al.*, 2013) est de fournir des documents utiles et pertinents pour les patients, en développant un système de recherche médical centré sur l'utilisateur ou basé sur le contexte de manière à satisfaire leurs besoins en informations médicales.

Avec l'évolution dans ce domaine, eHealth 2015 (Goeuriot *et al.*, 2015) s'est focalisé sur les cliniciens qui ont des problèmes à comprendre le jargon médical. Cette campagne propose deux tâches, à savoir : (1) extraction de l'information à partir du texte clinique ; et (2) la RI centrée sur l'utilisateur. Le Tableau 2.21 donne des statistiques sur les différentes collections de eHealth.

Année	Nb. de requêtes	Nb. de documents	Nb. documents pertinents
2013	50	1 million	37
2014	50	1 million	64
2015	66	1 million	66

Tableau 2.21 – Statistiques des collections CLEF eHealth.

7 Aperçu de systèmes de RI médicaux

Les systèmes informatiques médicaux ont connu une grande évolution depuis ces deux dernières décennies tant du point de vue de leur architecture que de la qualité et de la diversité des services autour du stockage de l'information, l'accès à l'information pertinente pour une médecine basée sur des niveaux de preuves, l'aide à la décision pour l'amélioration de la qualité des soins (Hersh, 2006). De ce fait, plusieurs moteurs de recherche ont été proposés, nous détaillons dans ce qui suit les quatre portails les plus utilisés dans le domaine médical : *PubMed*, *EMERSE*, *CisMef* et *Doctissimo*.

7.1 PubMed

PubMed est le moteur de recherche principal de données bibliographiques de l'ensemble des domaines de spécialisation de la biologie et de la médecine. Il a été développé par le centre américain pour les informations bio-technologiques (NCBI), et est hébergé par la bibliothèque américaine de médecine des instituts américains de la santé. *PubMed* est un moteur de recherche gratuit donnant accès à la base de données bibliographique MEDLINE, rassemblant des citations et des résumés d'articles de recherche biomédicaux. *PubMed* contient plus de 24 millions de citations et résumés de journaux biomédicaux et de santé. Les professionnels de santé considèrent *PubMed* comme une des plus importantes sources de preuves pour les soins de santé (Haux *et al.*, 1996). PubMed joue un rôle important dans le processus de découverte basée sur la littérature (Baker et Hemminger, 2010).

Le système *PubMed* inclut plusieurs fonctionnalités comme des champs d'exploration et de recherche, les termes MeSH (utilisés pour indexer les articles *PubMed*), des raccourcis et

des fonctionnalités utilisant l'historique de la recherche. Les données bibliographiques sont enregistrées dans des bases de données structurées concernant 65 domaines incluant les titres, résumés, auteurs, journaux, types de publication et dates de publication. *PubMed* fournit 48 tags de recherche dans les divers domaines de la base ; une description de chaque domaine est disponible sur le site *NLM*[18]. PubMed est donc un système de recherche orienté domaine avec les différents tags et annexes utilisant les opérateurs booléens (AND, OR, NOT). NLM indexe les documents PubMed en utilisant le vocabulaire *MeSH*. La Figure 2.5 présente une capture d'écran du système PubMed.

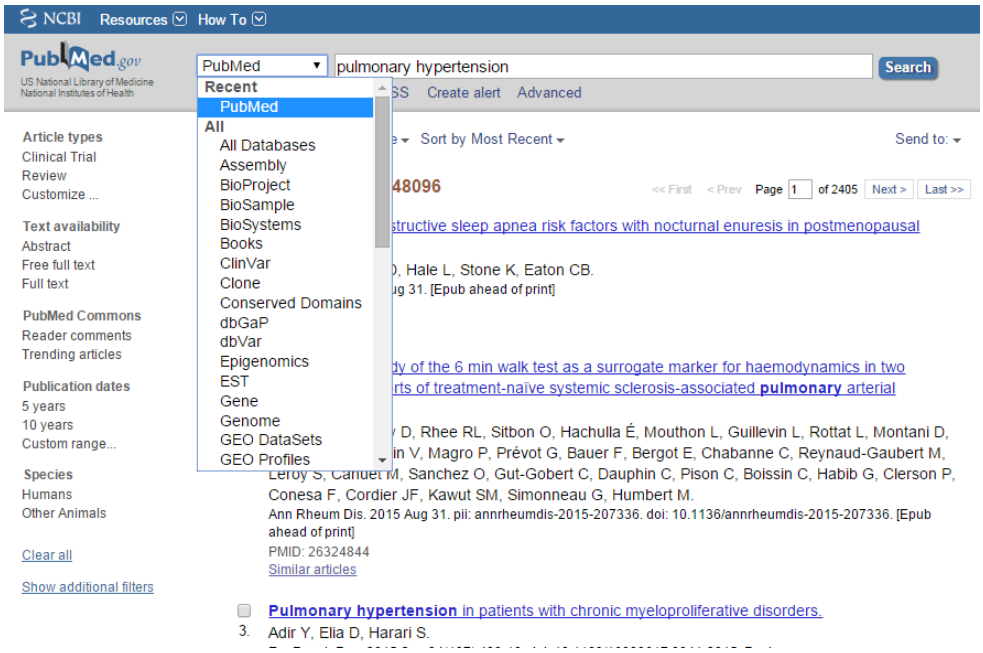


FIGURE 2.5 – Aperçu de domaines du système PubMed.

7.2 EMERSE

EMERSE (The Electronic Medical Record Search Engine) (Hanauer, 2006) est un moteur de recherche puissant et intuitif, pour la recherche de documents dans les dossiers électroniques médicaux (Hanauer, 2006). Il offre de multiples options pour créer des requêtes complexes pour réaliser la tâche de RI. Le moteur de recherche est idéal pour un examen rétrospectif des dossiers et revues ainsi que l'abstraction de données ; il peut également être utile pour les traitements cliniques. *EMERSE* concerne diverses tâches incluant la recherche clinique, l'amélioration de la qualité des initiatives ainsi que les tâches opérationnelles des hôpitaux. Le système a été utilisé par les agents de conformité pour gérer les risques et par différents établissements de prévention des infections. Il offre plusieurs fonctionnalités permettant un travail rapide et à jour. Par exemple, il comprend une liste large de synonymes et de mots clés pour aider à étendre la recherche si la personne ne connait pas les termes appropriés à utiliser. Cela inclut une large collection de noms de gènes, de médicaments, d'acronymes, d'abréviations et d'autres variations de mots prenant en compte même les fautes d'orthographe. La Figure 2.6 montre une capture d'écran du système *EMERSE* avec les termes d'expansion.

18. http ://www.nlm.nih.gov/bsd/mms/medlineelements.html

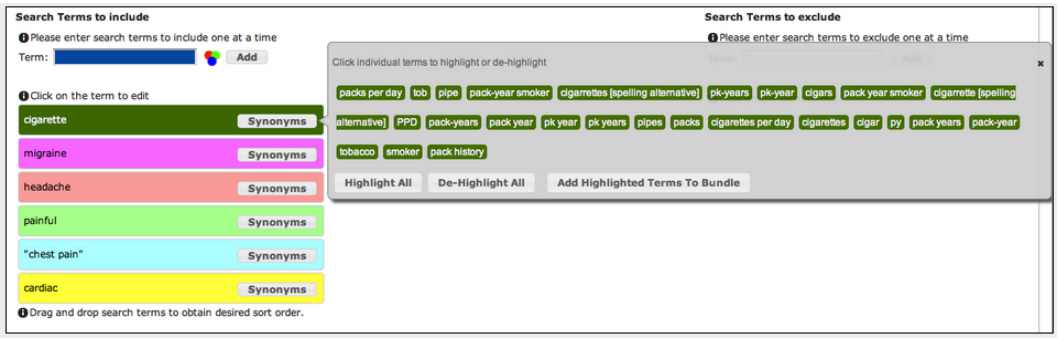


FIGURE 2.6 – Capture d'écran de EMERSE montrant les termes d'expansion du mot "cigarette".

7.3 CisMef

Le Catalogue et Index des Sites Médicaux Francophones (Darmoni et Joubert, 2000) du CHU de Rouen propose l'essentiel de la documentation biomédicale francophone en ligne (Figure 2.7). Il recense plus de 25248 documents (au 07/02/07) et sa mise à jour est hebdomadaire. Les documents sont classés selon le thésaurus MeSH. Il s'organise autour de 3 axes prioritaires :

— Ressources pour l'enseignement : cours, guides, QCM, cas cliniques, lecture critique d'articles, etc.

— Médecine factuelle (fondée sur des données validées) : recommandations de bonne pratique, conférences de consensus, méta-analyses, etc.

— Documents destinés aux patients et au grand public, pour participer à l'amélioration de l'éducation sanitaire.

FIGURE 2.7 – Aperçu du système CisMef

CISMeF est l'outil de choix pour une recherche sur un sujet médical général, car son corpus est limité aux sites institutionnels, universitaires et aux sociétés savantes. La recherche s'effectue en deux étapes : (1) Sélectionner l'axe correspondant au type de recherche, et (2) utiliser la fonction "Terminologie", le descripteur MeSH en langue française, pour la sélection des mots-clés et des limites de la recherche. CISMeF permet également de trier les ressources par

type de documents et d'étendre la recherche (vers d'autres mots-clés ou d'autres moteurs de recherche).

7.4 Doctissimo

Doctissimo[19] est un site Web francophone consacré à la santé et au bien-être. Le site comprend une encyclopédie médicale touchant à plusieurs domaines de santé, comme la grossesse, la forme, la nutrition, les médicaments, etc. Il propose également des articles sur différents phénomènes afin de répondre aux questions des utilisateurs sur un sujet de santé bien précis, par exemple pour arrêter de fumer ou encore les moyens pour perdre du poids.

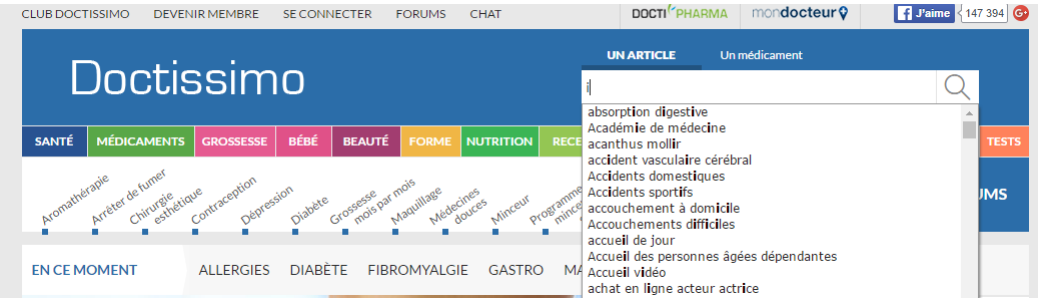


FIGURE 2.8 – Aperçu du site web Doctissimo.

Le site internet comprend des forums de discussions ce qui le rend interactif et permet d'être à jour sur les actualités et les questions du moment. La Figure 2.8 donne un aperçu des différents thèmes d'articles recherchés sur le site Doctissimo.

7.5 iMed

iMed est un moteur de recherche intelligent pour la recherche d'information médicale sur le web (Luo et Tang, 2008). C'est le premier moteur de recherche intelligent destiné à faciliter la tâche de la RI médicale, et qui utilise d'une manière extensive les connaissances médicales ainsi que les questionnaires interactifs. Ces derniers sont utilisés par le système pour guider les utilisateurs dans la formulation de leurs requêtes. De plus, iMed exploite les connaissances médicales à partir des réponses aux questions, pour constituer d'une manière automatiques un ensemble de requêtes susceptibles d'intéresser l'utilisateur. Dans (Luo, 2009), l'auteur a donné une évaluation du système en mettant en avant les avantages et les inconvénients, et a montré que *iMed* améliore considérablement la performance de recherche par rapport à d'autres systèmes comme *Healthline* et *Google Health*.

7.6 MedSearch

Proposé par (Luo *et al.*, 2008), *MedSearch*[20] est un moteur de recherche médical sur le web mis en place dans le but d'aider les utilisateurs à mieux formuler leur requêtes et à avoir des réponses pertinentes. Le système est basé sur un ensemble de techniques pour améliorer la qualité des résultats retournés, à savoir : (1) il accepte les requêtes longues des utilisateurs et

19. http ://www.doctissimo.fr/
20. http ://www.intelligence.tuc.gr/medsearch/

les transforme en requêtes courtes représentées par un ensemble de mots clés les plus représentatifs, et (2) il offre divers résultats de recherche. *MedSearch* est un système de recherche complet utilisant la base MEDLINE, et différentes méthodes comme SSRM et VSM.

8 Conclusion

Ce chapitre donne un aperçu de l'état de l'art de la RI biomédicale. Nous avons donc présenté une typologie de l'information biomédicale, y compris la littérature biomédicale et les dossiers médicaux personnels. La littérature biomédicale a fait l'objet de nouveaux travaux de recherche en RI biomédicale, en particulier dans le cadre de TREC Genomics 2003-2007.

Nous avons donc décrit les principales ressources termino-ontologiques les plus utilisées dans le domaine. Les typologies des requêtes médicales sont présentées dans le contexte de la recherche d'information médicale. Nous avons présenté dans ce cadre les différents travaux de catégorisation de requêtes médicales.

Dans le cadre de la RI biomédicale, nous avons présenté en catégorisant les différentes approches existantes, à savoir l'approche basée sur l'expansion des requêtes, l'approche basée sur la reformulation de la requête PRF ainsi que l'approche basée sur l'expansion documentaire automatique ou manuelle. En particulier, nous avons présenté les différentes approches de la RI biomédicale, à savoir : l'expansion conceptuelle de la requête, l'expansion de la requête par la méthode PRF ainsi que l'expansion documentaire.
De plus, les nouveaux modèles d'appariement basés sur les patrons de besoins cliniques, appelés modèle PICO, peuvent être utiles afin de mieux cerner les besoins en information spécifiques aux professionnels de la santé. En effet, le modèle PICO représente plusieurs challenges dans le domaine du point de vue de l' identification de ces éléments à partir des requêtes et des documents ainsi que leur utilisation dans la RI afin d'améliorer les systèmes de prise de décision médicale.

Enfin, nous avons présenté les campagnes d'évaluation spécifiques au domaine médical, comme TRECMed et TRECGenomics. Ces campagnes offrent des collections de test spécifiques au domaine en évoluant à travers les années afin d'évaluer au mieux les travaux proposés. Pour clôturer ce chapitre, nous avons donné un aperçu des systèmes de RI médicaux les plus utilisés dans le domaine et qui offrent une plateforme et interface de recherche avancées.

Nous allons présenter dans ce qui suit la deuxième partie du manuscrit qui comprend les deux chapitres de nos contributions dans le cadre de la RI médicale.

Deuxième partie

Analyse et évaluation de requêtes médicales

Introduction

La communauté des utilisateurs d'informations médicales est extrêmement variée selon plusieurs critères tels que, sa facon d'exploiter l'expertise dans la recherche, sa familiarité avec les différents vocabulaires d'indexation biomédicale et ses exigences en matière d'utilisation de l'information. Par exemple, l'expertise biomédicale varie entre les patients et les familles qui sont confrontés à des notions et des termes médicaux pour la première fois, à des spécialistes dans des domaines de recherche ciblés qui sont considérés comme des experts. Dans le domaine biomédical, l'analyse des besoins en information des utilisateurs de SRI biomédicaux ont été abordés sous l'angle de l'étude des stratégies de recherche d'une part et de l'analyse des requêtes d'autre part. De nombreux travaux ont cependant révélé des stratégies typiques des utilisateurs de SRI médicaux dépendantes de leur niveau d'expertise (Bhavnani, 2002). Dans le cadre de la RI sur le web les auteurs de (Eysenbach et Köhler, 2002) ont montré que les groupes de personnes qui cherchent l'information pour évaluer la crédibilité d'un site web, regardent la source, la conception, la langue utilisée et la facilité d'utilisation. Par ailleurs, les auteurs de (Haynes *et al.*, 2005) ont analysé des critères associés aux termes des requêtes cliniques ; ils ont identifié 4 critères stratégiques : la sensibilité, la spécificité, la précision et la justesse des termes.

D'une manière globale, l'expression des requêtes traduisant les besoins en information spécifiques est une tâche difficile aussi bien pour des novices que pour des experts (Spink *et al.*, 2004; White et Moris, 2008). En effet, avec la diversité des ressources disponibles en ligne ainsi que la multiplicité des SRI médicaux, il est difficile de trouver l'information pertinente. Cela est dû à deux problèmes majeurs : d'une part, la difficulté de formuler une requête claire, et l'interprétation des besoins en information induits par les SRI, d'une autre part. En effet, les utilisateurs se caractérisent par une variabilité significative à la fois sur le volet de l'âge que sur le niveau d'expertise, les besoins en information liés à la santé sont également très variés, couvrant des besoins liés à la médication, santé et bien-être, traitements et pathologies.

Plus proche de notre domaine d'étude, la tâche de recherche conduite par les experts a été étudiée dans le contexte de l'analyse du besoin exprimé par la formulation de requête. Cette analyse concerne les caractéristiques des requêtes lors de la formulation du besoin en information, comme le nombre de termes, le nombre de concepts, le vocabulaire utilisé, etc. D'une autre part, cela comprend le comportement de recherche comme le facteur temps, la navigation, la re-formulation de requête, etc.
Concernant la forme des requêtes, trois principales conclusions émergent des études antérieures, à savoir : (1) les requêtes médicales sont courtes (Hong *et al.*, 2002; Zeng *et al.*, 2002b; Natarajan *et al.*, 2010; Tracy Edinger *et al.*, 2012), (2) les termes de la requête ne sont pas fortement liés aux vocabulaires médicaux (Yang *et al.*, 2011; McCray et Tse, 2003), et (3) les sujets des requêtes sont ambigus (Song *et al.*, 2010; Liu *et al.*, 2012).

D'après notre état de l'art du domaine, peu de travaux se sont intéressés à des analyses spécifiques à la caractérisation des requêtes des experts médicaux, considérant une population spécifique d'utilisateurs ainsi que le contexte de la tâche de RI. Dans ce contexte, nous nous intéressons à l'analyse des besoins en information exprimés par les experts médicaux, nous traitons un axe de recherche dans le chapitre 3 qui concerne les *analyses exploratoires des requêtes expertes via l'analyse des attributs de requêtes*. Nous formalisons un ensemble d'attributs de requêtes, à savoir : la longueur en nombre de termes, la longueur en nombre de concepts, la spécificité terme-document, la spécificité hiérarchique, la clarté de la requête basée sur la pertinence et la clarté basée sur le sujet de la requête. Ces attributs sont utilisés dans

le cadre d'une méthodologie statistique dans le but de caractériser le besoin en information des experts et de mesurer l'impact de leur structure sur les résultats de recherche.

Un deuxième volet de contribution concerne la RI dans le cadre de la médecine basée sur les faits et preuves médicales. L'EBM consiste à fonder les décisions cliniques sur des connaissances théoriques et sur les preuves scientifiques tout en tenant compte des préférences des patients. La médecine basée sur les preuves est conçue comme une méthodologie explicite de recherche de meilleures preuves disponibles pour résoudre un problème clinique, elle s'effectue en quatre étapes :

1. la formulation du problème médical en une question claire et précise.
2. la recherche dans la littérature des articles les plus pertinents en rapport avec la question posée.
3. l'évaluation de la fiabilité et l'applicabilité des conclusions extraites des articles retenus, la pertinence clinique.
4. l'intégration des conclusions retenues pour répondre à la question initiale posée.

Ces étapes présentent la démarche systématique de l'EBM. La question se formule selon le modèle PICO élaboré par la *Cochrane Collaboration*. Les questions posées concernent le diagnostic, l'étiologie, le traitement, le pronostic selon le modèle PICO comme suit :

— **P** : Patient (âge, sexe..) et le problème qu'il pose *Who is affected ?*

— **I** : Intervention qu'est ce qui est envisagé ? Qu'est ce qui est fait ? L'intervention envisagée peut être une méthode de diagnostic ou de traitement.

— **C** : Comparaison par rapport à une autre intervention, *How effective are different interventions ?*

— **O** : *Output* qui représente l'issue clinique recherchée, *Does it work ?*

De nombreux travaux (Schardt *et al.*, 2007; Boudin *et al.*, 2010a; Boudin *et al.*, 2010b; Boudin *et al.*, 2010d) se sont intéressés à l'identification de ces éléments PICO à partir des requêtes et des documents dans le contexte de la RI médicale. Cette phase est primordiale pour une utilisation optimale de ces facettes dans les modèles de RI médicaux. Un deuxième volet des travaux est l'utilisation des éléments PICO dans la RI médicale. Peu de travaux ont été proposés dans ce cadre (Boudin *et al.*, 2010d), vu la complexité de l'identification pertinente des facettes à partir du texte, ainsi que la mise en évidence des preuves et leur intégration dans un modèle de RI qui intègre le contexte de recherche spécifique et assure la pertinence des résultats pour la prise de décision médicale.

Dans ce contexte, nous nous intéressons particulièrement dans cette partie du manuscrit, à la problématique de répondre au mieux aux questions cliniques PICO, qui comporte une identification explicite de ces facettes et leur utilisation dans la recherche d'information.

Dans ce cadre, nous présentons nos contributions sous forme de trois sections dans le chapitre 4 :

— Une représentation sémantique sous forme de graphes des requêtes PICO. Nous proposons pour cela un algorithme de génération de graphes conceptuels.

— Une approche d'expansion sémantique de requêtes pour mieux répondre aux questions PICO. Pour atteindre cet objectif, nous avons proposé deux algorithmes : un premier algorithme permettant de sélectionner les meilleurs concepts à partir des graphes sémantiques, un deuxième algorithme pour l'expansion de requêtes en se basant sur une approche de propagation de scores.

— Une méthode d'ordonnancement des documents permettant de calculer les scores de pertinence selon une approche d'agrégation prioritaire des scores selon les facettes PICO.

Dans le chapitre 3, nous détaillons les facteurs caractéristiques des requêtes expertes et les méthodes statistiques utilisées pour nos analyses exploratoires ainsi que l'étude de l'impact de ces facteurs sur la pertinence des résultats. Dans le chapitre 4, nous présentons nos contributions visant à mieux répondre aux requêtes PICO.

Chapitre 3

Analyses statistiques exploratoires des facteurs caractéristiques des requêtes expertes

Sommaire

1 Introduction . . . 96
2 Motivations et questions de recherche . . . 96
2.1 Motivations . . . 96
2.2 Objectifs et questions de recherche . . . 97
3 Définition et formalisation des attributs de requêtes . . . 99
3.1 Longueur de la requête . . . 99
3.2 Spécificité de la requête . . . 100
3.3 Clarté de la requête . . . 101
4 Caractérisation du besoin en information des experts du domaine médical . . . 102
4.1 Description des données d'analyses . . . 103
4.1.1 Collections de la campagne d'évaluation TREC . . . 103
4.1.2 Collections de la campagne d'évaluation CLEF . . . 104
4.2 Analyses et résultats . . . 106
4.2.1 Identification des Caractéristiques de requêtes . . . 106
4.2.2 Analyses de corrélation des attributs de requêtes . . . 108
4.3 Synthèse et discussion . . . 109
4.3.1 Analyses multidimensionnelles des corrélations entre les attributs . . . 110
4.3.2 Impact des attributs des requêtes sur les performances de recherche . . . 112
5 Caractérisation des requêtes PICO . . . 114
5.1 Analyses descriptives . . . 114
5.2 Analyses de corrélations entre attributs de requêtes PICO . . . 115
6 Bilan et conclusion . . . 118

1 Introduction

Dans ce chapitre, nous nous intéressons au contexte de la RI biomédicale, où les utilisateurs sont caractérisés par un niveau d'expertise vis-à-vis du besoin en information. Nous abordons le problème de la caractérisation des requêtes des experts médicaux. En effet, les investigations de recherche portant sur l'analyse de comportements des utilisateurs de SRI médicaux, sur l'analyse de la formulation des requêtes médicales ainsi que sur les pratiques de recherche ont montré que plusieurs facteurs contribuent au succès ou à l'échec de la recherche et impactent la pertinence des résultats (Richesson *et al.*, 2010; Cartright *et al.*, 2011; Lykke *et al.*, 2012). À la différence des précédents travaux dans le domaine, notre objectif n'est pas d'évaluer l'impact du degré d'expertise sur les résultats de recherche en considérant une population large d'utilisateurs, mais de caractériser les requêtes expertes du point de vue de la formulation, en considérant la tâche médicale associée qui constitue son contexte.

Nous proposons deux volets de travaux portant sur le besoin des experts médicaux : (1) l'analyse des requêtes expertes pour caractériser le besoin des experts médicaux, via l'analyse des attributs de requêtes et de leur impact sur la pertinence des résultats ; et (2) l'analyse des attributs de requêtes PICO. Nous avons mené des études exploratoires expérimentales pour la *caractérisation du besoin en information des experts du domaine médical* (Znaidi *et al.*, 2013a; Znaidi *et al.*, 2013b). Dans la littérature de la recherche d'information, nous avons constaté que comprendre les intentions des utilisateurs lors de la formulation de leurs requêtes est très important pour améliorer la phase de sélection des résultats pertinents. Plusieurs études ont analysé comment les utilisateurs effectuent des recherches exploratoires dans des environnements numériques, mais peu se sont intéressées à comprendre comment les requêtes sont formulées spécifiquement par les experts du domaine. Cette étude vise à combler cette lacune en analysant un ensemble de requêtes expertes issues de différentes tâches médicales.

Nous nous sommes intéressés aussi à l'analyse des besoins en information exprimés par des experts médicaux dans l'objectif de les caractériser puis mesurer l'impact de leur structure sur la performance des résultats de recherche. À cet effet, nous menons une étude exploratoire basée sur des analyses statistiques multidimensionnelles sur des collections de requêtes issues de campagnes d'évaluation internationales standards en l'occurrence, TREC et CLEF.

La suite du chapitre est organisée comme suit. La section 2 présente les motivations et introduit les questions de recherche de nos contributions. La section 3 détaille la formalisation des attributs de requêtes que nous avons défini. Nous présentons dans la section 4 les résultats des analyses statistiques, en détaillant les données, les méthodes et les résultats obtenus. La section 5 conclut le chapitre.

2 Motivations et questions de recherche

2.1 Motivations

Les professionnels de santé représentent une classe sensible de personnes en raison de leur besoin d'un accès à l'information médicale d'une bonne qualité afin d'améliorer leur qualité de service (Case, 2012). De ce point de vue, la qualité de l'information médicale disponible sur Internet est un facteur critique pour la réalisation de services de soins de santé de haute

qualité et la prise de décision (Hersh, 2009; Jamal *et al.*, 2009; Clarke *et al.*, 2013). Dans le contexte des soins de santé, cela nécessite une bonne connaissance et une demande accrue de savoir-faire spécialisé qui ont un effet sur la formulation des besoins en information pour le personnel médical et infirmier (Thain et Wales, 2005). En outre, cette complexité a généré une hétérogénéité des besoins en information, qui varient selon les différentes spécialités et les fonctions des professionnels de la santé. Par exemple, dans (Kostagiolas *et al.*, 2014), les auteurs ont conclu que plusieurs facteurs influencent la décision des professionnels de santé comme la motivation personnelle, en soulignant le rôle important des ressources d'information médicale.

Dans les deux dernières décennies, de nombreuses études et revues ont été publiées dans le domaine de la RI médicale, afin de souligner l'importance d'étudier le besoin en information des experts pour améliorer la prise de décision médicale (White *et al.*, 2008; Francke *et al.*, 2008; Xiao *et al.*, 2014).
En effet, exprimer des requêtes qui reflètent un besoin en information clair et précis est une tâche difficile dans tous les domaines et même pour les utilisateurs experts (Spink et Jansen, 2006; White et Moris, 2008). Ainsi, identifier l'intention des utilisateurs cachée derrière les requêtes qu'ils soumettent aux systèmes de RI est un problème complexe.

Malheureusement, la littérature a révélé qu'en dépit de la diversité des systèmes médicaux disponibles et la diversité des sources d'information utilisées, les utilisateurs se trouvaient toujours en difficulté pour sélectionner les informations pertinentes qui répondent à leurs besoins spécifiques (Zeng *et al.*, 2004; Arora *et al.*, 2007). Pour répondre à cette question, plusieurs études ont porté sur l'analyse du comportement des personnes qui recherchent l'information médicale, incluant les attitudes, les stratégies de recherche, les tâches et les requêtes (Hong *et al.*, 2002; Wildemuth, 2004; Toms et Latter, 2007). Ces études concernent un grand nombre de sujets sur les paramètres de recherche Web générale, avec des conditions expérimentales non contrôlées, ce qui rend difficile de généraliser leurs conclusions à des recherches d'experts impliqués par les médecins. En outre, la plupart de ces études ont porté sur le comportement de recherche grâce à des stratégies et des tactiques de recherche.

2.2 Objectifs et questions de recherche

Les études antérieurs dans le contexte de la caractérisation du besoin médical concernaient généralement des populations larges et/ou ciblées dans des scénarios de recherche qui sont cependant non reproductibles. Contrairement aux travaux de l'état de l'art, nous abordons plus spécifiquement la recherche basée sur le besoin expert des cliniciens du domaine à travers l'analyse des différentes facettes d'attributs de requêtes définis et formalisés, à savoir : la longueur de la requête (en termes et en concepts), la spécificité de la requête (spécificité terme-document et spécificité hiérarchique) et la clarté de la requête (clarté basée sur le sujet de la requête et clarté basée sur la pertinence), en utilisant des mesures appropriées proposées et construites selon différentes sources de données.

Notre objectif est de caractériser les requêtes expertes du point de vue de la formulation, en considérant la tâche médicale associée qui constitue son contexte. Nous nous sommes intéressés à l'analyse des requêtes d'experts médicaux, établies dans des scénarios de recherche reproductibles puisqu'elles sont issues de campagnes d'évaluation standards dans le domaine

de la RI. De plus, notre analyse est exploratoire et adossée à des tâches biomédicales bien spécifiques dans le cadre de ces mêmes campagnes.

Pour atteindre cet objectif, nous avons procédé à une analyse statistique approfondie des requêtes issues des campagnes d'évaluation de la recherche d'information, à savoir TREC et CLEF, destinées à différentes tâches médicales avec des paramètres d'évaluation contrôlés.

Notre étude expérimentale comprend une analyse statistique de corrélations entre paires d'attributs, une analyse multidimensionnelle à travers les différentes tâches caractérisant les collections, ainsi qu'une étude de l'impact des attributs de requêtes sur la performance de RI. Ainsi, l'objectif principal est d'identifier les particularités de la RI des experts médicaux dans le but de personnaliser la recherche qui impacte la RI médicale et les décisions cliniques.

Nos objectifs de recherche sont les suivants :

— Caractériser le besoin en information des experts du domaine médical ;

— Étudier l'impact du contexte de la tâche sur la performance de RI ;

— Étudier les attributs de requêtes expertes.

Pour atteindre ces objectifs, nous formalisons tout d'abord un ensemble de facettes d'attributs de requête, à savoir :

— **Attribut 1** : *La longueur de la requête.* Nous proposons d'étudier deux facettes de l'attribut longueur, qui sont la longueur en nombre de termes et la longueur en nombre de concepts. En effet, dans (Cartright *et al.*, 2011), les auteurs ont montré l'impact du nombre de termes utilisés pour formuler la requête sur la performance de recherche. Par ailleurs, la longueur en nombre de concepts permet de traduire en quelle mesure les utilisateurs utilisent les terminologies médicales en s'appuyant sur leurs connaissances du domaine pour formuler les requêtes médicales.

— **Attribut 2** : *La spécificité de la requête.* Nous proposons deux facettes de la spécificité : spécificité terme-document et spécificité hiérarchique. La spécificité est considérée comme un critère important pour identifier les descripteurs (Jones, 1972). Pour cette raison, nous avons considéré les deux facettes précédentes ; la première est basée sur les termes et représente la singularité des termes dans les documents et une deuxième spécificité conceptuelle est basée sur la profondeur du sens des termes selon la hiérarchie MeSH.

— **Attribut 3** : *La clarté de la requête.* Nous étudions deux facettes de la clarté, à savoir la clarté basée sur le sujet de la requête et la clarté basée sur la pertinence. Nous avons proposé deux facettes de l'attribut clarté vu son importance dans la mesure de l'ambiguïté du sujet de la requête. Les deux facettes de clarté représentent deux mesures différentes : la première représente la divergence entre le modèle de langue de la requête et celui du document, et la deuxième se base sur le nombre de termes en commun partagés entre les requêtes et les documents.

Nous menons ensuite des analyses statistiques exploratoires sur un ensemble de requêtes expertes issues de diverses tâches médicales, à savoir : TRECMedical, ImageCLEF et TRECGenomics. Cette diversité nous conduira à caractériser le besoin en information des experts du domaine médical du point de vue de leur formalisation ainsi que de leur utilisation dans la tâche de recherche d'information.

Nous mettons en œuvre des analyses descriptives, des analyses de corrélations entre facettes deux à deux, des analyses de corrélations multidimensionnelles basée sur une Analyse aux

Composantes Principales (ACP), et des analyses comparatives par tâche et toutes tâches confondues.

Nous nous sommes appuyés sur les résultats de performance des requêtes en terme de précision, d'une part pour décrire chaque tâche médicale selon son niveau de performance ; et analyser les éventuelles corrélations entre les mesures de performance et pour expliquer les performances de recherche en fonction des attributs via une analyse de covariance multivariée (MANCOVA), d'autre part.

3 Définition et formalisation des attributs de requêtes

Dans notre étude, nous considérons un scénario de recherche d'information médicale, où un expert du domaine soumet une requête Q pour une collection de documents C. Nous proposons trois attributs qui caractérisent les requêtes : 1) la longueur, 2) la spécificité et 3) la clarté. Nous proposons une formalisation de ces trois attributs, et nous justifions leur utilisation.

Nous avons formalisé un ensemble de facettes d'attributs de requêtes (*Cf.* Tableau 3.1) , définies et formalisées dans les sous-sections suivantes.

Attributs	**Facette 1**	**Facette 2**
Longueur de la requête	Nombre de termes	Nombre de concepts
Spécificité de la requête	Spécificité terme-document	Spécificité hiérarchique
Clarté de la requête	Clarté basée sur le sujet de la requête	Clarté basée sur la pertinence

Tableau 3.1 – Formalisation des facettes des attributs de requêtes

Nous considérons pour le reste du chapitre les notations définies par le Tableau 3.2.

3.1 Longueur de la requête

Comme mentionné dans l'état de l'art, le facteur de la longueur des requêtes médicales a été étudié dans le cadre de plusieurs travaux (Hong *et al.*, 2002; Zeng *et al.*, 2002b; Spink *et al.*, 2004). Ces derniers ont rapporté que les requêtes sont courtes et que cela impacte les résultats de recherche. En effet, le nombre de termes utilisés dans la formulation de la requête joue un rôle important dans les résultats retournés par les SRI, dans la mesure où ils peuvent augmenter le bruit s'il y a des termes non pertinents ou qui ne représentent pas le besoin spécifique exprimé. De plus, les experts formulent leurs requêtes en se référant à une ou plusieurs terminologies médicales comme MeSH ou UMLS (Zeng *et al.*, 2002a; Zeng *et al.*, 2002b; Jonnalagadda *et al.*, 2012), ce qui explique la présence des concepts médicaux. Ce procédé très utilisé par les experts n'est pas très efficace quand le concept concerne plusieurs traitements ou maladies, ce qui crée l'ambiguïté et donc induit des résultats de recherche non pertinents.
Dans le cadre de notre étude, nous définissons deux facettes de l'attribut longueur :

— **Longueur en nombre de termes (LgT(Q))** : concerne la longueur de la requête en nombre de termes. Nous considérons ainsi le nombre de mots significatifs utilisés pour formuler les requêtes.

Notation	Définition
Q	requête
C	collection
N	nombre total de documents dans la collection C
V	le vocabulaire de la collection
R	l'ensemble des documents qui contiennent au moins un terme de la requête
d	un document
t_i	un terme
c_i	un concept
n_i	le nombre de documents contenant le terme t_i
$LgT(Q)$	longueur de la requête en nombre de termes
$LgC(Q)$	longueur de la requête en nombre de concepts
$DSpe(Q)$	spécificité terme-document
$Hspe(Q)$	spécificité hiérarchique
$SCla(Q)$	clarté basée sur le sujet de la requête
$PCla(Q)$	clarté basée sur la pertinence
$termes(Q)$	ensemble de termes de la requête Q
$Concepts(Q)$	ensemble de concepts de la requête Q
$level(c_i)$	le niveau du concept c_i dans MeSH
$Maxlevel(MeSH)$	le niveau maximal de la hiérarchie MeSH
$P_C(t)$	la fréquence relative du terme t
$P(t\|Q)$	est estimé par $P(t\|Q) = \sum_{d \in R} P(t\|d)P(d\|Q)$
$R(Q)$	ensemble de documents pertinents pour la requête Q jugés par les experts
$\|Concepts(d)\|$	le nombre de concepts extraits des documents
$\|Concepts(Q)\|$	le nombre de concepts extraits des requêtes

Tableau 3.2 – Tableau des notations

— **Longueur en nombre de concepts (LgC(Q))** : c'est la longueur de la requête en nombre de concepts. Ces concepts représentent les entrées préférées (termes préférés désignant le concept) issues de la terminologie *MeSH*. Notre choix de cette terminologie se justifie par le fait que c'est l'ontologie la plus utilisée dans le domaine médical. Pour cela, chaque requête est liée à la terminologie *MeSH*, en utilisant la technique d'extraction de concepts contextuelle proposée dans (Dinh et Tamine, 2011a). Cette méthode s'appuie sur une technique d'indexation sémantique pour récupérer les documents médicaux à travers un processus d'identification de concepts du domaine extraits de MeSH.

Le Tableau 3.3 donne un exemple de requêtes avec le nombre de termes et de concepts extraits pour différentes tâches médicales.

3.2 Spécificité de la requête

La spécificité est une propriété sémantique des termes de l'index : la spécificité d'un terme est définie comme le niveau de détail et de précision de son sens.

Dans notre étude, nous nous sommes intéressés à deux facettes de la spécificité, à savoir :

— **Spécificité terme-document (Dspe(Q))** : la spécificité terme-document désigne l'unicité des termes de la requête dans l'index de la collection de documents. Notre hypothèse justifiant l'utilisation de la spécificité terme-document s'énonce comme suit : moins la collection contient des termes de la requête, plus les sujets de la requête sont

Id	Collection	Description	# Termes	# Concepts
102	TRECMedical 2011	*Patients with complicated GERD who receive endoscopy*	7	2
32	ImageCLEF 2011	*Pain and incapacity to move after an accident. Slight deformation can be seen in the x-ray*	16	2
6	TRECGenomics03	*phosphatase and tensin homolog (mutated in multiple advanced cancers 1)*	11	2
OHSU1	Ohsumed	*50 year old menopausal woman without hormone replacement therapy*	9	2

Tableau 3.3 – Exemples de requêtes avec les facettes de la longueur

spécifiques. Elle est calculée par :

$$DSpe(Q) = \frac{1}{LgT(Q)} * \sum_{t_i \in termes(Q)} -log(\frac{n_i}{N}) \tag{3.1}$$

Où $LgT(Q)$ est la longueur de la requête en nombre de termes, $termes(Q)$ est l'ensemble des termes de la requête, n_i est le nombre de documents contenant le terme t_i et N est le nombre total des documents de la collection C.

— **Spécificité hiérarchique (Hspe(Q))** : la spécificité hiérarchique dépend de la profondeur du sens du concept défini dans la terminologie MeSH. Nous avons exploité la position arborescente des concepts dans la hiérarchie MeSH. Notre hypothèse est qu'un concept fils est plus spécifique que le concept père dans la hiérarchie de la terminologie. Nous définissons la spécificité hiérarchique de la manière suivante :

$$HSpe(Q) = \frac{1}{LgC(Q)} * \sum_{c_i \in Concepts(Q)} \frac{level(c_i) - 1}{Maxlevel(MeSH) - 1} \tag{3.2}$$

où $LgC(Q)$ est la longueur de la requête en nombre de concepts, $Concepts(Q)$ est le nombre de concepts de la requête, $level(c_i)$ est le niveau du concept c_i dans MeSH, $Maxlevel(MeSH)$ est le niveau maximal de la hiérarchie MeSH.

Le Tableau 3.4 donne un exemple de requêtes avec les deux facettes de la spécificité pour les collections TREC Medical 2011 et ImageCLEF 2011.

3.3 Clarté de la requête

D'une manière générale, une requête claire dégage un sens pertinent et significatif du sujet, tandis qu'une requête ambiguë dégage une variété de sujets et de significations/sens qui ne sont pas corrélés. L'attribut de la clarté reflète d'une manière générale le degré d'ambiguïté de la requête par rapport au document retourné. Nous proposons deux facettes pour l'attribut de la clarté :

Id	Collection	Description	$DSpe(Q)$	$Hspe(Q)$
104	TRECMedical 2011	*Patients diagnosed with localized prostate cancer and treated with robotic surgery*	0.543	0.271
40	ImageCLEF 2011	*Young female gymnast presents with leg pain*	0.147	0.253

Tableau 3.4 – Exemples de requêtes avec les facettes de la spécificité

— **Score de clarté basé sur le sujet de la requête (SCla(Q))** : le score de clarté d'une requête est calculé par la divergence de *Kullback-Leibler* entre le modèle de langue de la requête et le modèle de langue de la collection, donnée par (Cronen-Townsend et Croft, 2002), soit :

$$SCla(Q) = \sum_{t \in V} P(t|Q) log_2 \frac{P(t|Q)}{P_C(t)} \tag{3.3}$$

où V est le vocabulaire de la collection, t un terme, $P_C(t)$ est la fréquence relative du terme t et $P(t|Q)$ est estimée comme suit :

$$P(t|Q) = \sum_{d \in R} P(t|D)P(D|Q) \tag{3.4}$$

où d est un document, R est l'ensemble des documents qui contiennent au moins un terme de la requête.

— **Score de clarté basé sur la pertinence (PCla(Q))** : une requête est supposée être d'autant plus claire qu'elle partage plus de concepts avec les documents jugés pertinents par les experts. Cette hypothèse est la base des modèles de recherche d'information. Par conséquent, nous calculons $PCla(Q)$ comme suit :

$$PCla(Q) = \frac{1}{R(Q)} * \sum_{d \in R(Q)} \frac{|Concepts(Q) \cap Concepts(d)|}{LgC(Q)} \tag{3.5}$$

où $R(Q)$ est l'ensemble de documents pertinents pour la requête Q jugés par les experts, $Concepts(d)$, (respectivement $Concepts(Q)$), est l'ensemble de concepts des documents (respectivement des requêtes).

Le Tableau 3.5 donne un exemple de requêtes avec les deux facettes de la clarté pour les collections TRECGenomics 2003 et ImageCLEF 2011.

4 Caractérisation du besoin en information des experts du domaine médical

Cette section détaille les méthodologies d'analyse avec lesquelles nous étudions comment les requêtes sont formulées spécifiquement par les utilisateurs experts du domaine médical ainsi que les résultats des analyses expérimentales conduites.

Id	Collection	Description	$SCla(Q)$	$PCla(Q)$
3	TRECGenomics 2003	*eukaryotic translation initiation factor 4E*	0,083	0,0071
38	ImageCLEF 2011	*68-year-old male, Glasgow Coma Score of 6. Found unconscious by family.*	0,091	0,0087

Tableau 3.5 – Exemples de requêtes avec leurs différents scores de clarté

Nous étudions 173 requêtes médicales expertes issues de 3 tâches de RI médicales dans le cadre de deux campagnes d'évaluation différentes, CLEF et TREC. Les tâches représentent différentes collections de plusieurs années de TRECMedical, ImageCLEF et TRECGenomics. Nous procédons à des analyses statistiques pour étudier, d'une part, les variations et les corrélations des attributs de requêtes médicales définis auparavant, à savoir : la longueur, la clarté et la spécificité , et, d'autre part, l'impact des attributs de requêtes sur les résultats de recherche.

4.1 Description des données d'analyses

Pour réaliser l'étude statistique, nous avons utilisé les données issues de deux campagnes d'évaluation du domaine de RI, à savoir : TREC et CLEF. Nous avons exploité les requêtes, les documents et les données de jugement de pertinence des médecins en suivant les différentes tâches de la RI médicale. Des statistiques des différentes collections de test utilisées dans notre étude sont illustrées dans le Tableau 3.6. Nous décrivons dans ce qui suit chacune des collections.

Collections	**Nb.requêtes**	**Nb.documents**
TREC Medical 2011	35	$95,701$
TREC Genomics 2003	50	$525,938$
TREC Genomics 2004	50	$4,591,008$
TREC Genomics 2006	28	55634
ImageCLEF 2011	10	$55,635$

Tableau 3.6 – Descriptions des collections utilisées pour l'analyse

4.1.1 Collections de la campagne d'évaluation TREC

Dans le cadre de nos analyses, nous nous sommes intéressés à deux collections de la recherche médicale ad hoc, à savoir : TREC medical et TREC Genomics, décrites dans ce qui suit.

— *TREC Medical record task* : la tâche de recherche consiste à identifier les groupes pour comparer l'efficacité de la recherche. Les requêtes décrivent un ensemble de maladies et états développés par les médecins et les documents représentent des rapports de visites médicales. Pour nos analyses, nous avons utilisé la tâche *TREC Medical 2011* qui comprend 35 requêtes et 95701 documents. Le Tableau 3.7 donne des exemples de requêtes de la collection TRECMedical 2011.

101 : Patients with hearing loss. **102 :** Patients with complicated GERD who receive endoscopy. **103 :** Hospitalized patients treated for methicillin-resistant Staphylococcus aureus (MRSA) endocarditis. **104 :** Patients diagnosed with localized prostate cancer and treated with robotic surgery.

Tableau 3.7 – Exemples de requêtes de la collection de test de la tâche TRECMedical 2011.

— *TREC Genomics series task* : *TREC Genomics* représente la tâche la plus couramment utilisée en RI biomédicale. La tâche modélise un cadre où un utilisateur qui cherche l'information sur le génome exprime une requête et la soumet à un moteur de recherche qui se base sur une collection de documents médicaux de la littérature scientifique MEDLINE. Pour nos analyses, nous exploitons la *TRECGenomics 2003* qui traite les noms de gènes (50 requêtes et 525928 documents), *TRECGenomics 2004* où les besoins en information sont exprimés en acronymes (50 requêtes et 4591008 documents) et *TRECGenomics 2006* qui concerne la question-réponse (28 requêtes et 55634 documents). Le Tableau 3.8 donne des exemples de requêtes des différentes collections de *TRECGenomics*.

TrecGeomics2003	Protein tyrosine phosphatase, non-receptor type 1.
TrecGeomics2004	How do AAA proteins mediate interaction with lipids or DNA and what is their functional impact ?
TrecGeomics2006	How do alpha7 nicotinic receptor subunits affect ethanol metabolism ?

Tableau 3.8 – Exemples de requêtes de différentes collections de test de la tâche TRECGenomics.

4.1.2 Collections de la campagne d'évaluation CLEF

ImageCLEF case-based task. Le but de cette tâche est de chercher et sélectionner les cas de patients en se basant sur des images que les médecins jugent pertinents par rapport à différents diagnostics. Les requêtes ont été créées à partir d'une base médicale de cas de patients existante, qui comprend la description démographique des patients, les symptômes, les résultats de tests ainsi que les études d'images.
Le Tableau 3.9 synthétise les descriptions des différentes collections de test utilisées pour l'analyse.

Collection-Requête		**TRECMedical11**		**ImageCLEF11**		**TRECGenomics03**		**TRECGenomics04**		**TRECGenomics06**	
		Moyenne	σ	Moyenne	σ	Moyenne	σ	Moyenne	σ	Moyenne	σ
Nombre de requêtes		35		10		50		50		28	
Critères	LgT(Q)	7,22	2,85	24	15,9	4,6	2,41	9,4	3,35	9	2,91
	LgC(Q)	3,22	1,28	5,2	2,61	1,36	0,72	2,52	1,31	2,46	1,31
	DSpe(Q)	0,44	0,07	0,17	0,08	0,25	0,08	0,0010	0,0016	0,0070	0,0076
	HSpe(Q)	0,20	0,13	0,13	0,06	0,26	0,21	0,27	0,31	0,20	0,07
	SCla(Q)	1.10	0,84	0,95	0,69	1,5	0,80	1,32	0,90	1,28	0,84
	PCla(Q)	0,026	0,03	0,041	0,027	0,19	0,16	0,003	0,001	0,007	0,004

Tableau 3.9 – Récapitulatif des critères selon la tâche médicale.

4.2 Analyses et résultats

Nous présentons dans ce qui suit les méthodologies des analyses statistiques ainsi que les résultats obtenus en utilisant deux outils d'analyse principaux : SAS statistiques et R. De plus, notre analyse est exploratoire et adossée à des tâches biomédicales bien spécifiques dans le cadre de ces mêmes campagnes. À notre connaissance, aucune étude de requêtes médicales capitalisées lors des différentes campagnes d'évaluation TREC et CLEF n'a été menée à ce jour.

4.2.1 Identification des Caractéristiques de requêtes

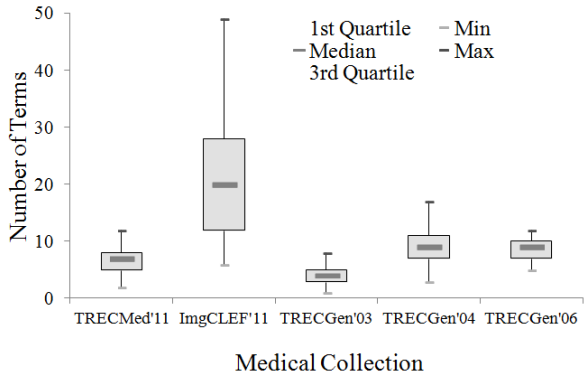
Pour identifier les différences entre les collections issues des différentes tâches médicales, nous avons commencé par des analyses statistiques descriptives. La Figure 3.1 montre la distribution des six facettes des attributs de requêtes par collection et pour l'ensemble des requêtes, présentées par des boites à moustache.

Nous avons procédé à des analyses de variance (*i.e., non-parametric Kruskall Wallis test* adapté aux petits échantillons) dans le but de comparer les moyennes des scores des facettes d'attributs de requêtes et détecter les différences significatives entre les collections étudiées (indiquées par *p-value<0,05*).

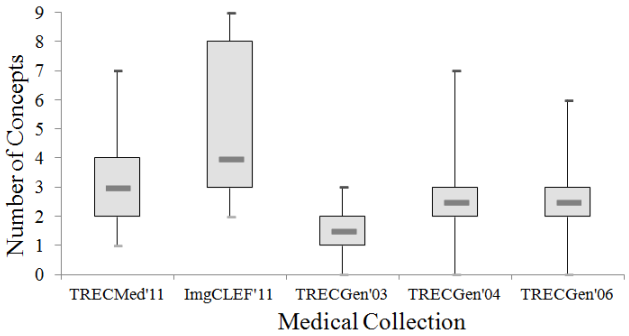
À partir des figures 3.1.(a) et 3.1.(b), nous observons des tendances similaires pour les deux attributs de la longueur. De plus, l'attribut longueur de la requête est significativement différent entre les 5 collections (*p-value<0,0001*) malgré le fait que toutes les requêtes représentent des besoins en information des experts. La valeur la plus élevée pour l'attribut longueur a été observée pour la collection *ImageCLEF* avec des nombres moyens de 24 termes et 5 concepts. À l'opposé, les requêtes les plus courtes ont été observées pour la collection *TRECGenomics 2003* avec en moyenne $4,6$ termes et $1,4$ concepts.

Cela peut s'expliquer par les différences majeures des tâches associées. En effet, pour *ImageCLEF*, les médecins expriment des requêtes longues qui décrivent les cas des patients incluant des images alors que pour *TRECGenomics 2003*, les experts médicaux expriment des requêtes courtes sur les noms de gènes. De ce fait, il est rare d'utiliser les concepts médicaux à partir d'une ressource terminologique.

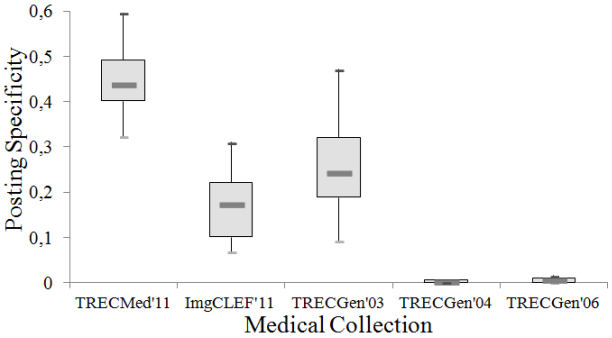
Les figures 3.1.(c) et 3.1.(d) représentent les distributions respectives des scores de spécificité terme-document et spécificité hiérarchique basée sur la terminologie MeSH. En considérant les définitions de ces deux facettes, les scores sont différents. On observe des écarts significatifs pour la spécificité terme-document à travers toutes les collections (*p-value<0,0001*), tandis que les spécificités hiérarchiques ne sont pas significativement différentes entre les collections. Comme le montre la figure 3.1.(c), les collections *TRECGenomics 2004* et *TRECGenomics 2006* sont caractérisées par des valeurs faibles de la spécificité terme-document (valeur maximale de $0,000964$ pour *TRECGenomics 2004* et $0,0032$ pour *TRECGenomics 2006*), comparés aux autres collections ($0,549$ pour *TRECMedical 2011* et $0,310$ pour *ImageCLEF 2011*). Cela peut être expliqué par la nature de la tâche : dans la collection *TRECGenomics 2004*, les experts utilisent d'une manière intensive les acronymes et les abréviations qui sont mal répartis dans les documents MEDLINE. Dans *TRECGenomics 2006*, les requêtes sont exprimées sous forme de questions basées sur des entités spécifiques portant sur les gènes et



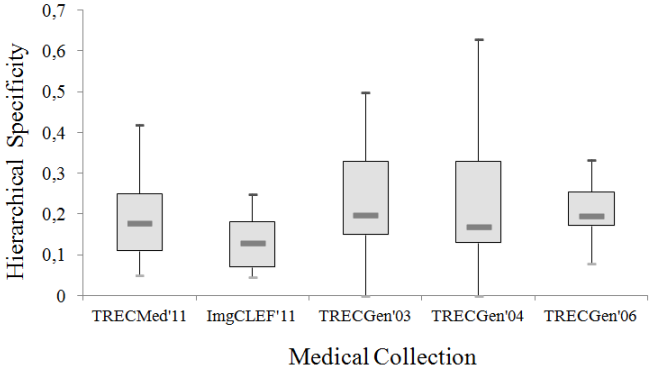
(a) Longueur en termes



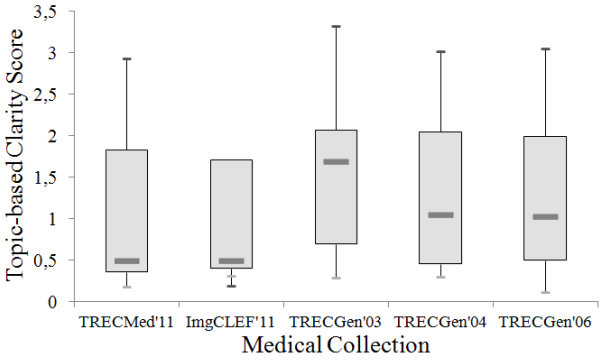
(b) Longueur en concepts



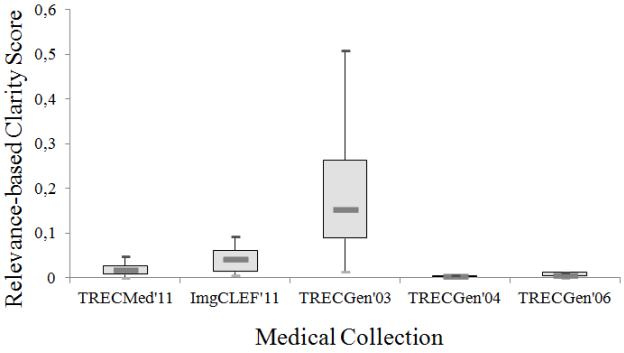
(c) Spécificité terme-document



(d) Spécificité hiérarchique



(e) Clarté basé sur le sujet



(f) Clarté basé sur la pertinence

FIGURE 3.1 – Distribution des facettes d'attributs de requêtes par collection

les protéines.

En ce qui concerne la spécificité hiérarchique, nous observons une plage de valeurs plus large pour *TRECGenomics 2003*, *TRECGenomics 2004* et potentiellement *TRECMedical 2011*. Cela indique que les experts médicaux utilisent des termes médicaux spécifiques à travers les terminologies. Les valeurs les plus élevées des attributs de la spécificité indiquent que les experts utilisent leurs connaissances du domaine médical afin d'exprimer leur besoin en information spécifique.

Par ailleurs, en analysant l'attribut de la clarté des Figures 3.1.(e) et 3.1.(f), nous observons des différences entre les collections. Ces dernières sont plus importantes pour la clarté basée sur la pertinence (*p-value<0,0001*) que pour la clarté basée sur le sujet (*p-value<0,05*). Cela est probablement dû à la large variance des scores de cette dernière comme illustrée par les boites a moustache dans la Figure 3.1.(e). Cependant, il y a des tendances similaires : les scores les plus élevés sont identifiés pour les requêtes *TRECGenomics 2003*, par opposition de ceux de *ImageCLEF*. Cela indique que la recherche sur les gènes et les protéines favorise l'expression de requêtes non ambiguës alors que les descriptions de cas de patients montre les différentes intentions des experts. On a précédemment découvert que les requêtes sur les gènes et les protéines sont courtes, on peut ainsi conclure que les requêtes courtes peuvent être claires et que cela dépend de la nature de la tâche de RI concernée.

4.2.2 Analyses de corrélation des attributs de requêtes

Nous visons à étudier les liens entre les facettes d'attributs de requêtes. En d'autres termes, notre but est d'identifier les corrélations significatives qui peuvent caractériser les requêtes médicales expertes. Dans un premier temps, nous analysons les corrélations entre chaque facette au sein de la même collection médicale. Dans un deuxième temps, nous étudions les corrélations entre attributs des différentes collections. Pour étudier les corrélations entre les attributs de requêtes, nous avons calculé le coefficient de Spearman (*Spearman correlation coefficient* ρ) entre les six attributs quantitatifs des requêtes. Nous présentons les corrélations fortement significatives dans le Tableau 3.10.

Pour les quatre collections les plus importantes en nombre de requêtes et documents, à savoir, *TRECMedical 2011*, *TRECGenomics 2003 2004 et 2006*, nous observons systématiquement des corrélations significatives positives entre les deux facettes de longueur en nombre de termes et en nombre de concepts (*pvalue <0,0001*). Intuitivement, nous l'expliquons ainsi :

1. Une entrée d'un concept biomédical dans la terminologie est généralement, par définition, un ensemble de termes. En effet, dans la terminologie MeSH, les concepts médicaux sont représentés par un ou plusieurs termes.
2. La deuxième raison, comme mentionné dans le précédent chapitre, est liée à la stratégie de recherche des experts médicaux, qui cherchent l'information médicale en favorisant l'utilisation des concepts en se basant sur leurs connaissances du domaine.

Nous observons également que toutes les corrélations significatives entre attributs impliquent la longueur de la requête en nombre de termes associés, reflétant l'importance de cette facette dans la caractérisation des besoins en information des experts du domaine. Nous pouvons également observer d'autres corrélations significatives sur les autres collections. Notons que ces corrélations ne sont pas systématiques.

Dans les collections *TRECMedical* et *TRECGenomics 2006*, nous observons des corrélations significatives positives entre le nombre de termes et la spécificité basée sur le sujet de la requête. Cela peut être expliqué en partie par le fait que, selon la Formule (3.3), plus la requête est longue, plus sa spécificité basée sur le sujet est importante. D'une manière générale, les scores de corrélation sont particulièrement élevés pour les deux collections, ce qui est dû à la nature de la tâche de recherche indépendamment de la représentation du besoin sous la forme

d'une requête simple ou d'une question factuelle liée à un cas de patient spécifique. Toutefois, la longueur de la requête en nombre de termes et la spécificité hiérarchique sont corrélées négativement pour la collection *TRECGenomics 2003* (*pvalue < 0,0001*). Cela s'explique par la nature de la tâche ainsi que du fait que la terminologie MeSH utilisée pour extraire les concepts n'est probablement pas appropriée. L'utilisation de l'ontologie *GENE Ontology* pour identifier les termes spécifiques est probablement la ressource la plus adéquate pour cette tâche. Enfin, pour la collection *ImageCLEF*, il n'existe aucune corrélation significative entre les attributs de requête. Cela s'explique par le nombre faible de requêtes de cette collection.

Collection	**Facette1**	**Facette2**	ρ	$p-value$
TREC Medical2011 ($N = 35$)	$LgT(Q)$	$LgC(Q)$	$0,69$	$< 0,0001$
	$LgT(Q)$	$DSpe(Q)$	$0,39$	$< 0,02$
TREC Genomics2003 ($N = 50$)	$LgT(Q)$	$LgC(Q)$	$0,55$	$< 0,0001$
	$LgT(Q)$	$HSpe(Q)$	$-0,54$	$< 0,0001$
TREC Genomics2004 ($N = 50$)	$LgT(Q)$	$LgC(Q)$	$0,47$	$< 0,001$
TREC Genomics2006 ($N = 28$)	$LgT(Q)$	$LgC(Q)$	$0,67$	< 0.0001
	$LgT(Q)$	$DSpe(Q)$	$0,58$	$< 0,001$

Tableau 3.10 – Résultats des corrélations entre les facettes d'attributs de requêtes des différentes collections

4.3 Synthèse et discussion

Les résultats d'analyses issues de cette étude donnent un aperçu détaillé sur les particularités des requêtes des experts du domaine médical selon les différentes tâches du domaine (Znaidi *et al.*, 2013a; Znaidi *et al.*, 2013b). Les principales conclusions montrent que contrairement aux requêtes des utilisateurs qui cherchent l'information de santé sur internet (Zeng *et al.*, 2002b), les requêtes des médecins sont relativement longues et dépendent de la nature de la tâche : la recherche de cas de patients se caractérise par des requêtes plus longues comparée à la recherche basée sur les entités. En effet, les médecins utilisent leurs connaissances du domaine ainsi que les ressources sémantiques pour formuler les requêtes, ce qui rend leur recherche plus spécifique, particulièrement pour la recherche des cas de patients.

Ces différentes constatations suggèrent de nouvelles fonctionnalités pour la conception des futurs systèmes de RI médicaux, telles que la reformulation et suggestion de requêtes utilisant les niveaux de terminologies appropriées pour améliorer la clarté de la requête, la personnalisation des résultats de recherche basés sur le niveau d'expertise, la catégorie des requêtes et la tâche de RI (Znaidi *et al.*, 2013b).

De plus, la spécificité de la requête des experts est influencée par la nature de la tâche indiquant, d'une part, l'utilisation d'acronymes et d'abréviations qui sont fréquents/répandus dans la collection MEDLINE, d'une autre part, l'usage significatif par les experts de leurs connaissances du domaine. Nos analyses ont également montré que rechercher les informations sur les gènes et les protéines favorise l'expression des requêtes clairement formulées alors que les descriptions des cas de patients sont traduites de façon plus ambiguë par les experts ce qui laisse supposer que l'appariement de cas pathologiques est une tâche experte.

Dans ce qui suit, nous présentons notre deuxième contribution dans le cadre des analyses

exploratoires du besoin en information des experts. La section 4 décrit le deuxième volet d'analyses sur les collections TREC et CLEF avec une étude de l'impact des attributs de requêtes sur la performance de la recherche. Il faut noter que nous utilisons les mêmes attributs de requêtes formalisés dans la section précédente (section 3.1). La section 5 présente des analyses statistiques conduites sur les requêtes PICO, spécifiques à la recherche d'information médicale basée sur les preuves.
Nous complétons la précédente étude en traitant les questions de recherche suivantes :

1. **Comment les attributs caractéristiques des requêtes sont corrélés en considérant les tâches médicales ?** Pour répondre à cette question, nous avons utilisé l'ensemble d'attributs de requête définis dans la première partie de l'étude, à savoir : longueur, spécificité et clarté. Nous proposons de mener une analyse de corrélations multidimensionnelles à l'aide d'une Analyse en Composantes Principales (ACP), toutes tâches confondues.

2. **Comment les attributs de requêtes impactent les performances de recherche ?** Pour répondre à cette question, nous nous sommes appuyés sur les performances des requêtes en terme de mesure de précision, d'une part, pour décrire chaque tâche médicale selon son niveau de performance et analyser les éventuelles corrélations entre les mesures de performance ; d'une autre part, pour expliquer les performances de recherche en fonction des attributs via une analyse de covariance multivariée (MANCOVA).

Pour répondre à ces questions, et en considérant l'ensemble d'attributs de requêtes déjà définis, nous avons conduit deux types d'analyses, à savoir : une première partie porte sur une analyse multidimensionnelle des corrélations entre les attributs et une seconde série d'analyses des mesures de performance de recherche et leurs corrélations avec les attributs de requêtes.

Nous considérons un cadre de recherche d'information où un expert du domaine soumet une requête Q à une collection de documents cibles C. Rappelons que nous utilisons trois attributs qui caractérisent les requêtes : 1) la longueur, 2) la spécificité, et 3) la clarté.

4.3.1 Analyses multidimensionnelles des corrélations entre les attributs

Dans le cadre de cette deuxième contribution, nous nous concentrons sur l'analyse des requêtes formulées par des experts du domaine médical. Notre analyse statistique est conduite en deux phases. La première a pour objectif l'analyse multidimensionnelle des corrélations entre les attributs. La seconde phase de notre analyse met en avant les mesures de performance et leurs éventuelles corrélations avec les attributs de requêtes. Pour ces deux phases, nous avons effectué des analyses en composantes principales (ACP). Les différences entre tâches ou entre requêtes cliniques et non-cliniques ont été testées par des analyses de variance ou des tests non-paramétriques de *Kruskall-Wallis* (adaptés aux petits échantillons). Dans un dernier temps, une analyse de covariance multidimensionnelle a permis de modéliser les mesures de performance en fonction des attributs quantitatifs et de la classification clinique ou non-clinique des requêtes.

Pour étudier les corrélations entre les six attributs des requêtes, nous avons réalisé une analyse en composantes principales (ACP) permettant de prendre en compte les corrélations entre

attributs d'un point de vue multidimensionnel. L'étude des trois premiers axes principaux a permis de dégager les tendances principales de chaque collection. La Figure 3.2 représente la projection des 173 requêtes selon les deux premiers axes. Elle illustre que :

— La collection *TRECGenomics 2003*, spécialisée dans la recherche sur les gènes et protéines, est caractérisée par des requêtes courtes en nombre de termes et de concepts, et claires (selon le score de clarté basée sur la pertinence).

— Les requêtes de la collection *TRECMed 2011* regroupant des compte-rendus médicaux de patients sont spécifiques, claires et écrites avec un nombre assez important de concepts, mais avec un nombre réduit de termes non identifiés comme des concepts. Elles s'opposent aux requêtes des collections *TRECGenomics 2004* et *TRECGenomics 2006* dont la spécificité est plus faible (comme nous l'avons mis en évidence Figure 3.1.(c)).

— La collection *ImageCLEF*, principalement axée sur la recherche de cas de patients, se distingue des autres collections par un faible nombre de requêtes qui peuvent être longues, voire très longues.

En outre, l'interprétation du troisième axe de cette ACP révèle l'indépendance du score de clarté (basé sur le sujet de la requête) vis-à-vis des autres attributs : l'ambiguïté d'une requête ne dépend ni de sa longueur, ni de sa spécificité, et ceci dans les cinq collections étudiées.

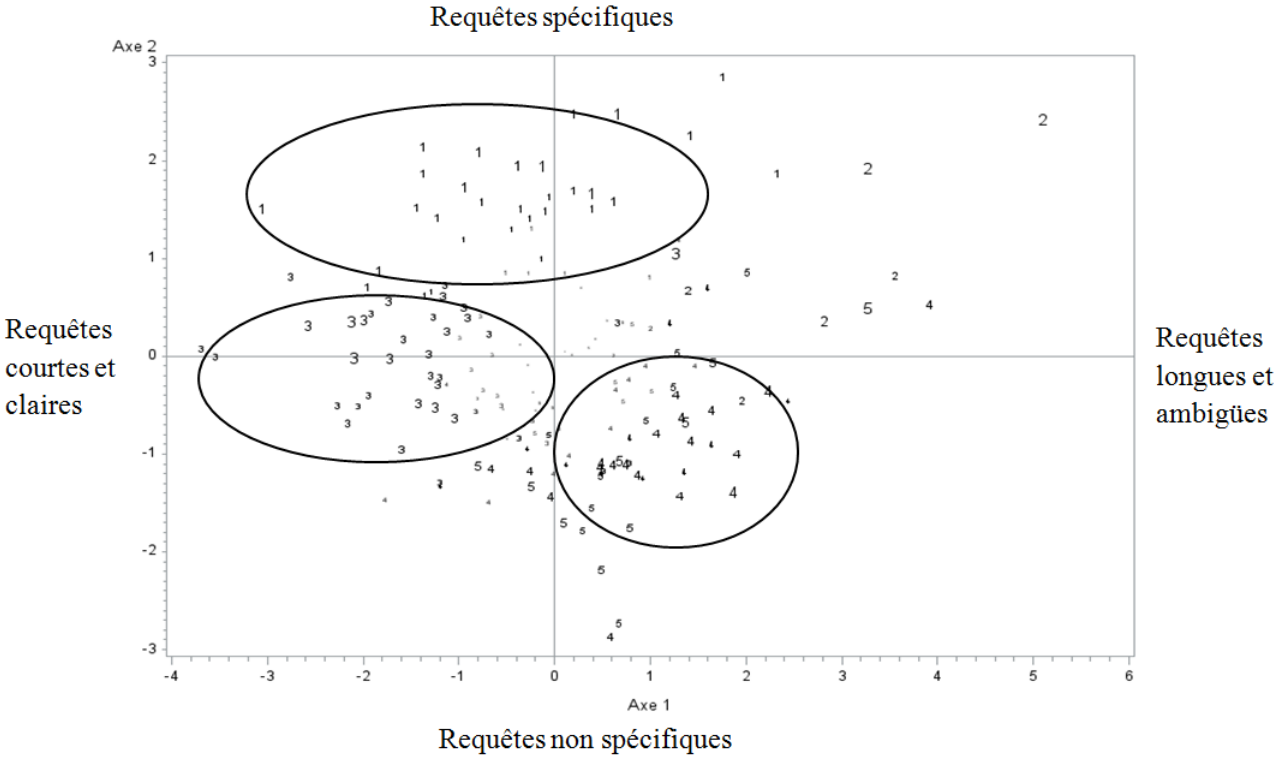


FIGURE 3.2 – Nuage de points résultats de l'ACP, 1 désigne la collection *TRECMedical* ; 2 désigne la collection ImageCLEF ; 3 désigne la collection *TRECGenomics03* ; 4 désigne la collection *TRECGenom04* et 5 désigne la collection *TRECGenom06*.

4.3.2 Impact des attributs des requêtes sur les performances de recherche

Cette phase des analyses est complémentaire à l'analyse des facettes des attributs présentée à la section précédente. Les performances de recherche sur l'ensemble des requêtes ont été mesurées suite à une évaluation avec le système *Terrier*[1], et l'outil d'évaluation standard *trec-eval*. Plus précisément, nous avons généré quatre scores de performance calculés sur la base du modèle *BM25*, à savoir : *MAP* (*Mean Average Precision*) et les précisions exactes à 10, 20 et 100 documents (notées respectivement P@10, P@20 et P@100).

Dans un premier temps, nous avons mené une étude comparative des scores afin de caractériser les performances relatives à chaque tâche et de mettre en évidence d'éventuelles différences entre les tâches. Le Tableau 3.11 présente les moyennes (m) et écart-types (sd) des quatre scores calculés pour chaque tâche, et le résultat du test non-paramétrique de *Kruskall-Wallis* de comparaison des scores (*p-value*).

Cette analyse révèle que les performances des requêtes issues de la collection *TRECGenomics 2003* sont caractérisées par des scores moyens significativement plus faibles (inférieurs à $0,07$) que les quatre autres collections (dont les scores moyens sont compris entre $0,28$ et $0,54$). Ceci s'explique par la difficulté de la tâche lors de son année de lancement en 2003 : les documents résultats sont jugés effectivement pertinents seulement dans le cas où ils s'apparentent avec des *Gene Reference Into Function (GeneRIF)*. Or, étant donné que peu d'annotations textuelles *GeneRIF* étaient disponibles en 2003, les performances des SRI sont de ce fait sous-estimées (Hersh et Voorhees, 2009).

Nous soulignons que les scores moyens de ces quatre collections ne sont pas significativement différents, suggérant une performance homogène des requêtes des quatre collections.
Par ailleurs, l'analyse des corrélations entre les quatre scores de performance a montré que tous les scores sont très fortement corrélés positivement deux-à-deux (*p-value* $< 0,0001$), et ceci sur l'ensemble des collections ; autrement dit, une requête évaluée comme performante par l'un des scores le sera également par les trois autres mesures.

	Scores de performances : m (sd)			
Tâche/Score	*MAP*	*P@10*	*P@20*	*P@100*
TRECMed 2011	0,32 (0,22)	0,45 (0,34)	0,43 (0,32)	0,28 (0,22)
ImageClef	0,30 (0,24)	0,54 (0,45)	0,48 (0,38)	0,31 (0,27)
TrecGenom 2003	0,067 (0,130)	0,038 (0,085)	0,045 (0,080)	0,028 (0,047)
TrecGenom 2004	0,36 (0,25)	0,53 (0,35)	0,47 (0,32)	0,34 (0,26)
TrecGenom 2006	0,34 (0,23)	0,45 (0,36)	0,11 (0,34)	0,31 (0,27)
p-value	† † †	† † †	† † †	† † †

Tableau 3.11 – Scores moyens de performance par collection (et écart-type) avec la *p-value* du test de comparaison des scores entre collections († † † : *p-value* $< 0,001$))

L'objectif final de cette étude est d'évaluer l'impact des attributs sur les performances des

1. http ://terrier.org

requêtes. Les performances étant mesurées par quatre scores (fortement corrélés), elle constitue une variable multidimensionnelle. Nous avons donc mis en œuvre un modèle d'analyse de covariance multivariée permettant d'expliquer les performances en fonction des six attributs quantitatifs. Une démarche de sélection descendante a permis de mettre en évidence les attributs pouvant avoir un impact significatif sur les performances. Nous avons choisi d'illustrer ces résultats sur le score de performance $P@100$, présentés dans le Tableau 3.12.

Attributs	**Paramètres estimés (s.e.)**	***p-value***
Longueur en concepts	0,04 (0,01)	††
Spécificité terme-document	0,74 (0,26)	††

Tableau 3.12 – Résultats de la modélisation du score de performance $P@100$ en fonction des attributs des requêtes (par une MANCOVA) : estimation des paramètres associés aux attributs significatifs (et erreur standard, *s.e.*) et *p-value* (*ns* : †† : $0,001 < p\text{-}value < 0,01$;)

La Figure 3.3 montre les scores moyens de la performance de recherche selon les mesures de la MAP, P@10, P@20 et P@100 des différentes collections médicales.

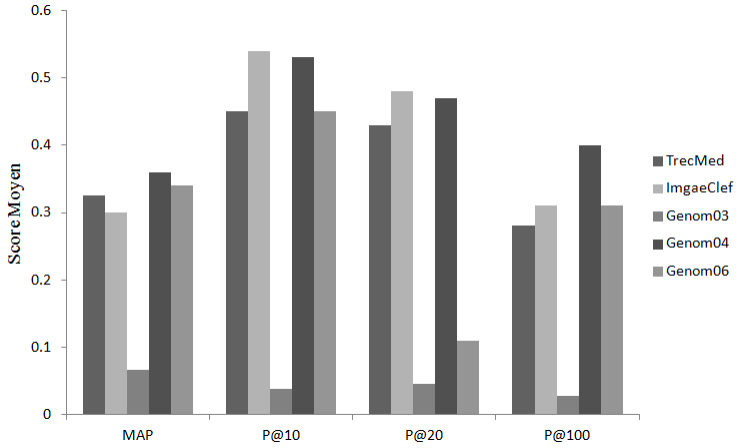


FIGURE 3.3 – Scores moyens des performances de recherche par tâche médicale.

Suite à ces analyses, nous avons pu mettre en évidence que les requêtes sont d'autant plus performantes qu'elles contiennent un nombre important de concepts (le score $P@100$ augmente en moyenne de $0,04$ point par concept supplémentaire). Toutefois, cet effet du type de requête est à mettre en relation avec la spécificité terme-document de la requête. En effet, les requêtes cliniques sont d'autant plus performantes qu'elles sont spécifiques. En revanche, des requêtes non-cliniques auront tendance à être moins performantes quand leur spécificité augmente.

Ces résultats montrent globalement que la recherche de cas pathologiques ou l'étude de cohorte constitue une tâche plus difficile qu'une recherche *ad hoc* sur des gènes. Nous pouvons conclure que :

1. La longueur de la requête favorise la clarification de la requête et par conséquent sa "facilité" ; ceci est à la base même des techniques d'expansion de requêtes qui ont montré leur efficacité en RI ;

2. Le type de la requête joue un rôle important conjointement à certains attributs caractéristiques du besoin en information des experts. Plus précisément, les requêtes cliniques,

même si elles sont naturellement longues, demeurent difficiles si leur vocabulaire est général, non suffisamment caractéristiques du cas pathologique considéré. En effet, les différences, même infimes, entre description du cas à travers la requête et cas retourné à partir de la base de cas, amèneront les experts à un jugement de pertinence négatif ;

3. En revanche, un besoin en information non clinique, ciblant des termes rares dans la collection, est plus difficile à satisfaire.

5 Caractérisation des requêtes PICO

Dans le cadre de nos analyses statistiques exploratoires, nous avons étudié les particularités des requêtes PICO, afin de les caractériser par rapport aux différentes tâches médicales. Nous avons utilisé les même attributs que pour nos analyses précédentes. Nous présentons dans ce qui suit les résultats des analyses.

5.1 Analyses descriptives

Nous menons une analyse descriptive afin d'identifier et mettre en avant les caractéristiques des requêtes PICO, issues de la collection CLIREC et les distinguer par rapport à d'autres besoins en information exprimés dans le cadre d'autres tâches médicales. La Figure 3.5 montre la distribution des six attributs de requêtes par collection médicale, présentés par des boîtes à moustache. Nous avons mené des analyses de variance sous forme de tests de Kruskal-Wallis non-paramétriques pour comparer les moyennes des attributs et distinguer la recherche médicale PICO des autres recherches issues de différentes tâches médicales (en terme de $p\text{-}value < 0,05$).
Dans la Figure 3.4.(a) et 3.4.(b), nous remarquons une variabilité de la distribution des termes et des concepts. Les deux attributs de la longueur sont significativement différents à travers toutes les collections malgré le fait qu'ils représentent tous le besoin en information des experts. Les valeurs les plus élevées en nombre de termes et en nombre de concepts sont observées pour les collections ImageCLEF et CLIREC (en moyenne 31 termes et 6 concepts pour CLEF, 15 termes et 5 concepts pour CLIREC). En effet, les cliniciens expriment des requêtes longues : ImageCLEF pour décrire les cas de patients et CLIREC pour décrire les questions PICO avec les différentes facettes.
Les Figures 3.4.(c) et 3.4.(d) représentent respectivement la distribution des scores de spécificité terme-document et spécificité hiérarchique basée sur la terminologie MeSH. Nous remarquons que la distribution des scores de spécificité terme-document pour les requêtes PICO est plus élevée avec les plus grandes valeurs par rapport aux autres collections (une moyenne $0,5188$), par contre une faible distribution avec les valeurs les moins importantes pour la spécificité hiérarchique comparée aux autres collections. Les premières conclusions montrent que les experts formulent des requêtes PICO utilisant des termes se caractérisant par une forte distribution dans les documents PubMed de la littérature médicale.
De plus, la faible valeur de la spécificité hiérarchique montre que les experts formulent des requêtes PICO avec une description de ces facettes (P, I, C et O) en se basant sur leurs connaissances du domaine, sans forcément se référer à une terminologie fiable du domaine médical. La facette de la spécificité hiérarchique est basée sur deux facteurs principaux : la longueur de la requête en nombre de concepts et le niveau de ces concepts dans la terminologie.
Ayant conclu que les requêtes PICO sont longues en nombre de concepts, nous pouvons constater que les requêtes malgré cette longueur ne sont pas spécifiques au contexte de la recherche.

En analysant les attributs de la clarté à partir des Figures 3.4.(e) et 3.4.(f), nous pouvons conclure que les requêtes PICO se caractérisent par une plus faible variabilité par rapport aux autres collections pour la pertinence basée sur le sujet et par une moyenne importante de valeurs. Cela montre que les requêtes PICO sont relativement claires.
Concernant la clarté basée sur la pertinence, la collection CLIREC se distingue des autres collections par les scores de clarté les plus importants. Cet attribut suppose que la requête est claire si elle partage le plus grand nombre de concepts avec les documents pertinents jugés par les experts. Cela s'explique par le fait que les cliniciens utilisent des concepts pertinents correspondant aux meilleurs documents répondant au mieux aux questions cliniques PICO.
Pour résumer, les requêtes PICO sont longues, les termes utilisés sont bien distribués dans la collection PubMed et elles sont claires par rapport à la présence des concepts utilisés dans les documents pertinents. La recherche PICO favorise l'utilisation des concepts qui reflètent les connaissances du domaine des experts. Toutefois, ces concepts ne sont pas assez spécifiques au contexte de la recherche où ils sont ambigus. En d'autres termes, il peut y avoir des concepts désignant un élément PICO mais qui peuvent être sémantiquement liés à un autre élément PICO de la requête, source ainsi d'ambiguïté. Se référer à leurs connaissances du domaine ne suffit donc pas.

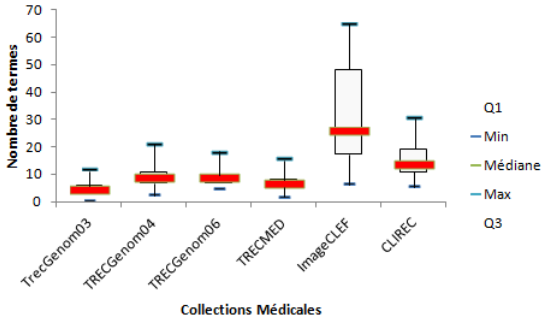
5.2 Analyses de corrélations entre attributs de requêtes PICO

Pour étudier les corrélations entre les attributs de requêtes, nous avons calculé le coefficient de corrélation de *Spearman* ρ entre les 6 attributs quantitatifs. Nous représentons uniquement les corrélations significatives entre paires d'attributs dans le Tableau 3.13.
Nous observons la plupart de corrélations avec les attributs de la longueur $LgT(Q)$ et $LgC(Q)$. Une forte corrélation positive entre la longueur en termes et la longueur en concepts (p-value<0,0001). De fortes corrélations positives significatives entre la longueur en termes et la spécificité terme-document (p-value<0,0001) d'une part, et la longueur en concepts et la spécificité terme-document (p-value<0,01) d'autre part.
La première corrélation, $LgT(Q)$ et $DSpe(Q)$ s'explique par le fait que, selon la formule (3.1), plus la requête est longue, plus sa spécificité terme-document est élevée. Ceci est justifié par le besoin exprimé sous forme de question clinique basée sur des termes bien distribués dans les documents. Toutefois, les deux attributs de la longueur sont négativement corrélés à la spécificité hiérarchique. Cela montre que l'utilisation de termes qui désignent les concepts de la terminologie *MeSH* ne sont pas appropriés pour désigner spécifiquement chacun des éléments PICO de la question.

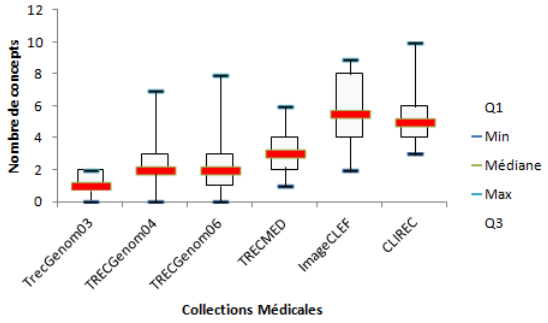
De plus, nous notons que les deux attributs de la longueur sont corrélés négativement à la clarté basée sur le sujet, avec une corrélation plus significative avec la longueur en nombre de concepts (*p-value<0,0005*). Cela peut être expliqué par le fait que les cliniciens qui formulent leurs requêtes PICO n'utilisent pas des termes qui désignent des concepts pertinents de la terminologie *MeSH*.

Enfin, la spécificité hiérarchique est positivement corrélée à la clarté basée sur la pertinence (*p-value<0,01*). Il en résulte que les concepts médicaux sont importants pour traduire le besoin PICO sans ambiguïté. La Figure 3.5 résume les différentes corrélations entre les attributs pour les questions PICO.

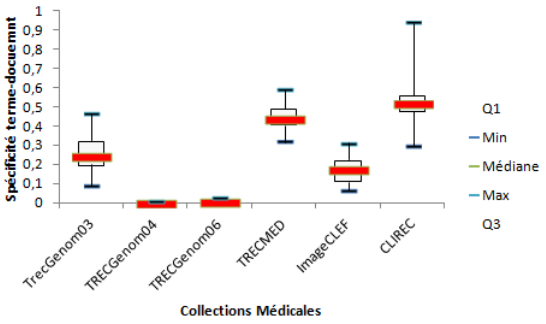
Pour synthétiser, les différentes analyses statistiques effectuées sur les requêtes cliniques PICO ont révélé que :



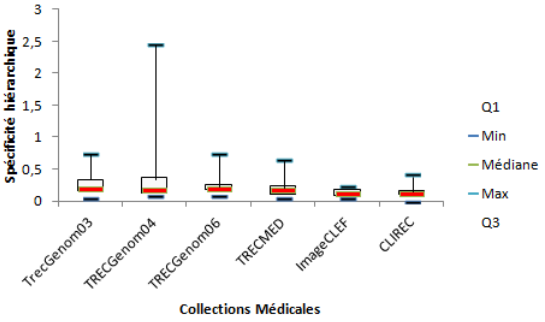
(a) Longueur en termes



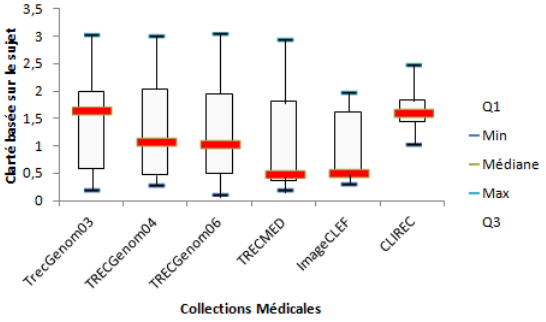
(b) Longueur en concepts



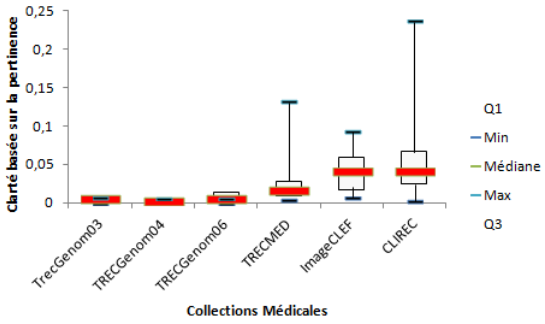
(c) Spécificité terme-document



(d) Spécificité hiérarchique



(e) Clarté basée sur le sujet



(f) Clarté basée sur la pertinence

FIGURE 3.4 – Distribution des attributs de requêtes par collection

— Les requêtes PICO sont longues en termes et en concepts. Plus particulièrement, les requêtes PICO se distinguent des autres requêtes médicales issues des campagnes TREC et CLEF par un nombre plus élevé de concepts. Cela s'explique par la nature de ces questions dont le but est d'identifier les meilleures preuves médicales, ce qui pousse les experts à utiliser d'une manière intensive les connaissances médicales et les ressources disponibles.

— Les scores de spécificité terme-document sont importants (les experts formulent des requêtes PICO utilisant des termes se caractérisant par une forte distribution dans les documents PubMed), et les scores de spécificité hiérarchique sont faibles (les experts formulent des requêtes PICO avec une description de ces facettes (P, I, C et O) en se basant sur leurs connaissances du domaine sans forcément se référer à une terminologie du domaine médical de référence pertinente) : les requêtes PICO sont longues en nombre de concepts, nous pouvons conclure

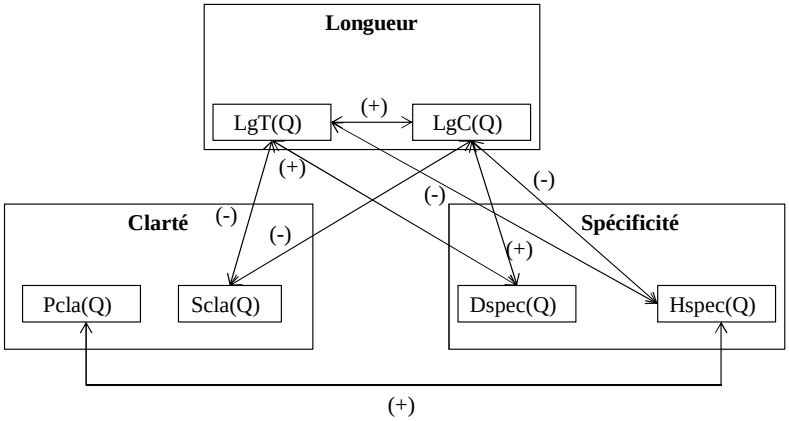


FIGURE 3.5 – Analyses de corrélations des attributs de requêtes PICO

Facette 1	**Facette 2**	ρ	***p-value***
$LgT(Q)$	$LgC(Q)$	$0,2156$	$< 0,0001$
$LgT(Q)$	$DSpe(Q)$	$0,2388$	$< 0,0001$
$LgT(Q)$	$HSpe(Q)$	$-0,1460$	$< 0,01$
$LgT(Q)$	$SCla(Q)$	$-0,1354$	$< 0,01$
$LgC(Q)$	$DSpe(Q)$	$0,1474$	$< 0,005$
$LgC(Q)$	$HSpe(Q)$	$-0,1041$	$< 0,05$
$LgC(Q)$	$SCla(Q)$	$-0,1805$	$< 0,0005$
$HSpe(Q)$	$PCla(Q)$	$0,1492$	$< 0,005$

Tableau 3.13 – Résultats de corrélations deux à deux entre les attributs de requêtes relatives à la collection CLIREC ($N = 423$)

que les requêtes, bien que longues, ne sont pas spécifiques au contexte de recherche.

— Les scores de clarté basée sur la pertinence sont élevés : les cliniciens utilisent des concepts pertinents qui permettent d'identifier les meilleurs documents qui répondent au mieux aux questions cliniques PICO.

— Corrélation positive significative de *LgT* et *LgC* avec *DSpe* : plus la requête est longue, plus sa spécificité terme-document est élevée et importante. L'hypothèse de ce score est la suivante : moins les documents sont concernés par les termes de la requête, plus le sujet de la requête est spécifique.

— Corrélation négative de *LgT* et *LgC* avec *HSpe* : l'utilisation des termes qui désignent les concepts de la terminologie MeSH ne sont pas appropriés pour désigner spécifiquement chacun des éléments PICO de la question. Les requêtes longues ne sont pas forcément spécifiques si les termes qui désignent les concepts ne sont pas appropriés pour le besoin PICO.

— Corrélation négative de *LgT* et *LgC* avec *SClar* : les cliniciens qui formulent leurs requêtes PICO n'utilisent pas des termes qui désignent des concepts pertinents de la terminologie MeSH.

— Corrélation positive significative entre *HSpe* et *PClar* : les concepts médicaux sont significatifs

pour traduire le besoin PICO sans ambiguïté ou d'une manière claire. Le sens du concept de la requête est bien traduit/représenté dans les documents pertinents.

La Figure 3.6 donne la distribution des deux facettes de la longueur pour les sous-graphes PICO.

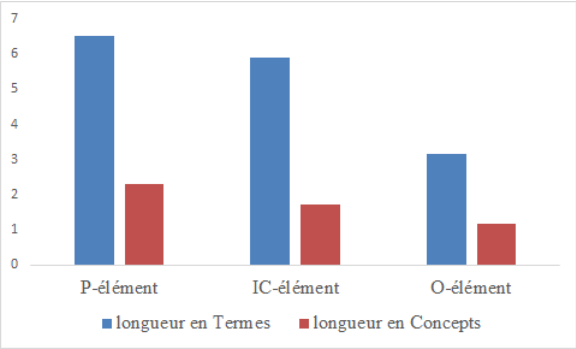


FIGURE 3.6 – Distribution des deux facettes de la longueur pour les sous-graphes PICO

6 Bilan et conclusion

Nous avons réalisé une étude statistique exploratoire sur les requêtes exprimées par les experts biomédicaux dans le cadre des campagnes d'évaluation TREC et CLEF. Les résultats de notre étude donnent un aperçu sur les spécificités des requêtes d'experts selon les différentes tâches. Trois attributs impactent les résultats de recherche, plus spécifiquement la longueur en termes, le score de clarté basé sur le sujet et la spécificité terme-document en fonction du type clinique ou non de la requête. Ces résultats suggèrent le besoin de contextualiser les modèles de RI médicale à la tâche. Plus précisément, un besoin de clarification et spécification par expansion/reformulation de requête serait appropriée pour les requêtes cliniques. Au delà, en effectuant un croisement avec les travaux de l'état de l'art, il en ressort un besoin de personnaliser la recherche, selon le niveau d'expertise des utilisateurs. Pour asseoir cette hypothèse, nous envisageons dans un futur proche de mener une analyse exploratoire des besoins en information des experts vs. novices du domaine médical.
Dans un premier temps, nous avons défini un ensemble d'attributs de requêtes au sein d'une collection de tâches médicales et nous avons analysé les corrélations des attributs de requêtes. Dans un deuxième temps, nous nous sommes intéressés à étudier l'impact des attributs de requêtes sur la performance de recherche. Dans ces deux contributions, nous avons défini les mêmes facettes des attributs et nous avons utilisé les mêmes collections d'analyse. Cette modélisation peut être utilisée dans le cadre de l'évaluation expérimentale moyennant d'autres collections ainsi que pour la formalisation de modèles de RI médicaux contextuels ou personnalisés.
De plus, les principales conclusions issues de ces analyses mettent en évidence les spécificités des requêtes médicales qui varient selon la nature de la tâche. Cela implique plusieurs recommandations pour les modèles de RI ainsi que les systèmes de RI ou de prise de décision médicale pour exploiter la sémantique, l'information contextuelle et personnalisée pour améliorer la performance et les résultats de recherche.

Nous avons analysé également les requêtes PICO, issues de la campagne CLIREC. Les principales conclusions soulignent la particularité des questions expertes PICO par rapport aux autres tâches

médicales, du point de vue des caractéristiques des attributs. Cela montre la nécessité de proposer des modèles et approches de RI spécifiques aux requêtes PICO.

Les études de l'état de l'art dans le cadre de la médecine basée sur les faits ont montré que la plupart des médecins pratiquent l'EBM pendant leur travail. Cependant, l'EBM présente plusieurs problèmes qui empêchent son utilisation dans la pratique. À ce titre, dans le chapitre suivant, nous présentons nos contributions portant sur la médecine basée sur les faits, où les utilisateurs se basent sur des requêtes PICO pour formuler leur besoin en information. Un premier volet détaille notre approche d'expansion de requêtes, où nous traitons le problème lié à la recherche des meilleures preuves cliniques qui correspondent aux facettes PICO. Un deuxième volet propose un modèle d'appariement sémantique basé sur une algorithme d'agrégation de scores de pertinence pour mieux répondre aux questions cliniques PICO.

Chapitre 4

Évaluation des requêtes cliniques : Modèles sémantiques pour mieux répondre aux questions PICO

Sommaire

1 Introduction . . . 123
2 Motivations et hypothèses . . . 125
3 Aperçu général de nos contributions . . . 127
4 Processus d'élicitation sémantiques des requêtes PICO . . . 128
4.1 Méthode de génération de graphes . . . 128
4.2 Identification de concepts les plus pertinents . . . 130
5 Évaluation des requêtes cliniques . . . 133
5.1 Expansion sémantique des requêtes cliniques . . . 133
5.2 Modèle d'ordonnancement basé sur un opérateur d'agrégation prioritaire des scores . . . 135
5.2.1 Cadre général . . . 135
5.2.2 Calcul de scores de pertinence des documents . . . 136
6 Évaluation expérimentale . . . 137
6.1 Données expérimentales . . . 137
6.2 Évaluation de l'approche d'expansion de requêtes . . . 137
6.2.1 Métriques d'évaluation et modèles de référence . . . 137
6.2.2 Ajustement des paramètres . . . 138
6.2.3 Analyse de l'efficacité de la recherche : évaluation comparative . . . 140
6.2.4 Discussion . . . 140
6.3 Évaluation du modèle d'ordonnancement de documents . . . 141
6.3.1 Objectifs . . . 141
6.3.2 Métriques et protocole d'évaluation . . . 141
6.3.3 Efficacité du modèle de pertinence basé sur l'agrégation des scores 142
6.3.4 Analyse de la robustesse du modèle d'agrégation de pertinence . . . 143
6.3.4.1 Analyse globale . . . 144
6.3.4.2 Analyse au niveau requête . . . 146

7 Bilan et conclusion 149

1 Introduction

De nombreuses études ont clairement montré que la recherche d'information médicale est largement pratiquée aussi bien par des experts que par des novices (Fox et Duggan, 2013; Zhang, 2014). Plus spécifiquement, la recherche d'information médicale pratiquée par les experts est une recherche verticale basée sur l'usage de ressources, comme les dossiers médicaux de patients ou la littérature scientifique médicale (*e.g.,* MEDLINE [1], Cochrane [2]), en vue de répondre à différents objectifs comme l'aide au diagnostic, l'aide à la prescription médicale ou encore la recherche clinique.

Ce type de besoins en information induit en grande partie l'évaluation de requêtes cliniques qui ont pour objectif de sélectionner, à partir de la littérature scientifique du domaine, des procédés cliniques et/ou des preuves d'études systématiques (Yang *et al.*, 2011). Plus globalement, les tâches de recherche d'information cliniques expertes sont généralement conduites par les cliniciens dans le cadre de la médecine basée sur les faits connue sous l'acronyme EBM (*Evidence-Based Medicine*) (Sackett *et al.*, 1996). Cette dernière consiste à utiliser de manière rigoureuse, explicite et judicieuse, les preuves scientifiques les plus récentes et plus pertinentes lors de la prise de décision concernant les soins à prodiguer à chaque patient. Sa pratique implique que l'on conjugue l'expertise clinique individuelle avec les meilleures preuves cliniques externes obtenues actuellement par la recherche systématique (Sackett *et al.*, 1996).

La recherche de ces meilleures preuves d'études cliniques, à partir de la littérature scientifique, est, à juste titre, l'objet de notre contribution présentée dans ce chapitre. Cela suppose d'abord la formulation d'un besoin en information clinique. Selon l'approche EBM, un moyen qui a été proposé aux experts en vue de clarifier leur besoin est de structurer leur requête selon la forme PICO, à savoir : Patient/Problem (P), Intervention (I), Comparison (C) et Outcome (O), appelés les éléments ou facettes PICO (Schardt *et al.*, 2007) détaillés dans le précédent chapitre.

En plus du verrou largement reconnu lié à l'ambiguité des expressions médicales et acronymes (Trieschnigg, 2010), l'évaluation de ce type de requêtes pose deux difficultés supplémentaires et non triviales à surmonter : (1) considérer la structure de la requête en facettes dans le processus d'appariement avec des documents qui ne sont pas ainsi structurés ; et (2) considérer leur importance relative dans l'estimation du score de pertinence. À notre connaissance, c'est un champ d'investigation peu exploré ; on recense en effet peu de travaux qui ont abordé spécifiquement le problème d'évaluation des questions cliniques PICO (Boudin *et al.*, 2010c; Boudin *et al.*, 2010b; Demner-Fushman et Lin, 2007). Dans l'ensemble des travaux précédents, une étape préliminaire à la recherche est la détection des facettes PICO dans les documents. Au niveau de l'appariement requête-document, les auteurs dans (Demner-Fushman et Lin, 2007), se sont basés sur une approche d'appariement sémantique entre les types sémantiques extraits d'*UMLS* de chacune des facettes détectées dans les requêtes et les documents. Cependant, cette méthode ne prend pas en compte l'importance des facettes PICO dans le texte, lors du calcul des scores de pertinence.

Cependant, plusieurs problèmes empêchent le développement et l'implémentation des meilleures preuves dans la pratique clinique (Francke *et al.*, 2008). Ces problèmes proviennent principalement de :

1. https ://www.nlm.nih.gov/bsd/pmresources.html accessible à l'aide PubMed
2. http ://www.ncbi.nlm.nih.gov/pubmed, Cochrane http ://www.cochranelibrary.com/

1. La complexité d'exprimer une requête précise, spécifique au contexte de la recherche permettant d'identifier facilement les preuves pertinentes ;
2. L'absence du niveau de compétence qui peut être utilisé pour effectuer l'évaluation des preuves.

Nous nous intéressons dans le contexte de nos recherches au premier problème.

Dans Boudin et al. (Boudin *et al.*, 2010c), les auteurs ont proposé un modèle d'appariement qui considère l'importance de chaque facette PICO pour calculer le score de pertinence des documents. Toutefois, les scores d'importance sont calculés d'une manière statique, sur la base de la distribution des mots appartenant à chaque facette sur l'ensemble de la collection de documents.

Dans ce contexte de recherche, nous proposons un modèle d'appariement requête PICO-document, qui, à la différence des précédents travaux, (1) ne requiert pas l'identification préalable des facettes P, I, C et O dans les documents ; et (2) utilise un opérateur d'agrégation prioritaire (Da Costa Pereira *et al.*, 2009) dans le calcul des scores d'appariement requête-document en personnalisant les poids de chaque facette selon la requête et le document en cours d'évaluation. Comme dans (Demner-Fushman et Lin, 2007), nous privilégions une représentation sémantique des requêtes et documents, qui est cependant basée sur la génération de graphes sémantiques par facette.

Ainsi, notre objectif est de contribuer à ce domaine de recherche qui reste insuffisamment exploré avec l'espoir de concevoir les systèmes de RI qui assistent les cliniciens dans leur pratique quotidienne de l'EBM. Les travaux pionniers de la littérature (Demner-Fushman et Lin, 2007; Boudin *et al.*, 2010c; Boudin *et al.*, 2010b) nous ont permis de conclure que les différentes techniques proposées reposent sur deux étapes principales, à savoir : (1) la détection des facettes PICO à partir des requêtes et des documents compris dans les entrepôts scientifiques médicaux ; et (2) l'intégration des éléments PICO dans le modèle d'appariement document-requête.
Cependant, nous proposons d'aborder la problématique différemment de la manière suivante : (1) faciliter les conditions d'identification des éléments PICO à partir des documents ; et (2) résumer la formulation de la question clinique en mettant en relief une image sémantique de chaque facette PICO.

À ce titre, nous proposons d'étendre séparément chaque facette de la question en utilisant les concepts extraits des documents les mieux classés. Ce classement est donné au préalable par un modèle de RI. Notre intuition est que ces documents comportent des indices de contexte utiles qui permettent de distinguer les documents pertinents des documents non pertinents en se basant sur ces différentes facettes.

La suite du chapitre est organisée comme suit : dans la section 2, nous donnons quelques motivations et les hypothèses de recherche. Ensuite, nous détaillons l'approche de l'élicitation sémantique des questions cliniques PICO dans la section 3, où nous proposons deux algorithmes : (1) un algorithme de génération de graphes sémantiques conceptuels, et (2) processus d'identification de concepts qui représentent au mieux les preuves cliniques entre la question et les documents. La section 4 détaille nos méthodes d'évaluation des requêtes cliniques, composée de deux sous-sections : la section 4.1 décrit l'approche d'expansion sémantique conceptuelle des requêtes ; et la section 4.2 détaille le modèle d'ordonnancement de documents basé sur l'agrégation prioritaire des scores pour calculer le score de pertinence des documents. Enfin, l'évaluation expérimentale menée dans l'objectif d'évaluer

la pertinence de nos contributions est détaillée dans la section 5. Nous présentons un bilan et concluons dans la section 6.

2 Motivations et hypothèses

Notre objectif est de répondre au mieux aux questions cliniques expertes. Plus spécifiquement, dans le cadre de la médecine basée sur les preuves, identifier la preuve médicale pertinente pour répondre à des questions PICO est une tâche difficile (Boudin *et al.*, 2010b; Schardt *et al.*, 2007).

En effet, la RI basée sur les requêtes PICO repose sur deux étapes principales : la première est d'identifier les éléments P, I, C et O à partir du texte des requêtes et des documents, une deuxième étape est d'intégrer ces éléments dans un modèle de RI afin de retourner un ensemble de documents qui constituent la preuve médicale, qui aidera les médecins à prendre une décision clinique.

Pour atteindre cet objectif, nous nous sommes basés sur les hypothèses suivantes :

— **Hypothèse 1.** Dans un cadre de recherche d'information clinique, plus la requête est courte en nombre de termes avec une faible spécificité hiérarchique (se réfère à la spécificité "*is-a*" extraite de la terminologie médicale), plus la requête est difficile (Tamine *et al.*, 2015).

— **Hypothèse 2.** Un document est d'autant plus pertinent qu'il s'apparie avec plus de facettes de la requête (Boudin *et al.*, 2010c; Demner-Fushman et Lin, 2007).

— **Hypothèse 3.** En phase d'évaluation de la pertinence des résultats, l'expert médical n'accorde pas la même importance à l'adéquation des facettes. La facette I/C est plus importante que la facette P qui, à son tour, est plus importante que la facette O (Weifield et Finkelstein, 1996; Boudin *et al.*, 2010c).

Avec ces hypothèses, et selon les conclusions des travaux antérieurs, nos motivations sont doubles :

— Pour l'approche d'expansion des requêtes PICO utilisant les termes les plus spécifiques issus des ressources sémantiques est justifiée. Afin d'alléger le sens du sujet attendu dû à l'expansion "aveugle" basée sur les concepts, on favorise l'utilisation de l'expansion guidée par les meilleurs concepts qui représentent le contexte local sémantique des questions issues d'une première recherche basée sur les termes.

— Pour le modèle d'appariement basé sur l'agrégation prioritaire de scores, nous avons intégré la représentation sémantique conceptuelle dans le calcul des scores de pertinence des documents. En effet, le calcul des scores selon l'ordre de priorité met en évidence le contexte de chaque facette PICO de la requête à l'égard du document. Par ailleurs, les travaux traitant les questions PICO n'ont pas intégré l'ordre d'importance des différentes facettes PICO dans le calcul des scores de pertinence. Nous avons donc proposé de calculer des poids dynamiques, qui représentent l'importance de chaque facette PICO et de les intégrer dans le modèle d'ordonnancement de documents.

Nos contributions se déclinent en quatre volets : (1) nous proposons une approche de génération de graphes sémantiques conceptuels selon chacun des éléments PICO ; (2) nous introduisons une approche spécifique d'expansion de requêtes PICO, en utilisant l'information sémantique extraite des documents les mieux classés par une première recherche selon chaque facette ; (3) nous proposons un modèle d'appariement sémantique qui repose sur une approche d'agrégation prioritaire des scores PICO, et (4) nous évaluons notre approche en utilisant une collection expérimentale CLIREC fournie

par (Boudin *et al.*, 2010b). Nous considérons une requête clinique ; notre objectif est de mieux répondre aux questions cliniques expertes PICO, et d'étudier comment considérer chacune des facettes de la structure PICO pour identifier et sélectionner les meilleures preuves adaptées à chacune des facettes de la question. Dans cette optique, les contributions majeures de ces travaux sont les suivantes :

1. Nous introduisons une approche de génération de graphes qui représente d'une manière efficace les facettes de la requête, en identifiant les parties sémantiques de l'information extraites à partir des meilleurs documents retournés, suite à une phase de recherche effectuée au préalable, selon chaque facette de la question.
2. Nous proposons une approche d'expansion de requêtes qui s'appuie sur une méthode sélective de concepts médicaux. Pour la sélection des concepts, nous introduisons un algorithme spécifique permettant de proposer les concepts pertinents qui représentent au mieux l'information sémantique commune entre la requête et le document.
3. Nous proposons un modèle d'ordonnancement utilisant un opérateur de calcul de scores de pertinence des documents, basé sur un ordre de priorité, qui s'appuie sur et complète un modèle d'ordonnancement de documents conçu spécifiquement pour répondre aux questions cliniques PICO.

Le Tableau 4.1 présente les notations utilisées pour les différents algorithmes proposés et pour le reste du chapitre.

Notation	**Description**
Q	une requête textuelle
d	un document
c	un concept actif
D_N^*	N top documents retournés par une première recherche
Q_{PICO}	requête annotée avec P, IC et O
Q_P	sous-requête correspondant au facette P
Q_{IC}	sous-requête correspondant au facette IC
Q_O	sous-requête correspondant au facette O
T	Terminologie MeSH
C	collection de documents
N_d	la liste de documents les mieux classés D_N^* de la collection C
N_c	ensemble de concepts
$MaxLevel$	niveau maximal considéré pour l'expansion
G_P	sous graphe correspondant au facette P
G_{IC}	sous graphe correspondant au facette IC
G_O	sous graphe correspondant au facette O
$Concepts(Q_P)$	concepts extraits correspondant à Q_P
$Concepts(Q_{IC})$	concepts extraits correspondant à Q_{IC}
$Concepts(Q_O)$	concepts extraits correspondant à Q_O
$HypG(G_x)$	fonction qui génère les graphes sémantiques, $x \in \{P, IC, O\}$
$Expand(G_x)$	fonction d'expansion de la requête avec les concepts candidats C_{expand}
$Score(c)$	score d'importance du concept actif c de la requête
$Normalisé(Score(c))$	score normalisé du concept c
$Entrée(Q_x^e)$	termes qui correspondent aux termes des entrées préférées dans MeSH selon la catégories sémantique x

Tableau 4.1 – Notations utilisées pour les différents algorithmes.

3 Aperçu général de nos contributions

Soit une requête Q avec les annotations PICO associées donnant lieu aux sous-requêtes Q_P, Q_{IC} et Q_O, manuellement ou automatiquement identifiées. Comme dans de précédents travaux (Boudin *et al.*, 2010c), nous considérons les facettes I et C regroupées étant donné qu'elles sont associées au même type sémantique. Notre objectif est de sélectionner les documents d qui sont pertinents au besoin en information expert véhiculé par la requête Q et ce en accord avec les facettes P, I/C et O. Nous avons proposé un algorithme de génération de graphes sémantiques conceptuels exploitant la ressource MeSH. Nous avons ensuite proposé un processus d'appariement sémantique document-requête afin d'identifier les meilleurs concepts associés, en appliquant un algorithme de propagation de scores concept-document. Ces scores ont permis de sélectionner les meilleurs concepts associés aux requêtes et aux documents. Nous décrivons ci-dessous les trois premières étapes, savoir :

1. **Étape 1** : Recherche d'une liste de documents pertinents utilisant un modèle de base de RI (Robertson et Sparck Jones, 1988). Cette étape repose sur une méthode d'appariement préliminaire document-requête basé sur les mots, qui retourne une liste de documents pertinents candidats ;
2. **Étape 2** : Représentation sémantique sous forme de graphe des éléments PICO de requêtes, permettant de représenter une requête textuelle annotée en graphes sémantiques soulignant les concepts extraits de chacun des éléments PICO. En effet, à partir des sous-requêtes facettes Q_P, Q_{IC} et Q_O, un algorithme génère les graphes sémantiques de requêtes Q_P^c, Q_{IC}^c et Q_O^c, en effectuant une extraction de concepts MeSH pour chacune des facettes en remontant de proche en proche concept commun de la hiérarchie de MeSH ;
3. **Étape 3** : Appariement sémantique document-requête sur la base des facettes PICO, basé sur les concepts actifs extraits, en commun entre les requêtes et les documents les mieux classés. Plus précisément, pour chaque arbre sous-requête associé à une facette Q_P^c, Q_{IC}^c et Q_O^c et chaque document d résultat de l'étape 1, identifier les N_c meilleurs concepts associés en appliquant un algorithme de propagation de scores d'appariement concept-document par accumulation des scores des concepts associés au document d jusqu'aux concepts feuilles de MeSH.

Comme indiqué précédemment, notre objectif est de répondre au mieux aux questions cliniques PICO. En d'autres termes, sélectionner les meilleures preuves cliniques (documents) permettant de satisfaire le besoin en information exprimé via la question médicale PICO. Pour atteindre cet objectif, et à partir des N_c meilleurs concepts identifiés, nous avons proposé deux approches, à savoir :

— Expansion conceptuelle sélective des requêtes, qui permet d'étendre chaque partie PICO de la requête avec les entrées préférées des concepts appartenant au document et à la requête. Nous avons conduit ensuite une évaluation des requêtes étendues, qui permet de retourner une liste finale de documents les mieux classés comme réponse aux questions cliniques.

— Application d'un opérateur d'agrégation prioritaire (Da Costa Pereira *et al.*, 2009) afin de calculer le score global de pertinence du document d par combinaison de ses scores partiels d'appariement avec les sous-requêtes facettes Q_P, Q_{IC} et Q_O.

4 Processus d'élicitation sémantiques des requêtes PICO

Afin d'identifier explicitement les parties sémantiques des questions PICO, et de mettre en évidence le contexte de recherche, nous avons proposé une approche d'élicitation sémantique. Notre méthode, comme décrite dans les sections qui suivent, est sous forme de deux algorithmes : le premier ayant pour objectif la génération de graphes sémantiques conceptuels selon chacune des facettes P, I, C et O. Le deuxième algorithme vient compléter le premier, et permet d'identifier à partie de ces graphes les meilleurs concepts pertinents permettant de représenter au mieux les preuves cliniques. Comme indiqué auparavant, ces concepts ont ensuite été utilisés dans l'approche d'expansion de requêtes ainsi que le modèle d'ordonnancement. La Figure 4.1 donne un aperçu du processus de représentation sémantique.

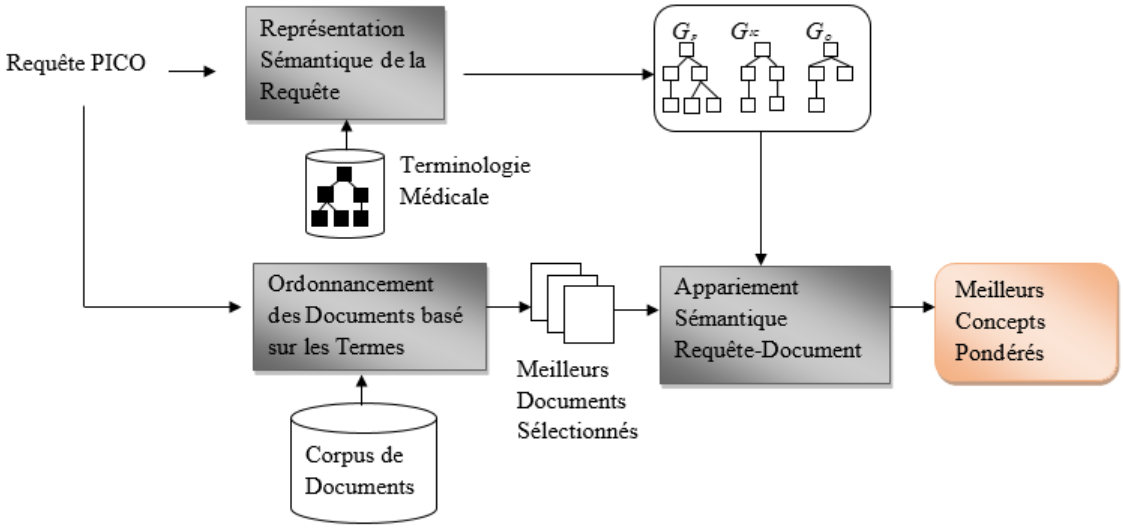


FIGURE 4.1 – Aperçu du processus d'élicitation sémantique de requêtes PICO.

Nous détaillons dans ce qui suit la méthode de génération de graphes ainsi que le processus d'identification des meilleurs concepts pertinents.

4.1 Méthode de génération de graphes

Nous avons proposé une représentation sémantique des requêtes PICO basée sur l'extraction de concepts médicaux à partir de la terminologie MeSH. Nous nous basons sur un algorithme de génération de graphes sémantiques décrit par l'Algorithme 1.

Pour une requête textuelle donnée Q, avec l'annotation associée Q_{PICO}, les différentes facettes de la requête Q_P, Q_{IC} et Q_O, ainsi que la liste N_d de documents les mieux classés D_N^* (de la collection C) retournés comme réponse à la requête Q, l'algorithme commence par construire des graphes sémantiques G_P, G_{IC} et G_O après un ensemble d'opérations. Ces opérations sont les suivantes : (1) extraction des concepts actifs de chaque facette de la requête en utilisant la terminologie MeSH avec la méthode d'extraction des concepts (*eg.,* (Dinh et Tamine, 2011a; Ruch, 2006)) ; chacun des concepts actifs c est considéré à un niveau relatif 0 et possède un score d'importance $Score(c)$ qui souligne la similarité de vraisemblance entre les entrées préférées de concepts et les termes

Algorithme 1 Génération de graphes et ordonnancement de documents

Entrées: $Q, Q_{PICO}, T, N_d, N_c, MaxLevel$
Sortie: G_P, G_{IC}, G_O, D_N^*

1: {Recherche initiale}
2: $D_N^* \leftarrow Top_D(Q, N_d, C)$;
3: {Construction du graphe de la requête}
4: $Q_P \leftarrow Substr(Q, P)$;
5: $Q_{IC} \leftarrow Substr(Q, IC)$;
6: $Q_O \leftarrow Substr(Q, O)$;
7: $Concepts(Q_P) \leftarrow Extract(Q_P, T)$;
8: $G_P \leftarrow HypG(Concepts(Q_P), T)$;
9: $Concepts(Q_{IC}) \leftarrow Extract(Q_{IC}, T)$;
10: $G_{IC} \leftarrow HypG(Concepts(Q_{IC}), T)$;
11: $Concepts(Q_O) \leftarrow Extract(Q_O, T)$;
12: $G_O \leftarrow HypG(Concepts(Q_O), T)$;
13: $Q_P^e \leftarrow Expand(G_P)$;
14: $Q_{IC}^e \leftarrow Expand(G_{IC})$;
15: $Q_O^e \leftarrow Expand(G_O)$;
16: $Words(Q^e) \leftarrow Words(Q) \cup Entries(Q_P^e) \cup Entries(Q_{IC}^e) \cup Entries(Q_O^e)$;
17: {Recherche finale}
18: $D_N^* \leftarrow Top_D(Q^e, N_d, C)$;

de la requête ; (2) construction des graphes associés G_P, G_{IC} et G_O (basées respectivement sur $Concepts(Q_P)$, $Concepts(Q_{IC})$ et $Concepts(Q_O)$) en joignant les concepts actifs aux hyponymes correspondants via la terminologie, avec la fonction $HypG$, jusqu'à atteindre le premier concept en commun.
La Figure 4.2 illustre les résultats de cette étape pour la requête Q donnée en introduction de l'algorithme, où les concepts actifs identifiés et les scores associés sont représentés en gras.

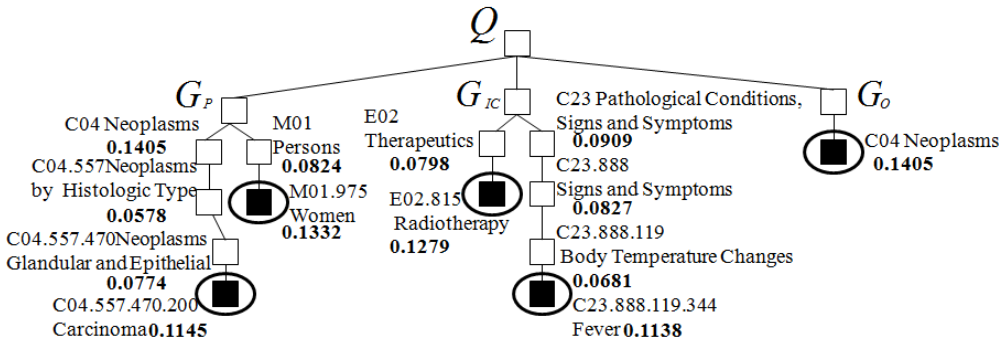


FIGURE 4.2 – Exemple de graphe sémantique de requête PICO

La Figure 4.3 donne un exemple d'appariement sémantique requête-document pour la requête suivante : "*In people with recurrent aggression having any antiepileptic drug in any dosage, what is length of time of placebo for observer reported aggression ?*" avec les éléments PICO suivants : P) "*people with recurrent aggression*" ; I)"*any antiepileptic drug in any dosage*" ; C) "*length of time of placebo*" ; O) "*reported aggression*".

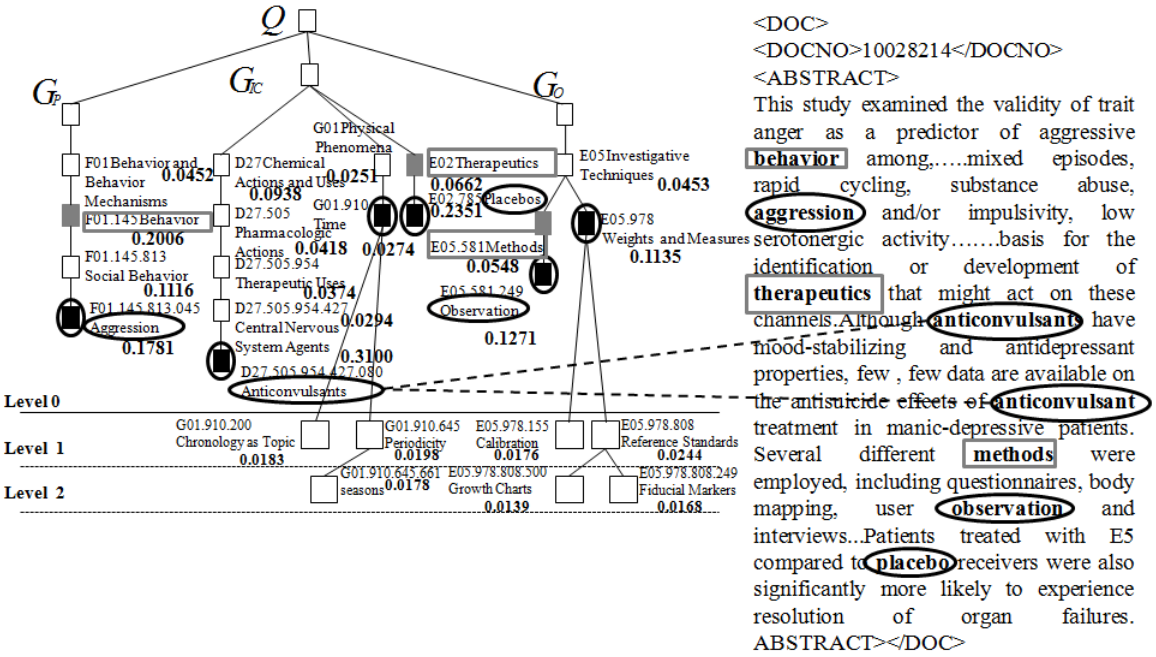


FIGURE 4.3 – Exemple d'un appariement sémantique requête-document

Les concepts actifs, identifiés en commun entre le document et la requête sont mentionnés par un carré noir. Nous avons proposé ensuite un algorithme de propagation de scores, qui propage les scores des concepts actifs de chaque sous-graphe du niveau 0 jusqu'au niveau Maxlevel d'une manière itérative. Nous détaillons cet algorithme dans ce qui suit.

4.2 Identification de concepts les plus pertinents

Nous décrivons dans l'Algorithme 2 le processus qui détaille l'identification des meilleurs concepts de la requête, à savoir : (i) extraction de concepts communs entre les meilleurs documents retournés et les sous-graphes PICO, et (ii) propagation des scores de poids de concepts pour identifier les meilleurs concepts $TopConcepts$, utilisés ensuite pour l'expansion.

Nos intuitions derrière cette approche sont les suivantes : (1) plus les concepts de la requête sont spécifiques au contexte de la recherche, plus ils sont significatifs ; et (2) plus le niveau de l'appariement est élevé entre les documents et chacune des facettes de la requête, plus ils sont pertinents. En se basant sur cette intuition générale, nous décrivons dans l'Algorithme 2, la fonction pour le classement des meilleures preuves (concepts médicaux) à retourner comme réponse pertinente pour les cliniciens.

Pour construire l'ensemble des concepts candidats $TopConcepts$ qui représentent au mieux la requête Q, nous considérons chaque document d dans D_N^* et ensuite, (1) nous procédons à l'extraction de concepts de chaque sous-graphe G_x (ou $x \in \{P, IC, O\}$) qui correspond à un document d ; (2) nous appliquons un algorithme de propagation de scores normalisés qui propage les scores des concepts actifs jusqu'aux sous-concepts ($Score(csub)$), en considérant chacun des sous-graphes de requêtes G_x à partir du niveau 0 jusqu'au niveau $Maxlevel$ et en ajoutant d'une manière itérative la somme

Algorithme 2 Propagation de scores et identification des meilleurs concepts

Entrées: G_x
Sortie: $TopConcepts$

1: {Traitement sur les meilleurs documents sélectionnés}
2: **Pour tout** $d \in D_N^*$ **Faire**
3: {Extraction des concepts des documents}
4: $TopConcepts \leftarrow Extract(d, G_x)$;
5: $level \leftarrow 0$;
6: {Propagation de scores}
7: **Pour tout** $c \in TopConcepts$ et $level < Maxlevel$ **Faire**
8: **Pour tout** $csub \in Hypo(c, G_x)$ **Faire**
9: $Score(csub) \leftarrow (Score(csub) + Lev(csub) * Score(c))$;
10: $Score(csub) \leftarrow Normalisé(Score(csub))$;
11: $level \leftarrow level + 1$;
12: **Fin Pour**
13: **Fin Pour**
14: **Fin Pour**
15: $TopConcepts \leftarrow Top_C(G_x, N_c)$; $TopConcepts$;

des scores de concepts hyponymes $Hypo(c, G_x)$ tout en respectant les sous-graphes G_x.
Selon l'hypothèse H1, l'idée de base est d'identifier les concepts les plus spécifiques qui représentent la facette x de la requête Q, ce qui les rend moins difficiles. La fonction $CSelect$ retourne les N_c concepts les mieux pondérés du sous-graphe G_x. La Figure 4.4 illustre cette étape sur la requête Q. Les concepts en gris sont ceux qui correspondent au document donné en exemple, en plus des concepts en gras entourés qui appartiennent à la requête. Les scores de concepts résultant de la propagation de scores sont représentés en gras.

La liste des concepts les mieux pondérés N_c retournés par $TopConcepts$ qui sont extraits de chaque sous-graphe G_x sont utilisés pour générer les sous-requêtes basées sur les concepts Q_P^c, Q_{IC}^c et Q_O^c liés respectivement aux sous-graphes G_P, G_{IC} et G_O.

Nous donnons dans ce qui suit un exemple de déroulement des différentes étapes du processus de génération de graphes ainsi que l'identification des meilleurs concepts.

Exemple illustratif
Nous allons présenter dans ce qui suit un déroulement des étapes de l'algorithme 1 et 2. Nous prenons comme exemple la requêtes textuelle présentée dans le Tableau 4.2 avec les annotations PICO correspondantes.

— Étape 1 : Les cinq documents les plus pertinents pour la requête A3.1 retournés par un modèle BM25 sont illustrés dans le Tableau 4.3.

— Étape 2 : À partir de la requête annotée, l'algorithme construit les sous-requêtes Q_P, Q_{IC} et Q_O comme le montre le Tableau 4.4 pour la requête A3.1.

  Le Tableau 4.5 montre les concepts actifs de la requête extraits en utilisant la terminologie MeSH et la Figure 4.4 donne les graphes sémantiques correspondant à P, IC et O.

  Le Tableau 4.6 donne le résultat de l'appariement sémantique entre les deux meilleurs do-

Id	**desc**	Q_{PICO}
$A3.1$	*Adult patients Primary Hypertension Beta-blockers and drug from another class of anti-hypertensive the other drug from another class of anti-hypertensive alone effect on blood pressure and heart rate*	***<P>****Adult patients Primary Hypertension Beta-blockers and drug from another class of anti-hypertensive****<P> <IC>****the other drug from another class of anti-hypertensive****</IC> <O>****alone effect on blood pressure and heart rate****</O>****.*

Tableau 4.2 – Exemple de requête textuelle avec les annotations PICO correspondantes.

UID	**Desc**	**MAP**
25761100	*Beta-blockers use for hypertension in the elderly.*	0.6314
26306578	*Blood pressure lowering efficacy of dual alpha and beta blockers for primary hypertension.*	0.6266
27048839	*Preoperative treatment with -blockers is associated with elevated postoperative mortality and cardiac morbidity in patients with uncomplicated hypertension undergoing non-cardiac surgery.*	0.6289
24911300	*Beta-blockers and depression in elderly hypertension patients in primary care.*	0.5704
26298028	*Thirty-day mortality leads to underestimation of postoperative death after liver resection : A novel method to define the acute postoperative period.*	0.5444

Tableau 4.3 – Top 5 documents retournés pour la requête A3.1 par le modèle BM25.

Q_P	*Adult patients Primary Hypertension Beta-blockers and drug from another class of anti-hypertensive.*
Q_{IC}	*the other drug from another class of anti-hypertensive.*
Q_O	*alone effect on blood pressure and heart rate.*

Tableau 4.4 – Exemple de sous requêtes PICO de la requête A3.1.

Requête	Concepts actifs (CUI)
A3.1	Adult(M01.060.116), Patients(M01.643), Hypertension(C14.907.489), Pharmaceutical Preparations (D26), Blood Pressure(E01.370.600.875.249), Heart Rate(E01.370.600.875.500)

Tableau 4.5 – Exemple de concepts médicaux actifs

cuments (*top ranked*) retournés par la première recherche et les sous-graphes de la requête avec les scores des meilleurs concepts sélectionnés pour la reformulation de la requête. Nous donnons aussi les concepts extraits de ces documents.

— Étape 3 : La propagation de scores d'appariement concept-document permet d'identifier les meilleurs concepts en sélectionnant les meilleurs poids.

Nous détaillons dans ce qui suit les deux contributions dans le cadre de l'évaluation des requêtes cliniques PICO.

G_P

<P>
<M01 concept="false">
Persons
<M01.060 concept="false"> Age Groups
<M01.060.116concept="true">
Adult
</M01.060.116>
</M01.060>
<M01.643concept="true">
Patients
</M01.643>
</M01>
<C14 concept="false">
Cardiovascular Diseases
<C14.907 concept="false">
Vascular Diseases
<C14.907.489 concept="true">
Hypertension
</C14.907.489>
</C14.907>
</C14>
</P>

G_{IC}

<IC>
<D26 concept="true">
Pharmaceutical Preparations
</D26>
</IC>

G_O

<O>
<E01 concept="false">
Diagnosis
<E01.370 concept="false">
Diagnostic Techniques and Procedures
<E01.370.600 concept="false">
Physical Examination
<E01.370.600.875 concept="false">
Vital Signs <E01.370.600.875.249
concept="true">
Blood Pressure
<E01.370.600.875.249>
<E01.370.600.875.500 concept="true">
Heart Rate
</E01.370.600.875.500>
</E01.370.600.875>
</E01.370.600>
</E01.370>

FIGURE 4.4 – Exemple de graphes sémantiques conceptuels

UID	concepts (CUI)[Scores]
25761100	*Adult*(M01.060.116)$[0,0914]$, *Patients*(M01.643)$[0.0972]$, *Hypertension*(C14.907.489)$[0,1058]$, *Nebivolol* (D03.438.150.755)$[0,0286]$, *Celiprolol* (D02.948.681.241)$[0,0356]$.
26306578	*Adult*(M01.060.116)$[0,0714]$, *Patients*(M01.643)$[0,0863]$, *Hypertension*(C14.907.489)$[0,0767]$, *Blood Pressure* (G09.330.380.076)$[0,0664]$, *Adrenergic beta-1 Receptor Antagonists* (D27.505.519.625.050.200.200.100)$[0,0418]$, *Placebos* (D26.660)$[0,557]$

Tableau 4.6 – Exemple de concepts de documents avec les scores associés.

5 Évaluation des requêtes cliniques

Une des plus grandes difficultés à développer chaque aspect de la question PICO est de fournir une quantité suffisante d'information où chaque composante de la question PICO doit être indiquée dans une phrase concise (Miller et Forrest, 2001). Du point de vue de la manière d'intégrer ces éléments avec les meilleures preuves cliniques pour mieux répondre à une situation spécifique est d'autant plus difficile vu le grand volume de la littérature médicale. Nous avons proposé d'évaluer les requêtes cliniques PICO en proposant deux approches : (1) une méthode d'expansion sémantique basée sur une technique spécifique d'extraction de concepts médicaux ; et (2) un modèle d'ordonnancement basé un opérateur d'agrégation prioritaire des scores. Nos contributions, dont le but est de mieux répondre aux questions cliniques PICO, sont décrites dans les sections qui suivent.

5.1 Expansion sémantique des requêtes cliniques

Nous nous sommes basés sur les hypothèses (H1 et H2 de la Section 2), ainsi que les conclusions établies par les travaux précédents, qui ont motivé la proposition de notre méthode d'expansion des

requêtes PICO utilisant les termes les plus spécifiques issus des ressources sémantiques médicales.

Pour pallier le problème de la dérive/ambiguïté du sujet de la requête, notre approche d'expansion repose sur l'utilisation des meilleurs concepts, et elle est guidée par le contexte sémantique local de la question PICO issus d'une phase de recherche conduite au préalable. En outre, pour mieux couvrir les différentes facettes PICO, nous avons utilisé chacune des facettes P, IC et O comme une source pertinente pour la sélection des concepts utilisés pour étendre la requête.

La Figure 4.5 donne un aperçu global de l'approche d'expansion sémantique des requêtes PICO.

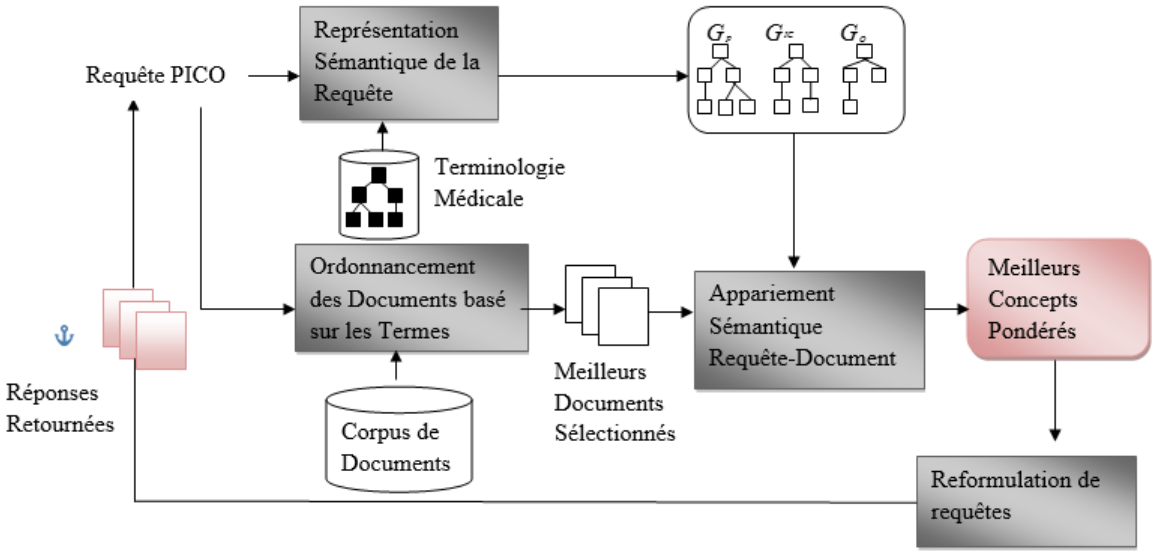


FIGURE 4.5 – Aperçu de notre modèle de traitement des questions PICO.

Pour chaque sous-graphe G_P, G_{IC} et G_O, on construit l'ensemble de concepts N_c utilisés plus tard pour l'expansion de la requête, utilisant la fonction $Expand(G_x)$, tout en tenant compte de *Max-level* qui est le niveau maximal considéré pour l'expansion de requête, commençant à partir du niveau 0. Afin de construire l'ensemble de concepts candidats $TopConcepts$, on considère chaque document d dans D_N^*, ensuite : extraction de l'ensemble des concepts pondérés communs avec G_x ou $x \in \{P, IC, O\}$ du niveau 0 jusqu'au niveau *Maxlevel* en additionnant les scores d'une manière itérative. Ces scores sont les hyponymes de concepts $Hypo(c, G_x)$ identifiés pour chacun des sous-graphes G_x (*Cf.* Algorithme 1). L'idée est d'identifier l'importance et la spécificité des concepts en assignant des scores normalisés $Nomalisé(Score(c))$ obtenu étape par étape du concept le moins spécifique jusqu'au concept le plus spécifique, considérant leur niveau *level(c)*.
Le score final du concept reflète son importance dans l'ensemble des meilleurs documents sélectionnés (*top-ranked*) en terme de spécificité et de degré d'appariement avec les documents D_N^*. Cela correspond à notre intuition qui favorise la sélection des concepts les plus spécifiques qui correspondent au mieux avec le contexte de recherche issu des documents les mieux classés retournés.
L'ensemble N_c de concepts les mieux pondérés C_{expand} de chaque sous graphe G_x est utilisé pour étendre respectivement les sous-requêtes Q_P Q_{IC} et Q_O (ayant abouti a Q_P^e Q_{IC}^e et Q_O^e) en ajoutant aux requêtes initiales Q les termes correspondant à leurs entrées préférées (Entrées(Q_P^e), Entrées(Q_{IC}^e) et Entrées(Q_O^e)) dans la terminologie T. La requête finale étendue Q^e est générée et permet la sélection de la liste finale des documents D_N^* qui sont retournés comme réponse à la requête initiale PICO Q.

5.2 Modèle d'ordonnancement basé sur un opérateur d'agrégation prioritaire des scores

5.2.1 Cadre général

De nombreux travaux en RI ont mis en exergue à la fois l'importance et la complexité du concept de pertinence (Da Costa Pereira *et al.*, 2009). Ces derniers se sont focalisés sur la modélisation d'approches théoriques pour la combinaison des critères de pertinence potentiels identifiés. D'autres travaux fondamentaux récents se sont intéressés en revanche à la définition d'opérateurs d'agrégation, indépendamment du cadre applicatif, qui permettent de traiter en partie le biais de l'interactivité (Pereira *et al.*, 2012).

Nous nous sommes basés sur un opérateur d'agrégation prioritaire, sur l'ensemble des facettes PICO, pour proposer un modèle d'ordonnancement de documents, qui donne un score de pertinence facette-document (Znaidi *et al.*, 2016b; Znaidi *et al.*, 2016a; Znaidi *et al.*, 2015). Notre modèle se base sur une approche d'agrégation prioritaire des scores tenant compte de l'importance des différentes facettes de la requête.

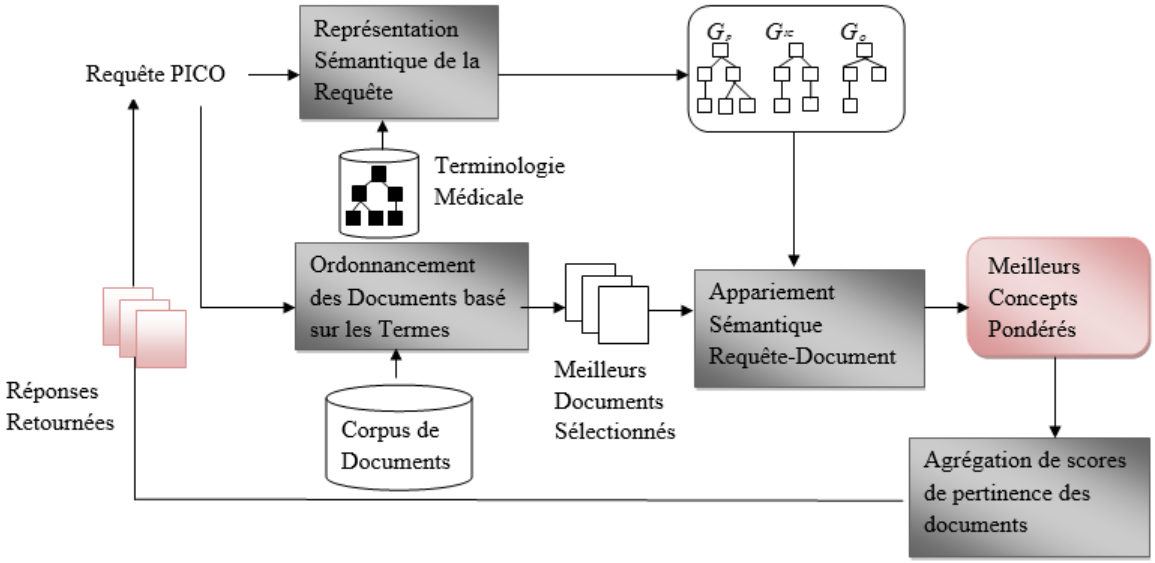


FIGURE 4.6 – Architecture de notre modèle de traitement des questions PICO.

Comme le montre la Figure 4.6, notre approche se base sur 4 étapes principales : (1) un appariement document-requête basé sur les termes, effectué en amont, qui retourne une liste initiale de documents pertinents candidats ; (2) une représentation sémantique des requêtes, qui retourne les graphes conceptuels associés à chaque facette PICO de la requête ; (3) un appariement requête-document qui permet de sélectionner les concepts les mieux pondérés de chaque facette du graphe ; et (4) un calcul de pertinence des documents basé sur un opérateur d'agrégation prioritaire (Da Costa Pereira *et al.*, 2009). Les étapes (1), (2) et (3) sont détaillées dans la Section 4. Nous détaillons dans ce qui suit l'étape 4 de calcul de pertinence en se basant sur une approche d'agrégation prioritaire des scores PICO.

5.2.2 Calcul de scores de pertinence des documents

Au niveau de l'étape (4), nous considérons une fonction d'agrégation F qui calcule, pour chaque document d, un score global PICO de pertinence, exprimé comme suit :

$$RSV_{PICO}(Q,d) = F(RSV_P(Q,d), RSV_{IC}(Q,d), RSV_O(Q,d)) \tag{4.1}$$

où F représente la fonction d'agrégation et P, IC et O sont les critères de pertinence PICO.

Nous calculons pour chacun des critères P, IC et O un score d'importance qui reflète son poids et qui prend en compte son importance dans le document et la nature de la tâche de RI. Nous nous sommes basés sur un opérateur prioritaire d'agrégation de scores (Da Costa Pereira *et al.*, 2009), qui permet de calculer un score global en prenant en compte un ensemble de critères. Le poids de chaque critère est calculé en fonction des poids ainsi que le degré de satisfaction du critère le plus important. Le principe est le suivant : plus le degré de satisfaction du critère le plus important est haut, plus le degré de satisfaction du critère le moins important influence le score global. Selon notre hypothèse (H2), on calcule le score PICO comme suit :

$$RSV_{PICO}(Q,d) = \lambda_P * RSV(Q_P,d) + lambda_{IC} * RSV(Q_{IC},d) + lambda_O * RSV(Q_O,d) \tag{4.2}$$

où $x \in \{P, IC, O\}$ et :

$$RSV_x(Q,d) = \frac{\sum_{c \in Q_x^c} SIM(c,d)}{\sum_{c \in \cup x Q_x^c} SIM(c,d)} \tag{4.3}$$

sachant que :

- Q_x^c est l'ensemble N_c de concepts pondérés, associés à chaque facette de la requête Q_x, résultat de l'Algorithme 2.
- $SIM(c,d)$ est la mesure de similarité à base de cosinus entre les vecteurs TF/IDF du document d et les entrées préférées associées au concept c (Boudin *et al.*, 2010a; Ruch, 2006).
- les poids d'importance des éléments PICO, sont calculés selon le principe de l'opérateur d'agrégation des scores (Da Costa Pereira *et al.*, 2009), en tenant compte de l'hypothèse H3, comme suit :

$$\begin{aligned} \lambda_P, \lambda_{IC}, \lambda_O \in [0..1], \quad \text{tels que} \quad \lambda_{IC} > \lambda_P > \lambda_O \quad et \quad \lambda_{IC} = 1 \\ \lambda_P = \lambda_{IC} * RSV(Q_{IC}^c, d) \\ \lambda_O = \lambda_P * RSV(Q_P^c, d) \end{aligned} \tag{4.4}$$

Pour prendre en compte la pertinence du document en se basant sur les termes et sur les concepts comme recommandé dans (Stokes *et al.*, 2009), le score de pertinence du document d par rapport à la requête Q est la combinaison linéaire des scores de pertinence basés sur les concepts ($Score_{PICO}(Q,d)$) et le score de pertinence basé sur les termes ($Score_w(Q,d)$). Le score global de pertinence est ainsi calculé comme suit :

$$RSV(Q,d) = \alpha * RSV_{PICO}(Q,d) + (1-\alpha) * RSV_w(Q,d) \tag{4.5}$$

où $\alpha \in [0..1]$ est un paramètre de combinaison linéaire identifié dans la phase des expérimentations.

6 Évaluation expérimentale

6.1 Données expérimentales

Nous utilisons la collection de données CLIREC dont la description est donnée dans le Tableau 4.7 ; construite dans le but d'évaluer la RI clinique (Boudin *et al.*, 2010c). Pour atteindre cet objectif, les auteurs dans (Boudin *et al.*, 2010c) ont construit la collection de test d'une manière semi-automatique à partir d'un ensemble de revues systématiques de la ressource *Cochrane*. Plus spécifiquement, les auteurs ont demandé à un groupe d'experts (un professeur et quatre étudiants en master de médecine) de générer les requêtes qui correspondent aux questions cliniques fournies par un sous ensemble du répertoire *Cochrane*. Chaque question est annotée manuellement avec les parties P (Patient,Problème), I (Intervention), C (Comparaison) et O (Outcome). Les citations de documents correspondants sont extraites de la section référence attachée à chaque revue, qui liste toutes les études pertinentes en traitant la question clinique considérée.
Enfin, une liste de documents pertinents (articles de journaux) pour chaque question clinique est générée de PubMed. Nous avons utilisé la terminologie MeSH, dont le vocabulaire est le plus utilisé pour indexer les citations biomédicales (Stokes *et al.*, 2009). Chaque nœud de la terminologie représente un concept qui fait référence à une entrée préférée dans la terminologie.

Nombre de documents	1.212.040 résumés PubMed
Longueur moyenne de document	246 termes
Nombre de requêtes	423
Nombre moyen de termes de la requête	4.3 termes
Longueur moyenne de la requête (PICO)	18.7 termes
Nombre moyen de doc pertinents	19

Tableau 4.7 – Statistiques de la collection de test CLIREC.

6.2 Évaluation de l'approche d'expansion de requêtes

6.2.1 Métriques d'évaluation et modèles de référence

Pour évaluer et comparer notre approche dans le cadre de la performance en RI, nous avons utilisé un ensemble de ressources sous la version 4.0 du moteur de recherche Terrier [3], à savoir :

1. *Mesures de performance* : la mesure de la *MAP (Mean Average Precision)* mesure de la précision moyenne sur l'ensemble de requêtes, et est utilisée pour fournir une mesure de performance globale unique. Plusieurs études antérieures ont montré la stabilité de la métrique MAP (Buckley et Voorhees, 2000). Les mesures de performance sont calculées en utilisant l'outil standard *TREC-eval* [4].

3. http ://www.terrier.org
4. http ://trec.nist.gov/trec-eval

2. Modèle de référence (*Baseline*) : nous avons comparé notre approche d'expansion de requêtes basée sur les graphes, appelée GQE, aux baselines suivantes :
   — Deux modèles de RI de l'état de l'art : *BM25 (Okapi Probabilistic model)* (Robertson et Sparck Jones, 1988) et le modèle de langue (LM) (Song et Croft, 1999). Nous avons paramétré le modèle Okapi comme recommandé dans la littérature $k1 = 1.2$, $k3 = 7$ et $b = 0.75$. Pour le modèle de langue, nous avons utilisé la méthode de lissage "*Dirichlet*" avec $\mu = 1000$.
   — Méthode de la ré-injection de la pertinence *PRF (Pseudo Relevance Feedback)* (Xu et Croft, 1996) que l'on considère proche de notre approche d'expansion proposée.

6.2.2 Ajustement des paramètres

Pour nos évaluations expérimentales, nous avons commencé par ajuster les paramètres des méthodes de comparaison PRF et BQE, et de notre algorithme d'expansion de requêtes basé sur les graphes, nommé GQE. Le modèle PRF dépend du nombre de documents et du nombre de termes utilisés pour l'expansion de la requête, variant respectivement dans un intervalle de valeurs [10..100] et [5..35] pour le nombre de documents et le nombre de termes, respectivement. Le modèle BQE dépend uniquement du niveau maximal de l'expansion de requête $MaxLevel$. Nous avons fait varier le paramètre "niveau" en choisissant différents *level* dans l'intervalle [0..4]. Nous avons fixé la limite de niveau à 4, compte tenue de la profondeur maximale de la propagation des scores qui peut être atteinte pour toutes les requêtes du test.
Nous visons à évaluer la précision des concepts en estimant leur impact sur l'efficacité de la recherche. Étant donnée l'absence ou le manque d'annotation conceptuelle humaine des requêtes PICO, nous avons évalué la précision des concepts résultats de l'Algorithme 1, en estimant leur impact sur la pertinence de la RI par l'approche d'expansion de requête.
En pratique, nous avons étendu la requête PICO initiale, avec les entrée principales des meilleurs concepts retournés par l'Algorithme 1. Ensuite, nous avons évalué la performance des requêtes étendues en utilisant le modèle probabiliste Okapi (BM25). Nous avons comparé la pertinence de cette méthode d'expansion conceptuelle, appelée GQE, aux deux modèles de référence suivants : 1) le modèle probabiliste Okapi (BM25), paramétré comme recommandé dans la littérature $k1 = 1.2$, $k3 = 7$ et $b = 0.75$, et 2) Pseudo-Relevance-Feedbak (PRF) (Xu et Croft, 2006). Pour effectuer des comparaisons équitables, nous commençons par paramétrer les algorithmes PRF et GQE puis nous avons comparé leur performance respective.

Comme le montre la Figure 4.7, les valeurs optimales du paramètre obtenues pour le modèle PRF sont 80 et 15, respectivement pour le nombre de documents et le nombre de termes d'expansion, permettant d'atteindre une valeur de MAP optimale de $0,1265$. Nous pouvons remarquer aussi à partir de la Figure 4.8 que le paramètre optimal $MaxLevel = 4$ pour le modèle BQE permet d'atteindre une valeur de la MAP de $0,1103$.

Notre algorithme GQE dépend de trois paramètres, à savoir : N_c le nombre de concepts utilisés pour l'expansion de requêtes, $MaxLevel$ utilisé pour limiter la profondeur lors du processus de propagation des scores et N_d qui est le nombre de premiers documents sélectionnés, utilisés pour l'extraction des concepts. Nous avons commencé par ajuster les paramètres N_c et $MaxLevel$, en les faisant varier respectivement dans un intervalle de [1..5] et [0..4], utilisant $N_d = 1000$. Comme le montre la Figure 4.10, la valeur optimale de la MAP est de $0,1334$, avec une configuration de 4 meilleurs concepts et une profondeur de propagation qui atteint le niveau 4. Nous avons retenu ce

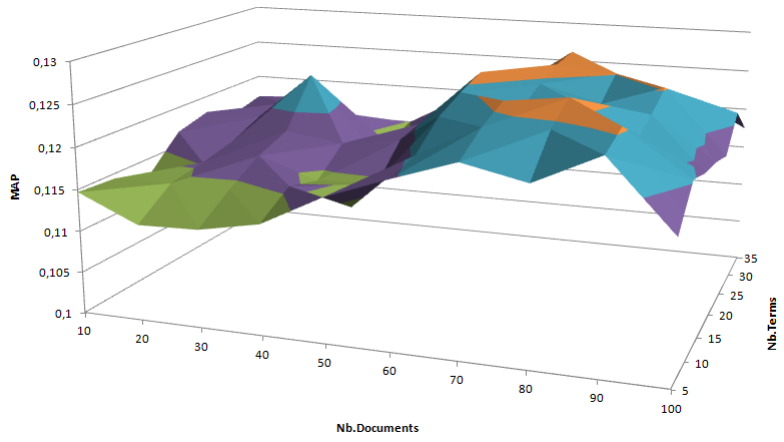


FIGURE 4.7 – Variation du paramètre pour le modèle PRF

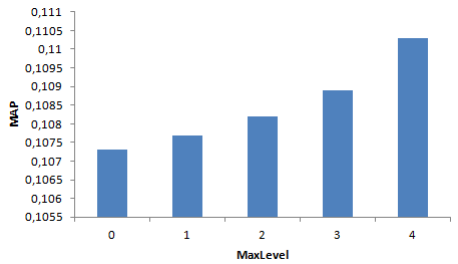


FIGURE 4.8 – Variation du paramètre pour le modèle BQE

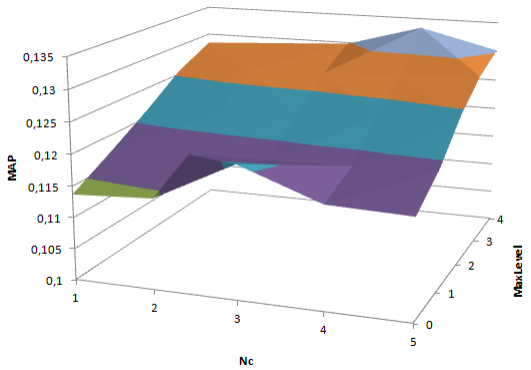


FIGURE 4.9 – Variation du paramètre pour l'expansion de requête

paramétrage optimal pour faire varier le paramètre N_d. Comme présenté par la Figure 4.9, la valeur optimale de la MAP est de $0,1346$ et est obtenue avec les 40 meilleurs documents retournés par le système.

Pour le reste de l'évaluation comparative, nous utilisons ces valeurs de paramètres optimales.

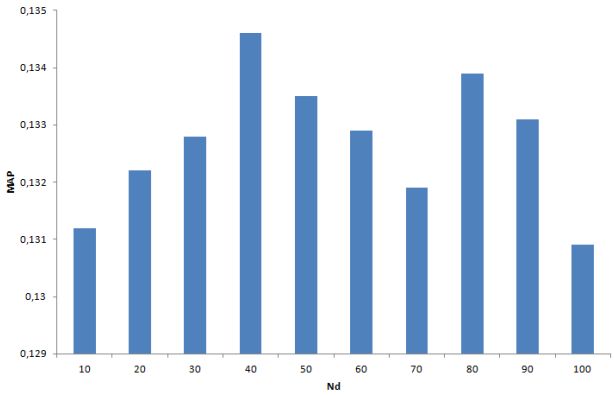


FIGURE 4.10 – Paramétrage du modèle GQE

6.2.3 Analyse de l'efficacité de la recherche : évaluation comparative

Nous avons comparé l'efficacité de la recherche en terme de *MAP* de notre algorithme d'expansion de requêtes basé sur les graphes (GQE), avec les modèles d'ordonnancement de l'état de l'art *BM25*, *LM*, *PRF* et *BQE*. Le Tableau 4.8 montre les résultats obtenus en terme de MAP, du nombre de documents pertinents sélectionnés ainsi que le taux d'accroissement et le taux de significativité t de la mesure statistique t-test. Notre modèle donne de meilleurs résultats par rapport à toutes les baselines. Plus précisément, en considérant la mesure de la MAP, nous avons observé que la performance de notre algorithme dépasse les approches d'ordonnancement de documents basées sur les termes ($BM25, LM$ et PRF) de $6,40\%$ à $27,94\%$, ainsi que l'approche d'évaluation basée sur les concepts de $22,03\%$.

Ces résultats montrent que :

1. notre approche sémantique permet d'obtenir de meilleurs résultats par rapport aux modèles de RI basés sur les termes de l'état de l'art, qui ne prennent pas en compte la structure PICO ($BM25, LM$ et PRF). Cela confirme notre intuition de la représentation des questions sur la base de la sémantique cachée derrière chacune des facettes PICO de la requête ;
2. notre algorithme de propagation de scores permet de sélectionner des concepts plus précis pour l'expansion de requêtes, par rapport à chaque facette PICO, puisque il permet de dépasser d'une manière significative le modèle BQE. Cependant, nous observons que les taux d'amélioration sont moins importants comparé au modèle PRF. Selon l'hypothèse *Hypothèse 3* annoncée au-dessus (Section 2) , cela suggère que les requêtes ne sont pas au même niveau de difficulté, et cela nécessite donc un algorithme approprié de propagation de scores.

6.2.4 Discussion

Nous avons présenté dans cette section une nouvelle approche pour répondre aux questions cliniques PICO. L'idée sous-jacente est d'améliorer chaque facette de la requête avec les concepts terminologiques les plus représentatifs du contexte local de la recherche. De plus, nous appliquons un algorithme de propagation de scores, qui permet de sélectionner les concepts avec le degré le plus élevé de correspondance avec le contexte de recherche et à travers les différentes facettes. Les expérimentations utilisant une collection de données standard ont montré que l'approche proposée dépasse significativement les modèles de RI proposés de l'état de l'art. Cette approche n'est

Model	MAP	% Acc	t	Rel. Ret	% Acc
$BM25$	$0,1073$	$+25,44\%$	††	4783	$+15,28\%$
LM	$0,1052$	$+27,94\%$	††	4685	$+17,69\%$
PRF	$0,1265$	$+6,40\%$	†	4983	$+10,65\%$
BQE	$0,1103$	$+22,03\%$	†	4819	$+14,42\%$
GQE	0.1346	-	-	5514	-

Tableau 4.8 – Comparaison de l'impact sur la performance de recherche de l'expansion de requêtes basée sur le graphe sémantique. $\%Acc$: Test de significativité de Student sur la métrique MAP † : $0,01 < t \leq 0,05$; †† : $0,001 < t \leq 0,01$; † † † : $t \leq 0,001$.

cependant pas sans limite. Nous supposons que les différentes facettes des questions PICO ont la même importance lors de l'évaluation de la pertinence des documents, ce qui n'est pas le cas dans la pratique EBM dans la vie réelle.

En conséquence, il est instructif de proposer une approche pour améliorer le modèle d'appariement sémantique requête-document avec le calcul de différents poids des facettes, afin d'entreprendre une comparaison équitable avec les travaux antérieurs qui ont pris en compte cet aspect (Boudin *et al.*, 2010c; Demner-Fushman et Lin, 2007).

Dans ce contexte, et dans le cadre de notre deuxième contribution, la section suivante décrit l'évaluation expérimentale de notre modèle d'ordonnancement pour mieux répondre aux questions cliniques PICO.

6.3 Évaluation du modèle d'ordonnancement de documents

6.3.1 Objectifs

Les objectifs principaux de l'évaluation expérimentale sont : (1) évaluer la précision de la représentation sémantique des requêtes PICO sous forme de graphes ; (2) évaluer l'efficacité du modèle d'ordonnancement basé sur une approche d'agrégation prioritaire des scores PICO ; (3) mesurer l'efficacité du modèle d'évaluation des requêtes PICO en analysant l'effet de chacun des éléments contributifs (représentation conceptuelle, principe de combinaison des facettes, principe de pondération des facettes) ; et (4) analyser la robustesse du modèle en identifiant les raisons possibles d'échec vs. succès des requêtes.

Nous détaillons dans ce qui suit le cadre expérimental ; nous présentons et nous discutons par la suite les résultats obtenus. Nous précisons que nous utilisons la collection CLIREC décrite dans la Section 6.1 pour les expérimentations.

6.3.2 Métriques et protocole d'évaluation

Dans le but d'évaluer la précision de la représentation sémantique des requêtes PICO et de comparer l'efficacité de recherche de notre opérateur de calcul de score, nous avons utilisé les mesures de la précision moyenne (MAP) et de précision exacte $P@X$ ($X = 5, 10$), calculées utilisant l'outil

standard TREC-eval [5]. Nous avons mis en œuvre un protocole de validation croisée *10-fold* (pour la phase de paramétrage et de test) afin de mesurer l'efficacité et la performance sur les 10 parties après une recherche de la valeur optimale qui maximise la MAP.

6.3.3 Efficacité du modèle de pertinence basé sur l'agrégation des scores

Dans un premier temps, nous avons identifié la valeur optimale du paramètre α utilisé dans l'équation (6) et ce en le faisant varier dans l'intervalle [0..1]. Comme le montre la Figure 4.11, la valeur optimale du paramètre α est de 0.7 permettant d'atteindre un score de précision MAP égal à 0.170. Ces résultats montrent que le score basé sur les éléments PICO confirme que ces derniers contribuent de façon significative au calcul du score global optimal. Nous retenons cette valeur pour le reste des expérimentations.

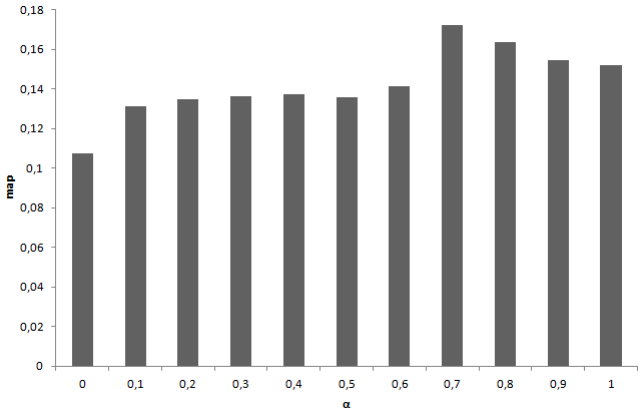


FIGURE 4.11 – Variation du paramètre α.

Nous avons comparé ensuite notre méthode d'ordonnancement de documents prioritaire, notée PSM, aux modèles de référence suivants :

— Deux modèles de RI de l'état de l'art : (1) Okapi probabilistic model ($BM25$) (Robertson et Sparck Jones, 1988) et (2) le modèle de langue (LM) (Song et Croft, 1999). Pour le modèle de langue, nous avons utilisé la méthode de lissage de Dirichlet avec $\mu = 1000$.

— Un calcul de scores basé sur un opérateur de priorité appliqué à un ensemble de documents ordonnés par un modèle BM25 noté ($PSBM25$), résultat de l'évaluation de chacune des sous-requêtes Q_P, Q_{IC} et Q_O.

— Deux modèles de RI conçus spécifiquement pour répondre aux requêtes PICO : (1) un modèle d'agrégation sans pondération noté (AGM), comme les travaux de (Demner-Fushman et Lin, 2007), où le modèle BM25 est utilisé pour calculer la similarité entre les facettes des requêtes et les documents, et (2) Positional Language Model (PLM) décrit dans (Boudin *et al.*, 2010c), basé sur une extension du modèle de langue (LM). Les résultats présentés sont ceux obtenus dans (Boudin *et al.*, 2010c) sur la même collection de test qui incluent uniquement les mesures de MAP et $P@5$.

Le Tableau 4.9 compare les résultats de l'efficacité de recherche de notre modèle d'agrégation prioritaire sémantique, notée PSM, avec ceux des cinq modèles de référence cités ci-dessus, selon les

5. http//trec.nist.gov/trec_eval

mesures de précision et du nombre de documents pertinents sélectionnés ($\#RR$). Les résultats montrent d'importantes améliorations significatives pour toutes les mesures d'évaluation, sauf pour le modèle PSM où l'amélioration est basse et non significative. Plus spécifiquement, pour la mesure de la MAP, les améliorations varient de $+4,60\%$ à $+52,36\%$. De plus, le nombre de documents pertinents retournés ($\#RR$) est plus élevé pour notre modèle.

Modèle	**Précision**			**% Acc**	**# RR**
	MAP	**P@5**	**P@10**		
$BM25$	$0,112$	$0,1561$	$0,127$	$+51,42\%^{\dagger\dagger}$	4574
LM	$0,111$	$0,156$	$0,130$	$+52,36\%^{\dagger\dagger\dagger}$	4491
$PSBM25$	$0,123$	$0,151$	$0,139$	$+37,94\%^{\dagger\dagger}$	4904
AGM	$0,121$	$0,148$	$0,135$	$+40,09\%^{\dagger\dagger}$	4835
PLM	$0,163$	$0,240$	–	$+4,60\%$	5770
PSM	**0,170**	**0,254**	**0,198**	–	**5894**

Tableau 4.9 – Comparaison de l'impact de notre modèle sur l'efficacité de recherche. $\%Acc$ indique les taux d'amélioration du modèle PSM en terme de MAP. Les Symboles †, †† et † † † indiquent le test de significativité de Student : † : $0,01 < t \leq 0,05$; †† : $0,001 < t \leq 0,01$; † † † : $t \leq 0,001$.

D'une manière générale, ces résultats mettent en avant quatre observations majeures :

1. Exploiter la sémantique cachée derrière les facettes des requêtes à travers les concepts est efficace étant donné que le modèle PSM dépasse les modèles $BM25$ et LM mais aussi le modèle $PSBM25$ qui se base sur l'opérateur d'agrégation prioritaire.
2. L'importante amélioration de performance du modèle PSM par rapport au modèle GQE montre l'intérêt d'intégrer la structure de la requête en facettes P, I/C et O dans le calcul des scores de pertinence des documents.
3. Le modèle PSM donne de meilleurs résultats que le modèle AGM basé sur la représentation sémantique de la requête ainsi que l'agrégation additive des scores d'importance des facettes ; ces résultats montrent l'avantage d'assigner des scores partiels de pertinence, en adéquation avec l'importance des facettes PICO avec le document en cours d'évaluation.
4. Le modèle PSM est légèrement mais pas significativement plus performant que le modèle PLM ($+4,60\%$ en terme de MAP). Ce point sera particulièrement exploré lors de l'analyse de robustesse du modèle PSM, présentée dans ce qui suit.

Pour mieux comprendre cette dernière observation, nous présentons une analyse de *gain* et *d'échec* de notre modèle, noté PSM en comparaison avec le modèle PLM. Nous étudions ensuite les performances selon les cas de requêtes.

6.3.4 Analyse de la robustesse du modèle d'agrégation de pertinence

Ici, notre objectif est d'analyser la robustesse du modèle PSM. Rappelons qu'un modèle de recherche d'information robuste doit impacter positivement la plupart des requêtes (Wang *et al.*, 2012). Pour cela, nous menons tout d'abord une analyse globale d'amélioration/baisse des performances sur l'ensemble des requêtes en comparaison avec les modèles de recherche d'information PICO de l'état de l'art, en l'occurrence les modèles PLM et AGM. Ensuite, nous nous focalisons sur l'étude de cas

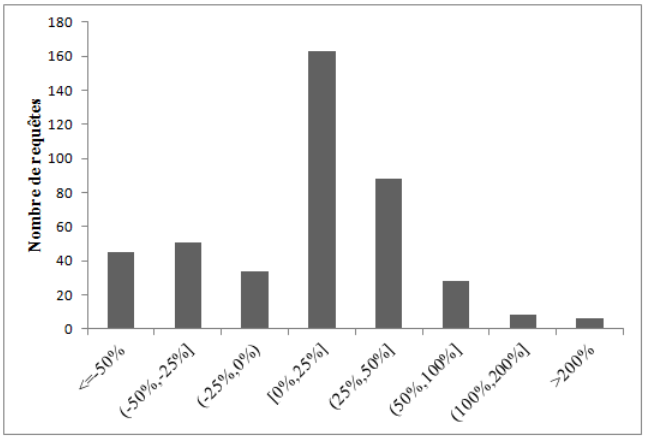


FIGURE 4.12 – Statistiques sur l'amélioration/dégradation en terme de MAP comparée au modèle PLM

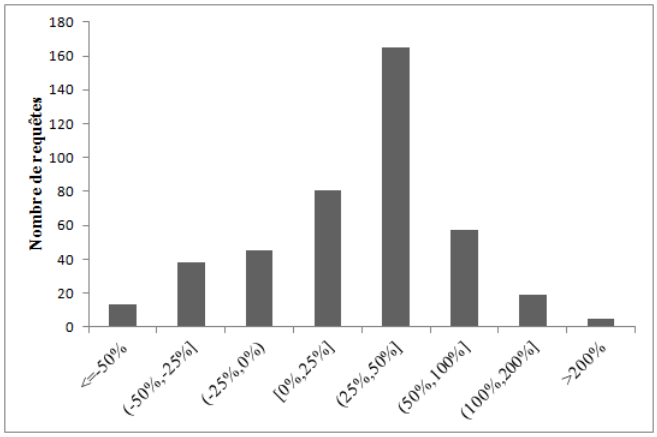


FIGURE 4.13 – Statistiques sur l'amélioration/dégradation en terme de MAP comparée au modèle AGM

de requêtes typiques du succès vs. échec du modèle PSM comparativement au modèle de référence le plus performant pour lequel les améliorations observées sont non significatives.
Les Tableaux 4.10 et 4.11 présentent le nombre (# requêtes), le pourcentage de requêtes (%) et le rapport moyen de changement positif (+) et négatif (−) : ($Avg.\ \%Acc$) pour trois ensembles de requêtes, à savoir : $R+$, R et $R-$, pour lesquels le modèle PSM donne un niveau de performance respectivement plus important, équivalent ou plus faible que le modèle PLM, en terme de MAP. L'équivalence de la performance est évaluée sur un intervalle de valeurs de $\%Acc$ dans l'intervalle $[-5\%... + 5\%]$.

6.3.4.1 Analyse globale

Les Figures 4.12 et 4.13 résument les statistiques concernant le nombre de requêtes pour lesquelles la performance du modèle PSM a été plus élevée vs. dégradée en comparaison avec les modèles PLM et AGM. L'axe des abscisses représente le taux l'amélioration (+) vs. dégradation (-) de la performance en terme de MAP. L'axe des ordonnées représente le nombre de requêtes concernées par cette amélioration vs. dégradation. Les barres à gauche de [0%..25%] représentent les requêtes

dont les performances sont plus élevées en considérant les modèles de comparaison AGM et PLM. Les barres à droite (incluant [0%..25%]) représentent les requêtes dont la performance est plus élevée pour le modèle PSM. On peut observer d'après ces figures que la proportion des requêtes avec une amélioration de la performance plus élevée en faveur du modèle PSM est plus importante.

À partir de la Figure 4.12, nous constatons plus particulièrement un grand nombre de requêtes avec une amélioration de performance dans l'intervalle [0%..25%] suivi de (25%..50%] par rapport au modèle PLM. Dans la Figure 4.13, nous observons un plus grand nombre de requêtes avec une amélioration de performance dans (25%..50%] suivi de l'intervalle [0%..25%] par rapport au modèle AGM. Nous remarquons également que l'amélioration est plus importante par rapport au modèle AGM avec un plus grand nombre de requêtes dans [25%..50%]. Cela montre l'utilité et l'efficacité d'appliquer l'opérateur d'agrégation de scores avec des poids d'importance variable d'une requête à l'autre. La performance de quelques requêtes dépasse les 100% d'amélioration.

Pour résumer, comparé aux modèles PLM et AGM, les résultats montrent que le modèle PSM est robuste.

Nous pouvons constater à partir du Tableau 4.11 que le pourcentage des requêtes pour lesquelles notre modèle PSM dépasse le modèle PLM est considérablement élevé ($64,3\%$ vs. $30,73\%$). Afin d'analyser et expliquer d'une manière précise les raisons de l'amélioration (vs. échec), nous avons analysé les deux meilleures requêtes de $R+$ et les deux plus mauvaises de $R-$, présentées dans le Tableau 4.13. Nous mentionnons la requête (Id), sa description (Des) ainsi que le ratio d'augmentation vs. diminution par rapport au modèle PLM ($\%Acc$), la longueur de la requête en terme de mots ($\#W$) et de nombre de concepts ($\#C$). Notons que la longueur de la requête peut expliquer partiellement les différences observées entre les deux ensembles de requêtes. Nous pouvons clairement remarquer que le modèle PSM présente une meilleure performance pour les requêtes relativement courtes ($M6.2$ *et* $Q37.2$) et inversement, le modèle PLM est performant pour les requêtes longues ($C21.1$ *et* $Q48.3$).

	#Requêtes (%)	$Avg.\%Acc$
R+	$272\ (64,3\%)$	$+26.97\%$
R	$21\ (4,96\%)$	-
R-	$130\ (30,73\%)$	-22.41%

Tableau 4.10 – Statistiques pour comparer la performance entre le modèle PSM et le modèle PLM.

	# Requêtes (%)	$Avg.\%Acc$
R+	$315\ (74,4\%)$	$+34.82\%$
R	$12\ (2,83\%)$	-
R-	$96\ (22,69\%)$	-15.43%

Tableau 4.11 – Statistiques pour comparer la performance entre le modèle PSM et le modèle AGM.

Cela peut être expliqué par le fait que, en comparaison avec le modèle PLM, la représentation conceptuelle des facettes de requêtes et des documents dans notre modèle PSM permet de réduire l'effet négatif de défaillance ou défaut de *mapping* entre les représentations basées sur les termes. À l'opposé, pour les requêtes longues, le problème de défaut de *mapping* est clairement réduit en faveur

du modèle PLM. Il est plus probable que le document corresponde à la requête selon les différentes facettes avec des interprétations trompeuses des intentions de recherche. À titre d'exemple, pour la requête $Q48.3$, les termes comme *older*, *rheumatoid* et *arthritis methotrexate* peuvent correspondre aux facettes implicites qui apparaissent dans les documents candidats. Cette ambiguïté n'existe pas pour le modèle PLM car les auteurs traitent des documents annotés avec les éléments PICO. Par contre, dans notre modèle, ne pas prendre en compte le contexte dans le calcul des poids d'importance des facettes PICO rend les résultats de recherche moins pertinents.

6.3.4.2 Analyse au niveau requête

Dans le but de comprendre les raisons de l'amélioration vs. dégradation de la performance des requêtes du modèle PSM comparativement au modèle de référence PSM, nous avons analysé les deux meilleures requêtes $R+$ et les deux plus faibles $R-$ en terme de performance MAP. Le Tableau 4.12 présente les ensembles $R+$ et $R-$ en mentionnant l'identifiant de la requête (Id), sa description (Des) ainsi que le taux de changement (augmentation vs. dégradation) par rapport au modèle PLM ($\%Acc$), la longueur en nombre de mots ($\#T$), le nombre de concepts extraits ($\#C$) et le score de clarté ($\#Cla$) de la requête. Ce score traduit le degré d'appariement de la requête Q avec la collection. Il est calculé sur la base d'une mesure de divergence entre le modèle de langue de la requête et celui de la collection (Steve et Croft, 2002), soit :

$$Cla(Q) = \sum_{t \in V} P(t|Q) log_2 \frac{P(t|Q)}{P_{coll}(t)} \tag{4.6}$$

où V est le vocabulaire de la collection, t est un mot, $P_{coll(t)}$ est la fréquence relative du mot t et $P(t|Q)$ est estimée $P(t|Q) = \sum_{d \in R} P(t|d)P(d|Q)$ où d est un document et R est l'ensemble des documents indexés par au moins un mot de la requête Q.

Les résultats présentés dans le Tableau 4.12 montrent que le modèle PSM présente une meilleure performance pour les requêtes relativement courtes ($M6.2$ *et* $Q37.2$) ; inversement, le modèle PLM est plus performant pour les requêtes longues ($C21.1$ *et* $Q48.3$). L'amélioration de la MAP atteint $78,08\%$ et $74,60\%$ respectivement pour les requêtes $M6.2$ et $Q37.2$. À l'opposé, pour les requêtes longues, la dégradation de la MAP atteint respectivement $-45,07\%$ et $-83,21\%$ pour les requêtes $C21.1$ et $Q48.3$. On remarque que les requêtes sont bien plus longues en mots (7, 8 vs. 18, 26) mais pas nécessairement plus longues en concepts avec le même ordre de grandeur (4, 5 vs. 7, 6). Même si on constate que ces requêtes sont moins claires ($0,062$ et $0,020$) que les requêtes plus longues ($0,080$ et $0,071$) en défaveur du modèle PLM, ce dernier s'avère plus performant. Ceci peut être expliqué par le fait que plus la requête est longue, plus le risque d'appariement des mots avec les documents candidats est grand. Cependant la non prise en compte du contexte de la facette, comme cela est fait dans le modèle PSM, conduit au calcul d'un score de pertinence selon la formule (3) qui pourrait être erroné. À titre d'exemple, pour la requête $Q48.3$, des mots comme *older*, *rheumatoid* et *arthritis methotrexate* présents dans un document candidat peuvent s'apparier indifféremment avec les facettes IC et P dans le cas du modèle PSM. En revanche, dans le modèle PLM, ce défaut d'appariement lié à la facette ne peut survenir puisque les documents sont préalablement annotés et l'appariement est effectué facette à facette.

Afin de comprendre davantage les raisons de l'amélioration/baisse de performance des requêtes de notre modèle par rapport aux autres modèles, nous avons analysé les deux meilleures requêtes de

Q	Id	Desc	%*Acc*	#*T*	#*C*	#*Cla*
R^+	$M6.2$	*(**P**) In obese patients diabetes(**P**)(**IC**) orlistat Placebo(**IC**) (**O**)Weight loss(**O**).*	$+78,08\%$	7	4	$0,080$
	$Q37.2$	*(**P**) Adults 18 years or more migraine(**P**)(**IC**)aspirin plus an antiemetic placebo (**IC**) **O**)pain free(**O**).*	$+74,60\%$	8	5	$0,071$
R^-	$C21.1$	*(**P**) Adults 14 years and older GORD (**P**)(**IC**)Medical management : proton pump inhibitors/histamine receptor antagonists Laparoscopic fundoplication surgery(**IC**)(**O**) Health-related quality of life (**O**).*	$-45,07\%$	18	7	$0,062$
	$Q48.3$	*(**P**) Adults 18 years or older rheumatoid arthritis (**P**)(**IC**) methotrexate combined with other non-biologic disease modifying anti-rheumatic drugs (DMARDs) methotrexate alone(**IC**)(**O**)ACR response of non-MTX DMARDS inadequate response(**O**).*	$-83,21\%$	26	6	$0,020$

Tableau 4.12 – Analyse comparative de cas de requêtes types PSM vs. PLM

$R+$ et les deux les plus faibles de $R-$, présentés dans les Tableaux 4.13 et 4.14 respectivement comparé aux modèles PLM et AGM. Nous mentionnons la requête (Id), sa description (Des) ainsi que le ratio d'augmentation vs. diminution par rapport au modèle PLM et AGM ($\%Acc$), la longueur de la requête en terme de mots ($\#W$), le nombre de concepts ($\#C$), la spécificité terme-document ($\#Tspe$) et hiérarchique ($\#Hspe$), la clarté basée sur le sujet($\#Scla$) et celle basée sur la pertinence ($\#Pcla$). On peut mentionner que la longueur de la requête peut expliquer partiellement les différences observées entre les deux ensembles de requêtes. Nous pouvons clairement remarquer que le modèle PSM présente une meilleure performance pour les requêtes relativement courtes ($M6.2$ *et* $Q37.2$) et inversement, le modèle PLM est performant pour les requêtes longues ($C21.1$ *et* $Q48.3$). Cela peut être expliqué par le fait que, en comparaison avec le modèle PLM, la représentation conceptuelle des facettes de requêtes et des documents dans notre modèle PSM permet de réduire l'effet négatif du défaut de *mapping* entre les représentations basées sur les termes. À l'opposé, pour les requêtes longues, le problème de défaut de *mapping* est clairement réduit en faveur du modèle PLM, il est plus probable que le document soit associé à la requête selon les différentes facettes avec des interprétations non précises des intentions de recherche. Par exemple pour la requête $Q48.3$, les termes comme *older*, *rheumatoid* et *arthritis methotrexate* peuvent correspondre aux facettes implicites qui apparaissent dans les documents candidats. Cette ambiguïté n'existe pas pour le modèle PLM car les auteurs traitent des documents annotés avec les éléments PICO. Cela peut induire à des poids d'importance de facettes PICO inappropriés (selon la formule X), et par conséquence diminue l'efficacité de recherche par rapport au modèle PSM.

En terme de spécificité : les scores de la spécificité terme-document sont légèrement différents pour les deux catégories de requêtes et sont plus importants pour les requêtes plus performantes pour

Q	Id	$\%Acc$	$\#T$	$\#C$	$\#Tspe$	$\#Hspe$	$\#Scla$	$\#Pcla$
R^+	$M6.2$	$+78,08\%$	7	4	$0,603$	$0,212$	$0,080$	$1,642$
	$Q37.2$	$+74,60\%$	8	5	$0,520$	$0,136$	$0,071$	$1,658$
R^-	$C21.1$	$-45,07\%$	18	7	$0,491$	$0,106$	$0,062$	$1,5851$
	$Q48.3$	$-83,21\%$	26	6	$0,519$	$0,097$	$0,020$	$1,367$

Tableau 4.13 – Analyses de l'échec et de l'amélioration de la performance des requêtes avec le modèle PLM.

Id	Desc
$M6.2$	*(**P**) In obese patients diabetes(**P**)(**IC**) orlistat Placebo(**IC**) (**O**)Weight loss(**O**).*
$Q37.2$	*(**P**) Adults 18 years or more migraine(**P**)(**IC**)aspirin plus an antiemetic placebo (**IC**) **O**)pain free(**O**).*
$C21.1$	*(**P**) Adults 14 years and older GORD (**P**)(**IC**)Medical management : proton pump inhibitors/histamine receptor antagonists Laparoscopic fundoplication surgery(**IC**)(**O**) Health-related quality of life (**O**).*
$Q48.3$	*(**P**) Adults 18 years or older rheumatoid arthritis (**P**)(**IC**) methotrexate combined with other non-biologic disease modifying anti-rheumatic drugs (DMARDs) methotrexate alone(**IC**)(**O**)ACR response of non-MTX DMARDS inadequate response(**O**).*

Tableau 4.14 – Description des requêtes de l'analyse (*Cf.* Tableau 4.13).

notre modèle. Cela montre que notre modèle traite bien les concepts qui représentent le sujet de la requête dans les documents pertinents. Le modèle PLM se base sur la distribution de termes par facette PICO dans les documents sans prendre en compte leur importance ce qui réduit leur pertinence. En terme de spécificité hiérarchique, les requêtes plus performantes pour notre modèle sont plus spécifiques par rapport au modèle PLM et AGM, avec des scores de spécificité plus importants. Malgré le fait que les requêtes les plus performantes pour notre modèle contiennent moins de concepts, ces derniers représentent mieux la sémantique de la requête appariée à celle du document. Les scores de clarté sont plus importants pour les requêtes plus performantes pour notre modèle. On remarque aussi qu'il y a une différence en comparant les scores de clarté entre les meilleures requêtes/plus faibles via les deux modèles PLM et AGM. Pour le modèle AGM, les scores de clarté sont clairement plus faibles pour les requêtes moins performantes comparés au modèle PLM. Ce modèle ne prend pas en compte l'importance des termes par facette, ce qui les rend plus difficile à traiter. Les Tableaux 4.15 et 4.16 donnent les résultats d'analyses de performance des requêtes avec le modèle PLM.

Q	Id	$\%Acc$	$\#T$	$\#C$	$\#Tspe$	$\#Hspe$	$\#Scla$	$\#Pcla$
R^+	$A24.2$	$+82,36\%$	11	4	$0,604$	$0,155$	$0,055$	$1,931$
	$C9.3$	$+77,49\%$	19	5	$0,523$	$0,145$	$0,050$	$2,144$
R^-	$C13.2$	$-18,65\%$	19	4	$0,495$	$0,106$	$0,047$	$1,164$
	$Q26.3$	$-51,23\%$	20	10	$0,465$	$0,136$	$0,004$	$1,4497$

Tableau 4.15 – Analyse de l'échec et de l'amélioration de la performance des requêtes avec le modèle AGM.

Nous avons effectué des analyses plus approfondies sur les requêtes ou il y a eu une baisse de

Id	Desc
$A24.2$	*(**P**) Adults Advanced renal cell carcinoma(**P**)(**IC**) Targeted agents Standard interferon-alfa(**IC**) (**O**)Overall survival(**O**).*
$C9.3$	*(**P**) Adults aged 18 or older Lower back pain (LBP)(**P**)(**IC**)Post-treatment exercise consisting of strength, endurance and aerobic training (**IC**) **O**)No intervention Number of recurrences of LBP(**O**).*
$C13.2$	*(**P**) Adults (18 years and older) MI, angina, or who had undergone revascularisation or heart failure (**P**)(**IC**)Home-based cardiac rehabilitation(**IC**)(**O**) Centre-based cardiac rehabilitation Systolic blood pressure (**O**).*
$Q26.3$	*(**P**) adults 18 years and over any type of wound in any care setting (**P**)(**IC**) wound dressings containing silver wound dressings(**IC**)(**O**)without silver number of patients that developed wound infection(**O**).*

Tableau 4.16 – Description des requêtes de l'analyse (*Cf.* Tableau 4.15).

performance par notre modèle :

— Par rapport au modèle PLM (**C21.1** et **Q48.3**) : la requête $C21.1$ contient un acronyme médical *GORD (Gastro Oesophageal Reflux Disease)*, notre modèle ne traite pas les acronymes d'une manière à bien identifier le problème gastrique dans la facette P (patient/problème). De plus, les termes que contient la facette IC, qui est la facette la plus importante de la question PICO, et qui sont *proton pump inhibitors/histamine receptor antagonists Laparoscopic fundoplication surgery*, sont représentés dans notre approche sémantique par l'ensemble de concepts *Receptors Histamine, General Surgery, Fundoplication, Proton Pump Inhibitors*. Vu que le modèle PLM se base sur la distribution des termes dans un modèle de langue, tenant compte de la position de chacun des termes dans les documents, ceci explique la meilleure performance de ce modèle par rapport au notre. De même, la requête $Q48.3$ contient plusieurs acronymes *DMARDs, MTX, DMARDS, ACR*, or notre approche sémantique basée sur l'extraction de concepts de MeSH ne prend pas en compte ces cas de figure ce qui rend le sujet de la requête peu claire.

— Par rapport au modèle AGM (**C13.2** et **Q26.3**) : nous remarquons plusieurs termes redondants dans ces requêtes, comme pour $C13.2$ les termes *Cardiac* et *rehabilitation* qui apparaissent dans les facettes P et IC de la requête. De même pour la requête $Q26.3$, le terme *wound* apparait 3 fois dans les trois facettes P, IC et O de la requête. La méthode AGM donne de meilleurs résultats vu qu'elle n'utilise pas l'algorithme de propagation de scores qui désigne l'importance de chacun des concepts de la requête.

7 Bilan et conclusion

Dans cette partie, nous nous sommes intéressés à la RI clinique basée sur les requêtes PICO dans le contexte de la médecine basée sur les preuves (EBM). La principale motivation de nos contributions repose sur la complémentarité des parties/informations sémantiques dans le cadre de l'appariement requête-document. Elles permettent d'apporter plus de couverture au niveau sujet et contexte entre requête et document (Demner-Fushman et Lin, 2007; Boudin *et al.*, 2010b), pour apparier la requête avec les documents pertinents et donc mieux répondre aux questions cliniques expertes PICO.

Dans un premier temps, nous avons proposé une approche d'expansion de requêtes cliniques PICO dans l'objectif d'améliorer les résultats retournés par le système. Nous avons également mis en évidence l'importance des caractéristiques sémantiques dans l'identification de meilleures preuves médicales, capables d'améliorer les résultats de recherche. Un premier volet a traité l'exploitation des parties sémantiques représentées par la reformulation conceptuelle des requêtes PICO afin d'améliorer la performance de la recherche. Le deuxième volet est complémentaire au premier dans la mesure ou les deux algorithmes se basent sur un une approche de génération de graphes sémantique. Cette partie propose un modèle d'appariement requête-document basé sur une méthode d'agrégation prioritaire des scores pour classer d'une manière optimale les documents selon leur pertinence à une requête.

Nous avons proposé l'application d'un opérateur d'agrégation prioritaire pour l'évaluation de requêtes cliniques PICO. L'opérateur ne requiert pas une annotation préalable des facettes PICO dans les documents et permet d'adapter le score d'importance des facettes aux documents et requêtes en cours d'évaluation. Les expérimentations conduites sur la collection *CLIREC* ont montré que l'opérateur proposé est significativement plus performant que la majorité des modèles de référence basés sur l'appariement mot-mot, la reformulation sémantique des requêtes et des modèles d'agrégation classiques des ordonnancements issus de l'évaluation de chaque sous-requête associée à une facette. Bien que robuste, l'analyse des performances de recherche au niveau requête montre que l'opérateur PSM présente des limites. En effet, le modèle d'ordonnancement des documents ne prend pas en compte le lien entre le contexte des mots dans les documents et leur contexte dans la requête, représenté par les facettes auxquelles les mots font référence. Nous prévoyons de vérifier le bien fondé de ces limites en menant une analyse statistique sur l'ensemble de la collection qui permettrait de déterminer les facteurs d'échec des requêtes. Il en suivra une réflexion concernant l'intégration de ces facteurs comme éléments contextuels dans l'opérateur d'agrégation prioritaire des scores.

Conclusion générale

Synthèse des contributions

Les travaux présentés dans ce manuscrit s'inscrivent dans le contexte de la recherche d'information biomédicale, qui correspond à un des domaines émergents de la recherche d'information avec de nombreux enjeux multidisciplinaires tels que la formalisation des modèles d'ordonnancement, analyse de l'expertise dans le domaine médical, la prise en compte des caractéristiques et conditions du besoin exprimé pour améliorer les systèmes de prise de décision dans le domaine médical.

Dans cette thèse, nous nous sommes particulièrement intéressés à l'analyse des requêtes expertes du domaine médical, de point de vue de leur formulation ainsi que les caractéristiques linguistiques. L'importance de caractériser les requêtes médicales des experts s'est manifestée à travers les travaux de l'état de l'art.

En effet, la plupart des études ont analysé les caractéristiques des requêtes dans un cadre spécifique qui ne peut pas être généralisé. Nous avons levé une partie d'ambiguïté sur le besoin en information des expert du domaine médical. Nous nous sommes focalisés sur la proposition de modèles de recherche d'information médicaux, spécifiques aux requêtes cliniques PICO dans le cadre de la médecine basée sur les preuves. D'une part, l'identification des éléments PICO à partir des questions cliniques et des documents ; et l'utilisation de ces éléments pour accomplir la tâche de recherche d'information en vue de mieux répondre aux questions cliniques PICO, d'autre part.

Dans ce contexte, nous avons proposé deux types de contributions, respectivement focalisées sur les requêtes expertes des cliniciens et le cadre de la recherche basée sur les preuves cliniques utilisant les requêtes PICO. Nous les rappelons brièvement ci-dessous :

1. *Analyse statistique comparative de requêtes expertes, issues de différentes tâches de RI médicaux* : nous avons mis en évidence le fait que : (a) Il y a des variabilités significatives à la fois dans la morphologie des requêtes que des besoins et des performances, (b) Trois attributs impactent les résultats de recherche, plus spécifiquement la longueur en termes, le score de clarté basé sur le sujet et la spécificité terme-document, (c) contrairement aux utilisateurs qui cherchent l'information en ligne, les requêtes des médecins sont relativement longues, et cette longueur dépend de la nature de la tâche de RI, et (d) les experts médicaux se basent sur leurs connaissances du domaine et utilisent des ressources sémantiques pour formuler les requêtes, ce qui les rends spécifiques surtout pour la recherche des cas médicaux de patients.

Ces résultats suggèrent le besoin de contextualiser les modèles de RI médicale à la tâche. Plus précisément, un besoin de clarification et spécification par expansion/reformulation de requête serait approprié pour des requêtes cliniques expertes.

De plus, les résultats de cette étude ont un impact sur la conception de systèmes de recherche d'information médicaux, plus spécifiquement celle de prise de décision médicale.

2. *Analyse des requêtes cliniques PICO, dans le cadre de la recherche d'information médicale basée sur les preuves cliniques (EBM).* Nous nous sommes intéressés à l'évaluation de requêtes cliniques exprimées avec les facettes PICO (Population/Problem (P), Intervention (I), Comparaison (C) et Outcome (O)). D'après les travaux de l'état de l'art, la recherche médicale basée sur le paradigme PICO, en vue d'identifier les preuves cliniques pertinentes, est un domaine sous exploré Ainsi, nous avons proposé de considérer la recherche basée sur les requêtes PICO selon deux axes, à savoir : (i) identification des facettes PICO, à partir des documents et des requêtes et (ii) l'utilisation des facettes dans la RI pour rechercher les meilleures preuves correspondant à chacune des facettes P, IC et O. Nous avons ainsi formulé un ensemble d'hypothèses, dans la mesure où un document est d'autant plus pertinent qu'il s'apparie avec plus de facettes de la requête (Boudin *et al.*, 2010c; Demner-Fushman et Lin, 2007) ; et en phase d'évaluation de la pertinence des résultats, l'expert médical n'accorde pas la même importance à l'adéquation des facettes. La facette I/C est plus importante que la facette P qui, à son tour, est plus importante que la facette O (Weifield et Finkelstein, 1996; Boudin *et al.*, 2010c).

   Nous avons donc proposé : (1) un algorithme de classement basé sur une expansion sémantique de requêtes qui prend en compte chacune des facettes PICO ; (2) une représentation sémantique sous forme de graphes conceptuels correspondant à chacun des éléments P IC et O ; (3) une approche d'appariement sémantique requête-document pour mieux répondre aux questions cliniques PICO. Ces contributions se basent sur l'application d'un opérateur d'agrégation prioritaire des scores permettant de : (a) agréger les scores de pertinence partiels issus de l'évaluation de représentations sémantiques associées aux sous-requêtes facettes et (b) contextualiser le score d'importance des facettes au document et requête en cours d'évaluation.

Perspectives

Les différentes analyses et évaluations expérimentales menées pour évaluer nos différentes contributions ont montré leur efficacité vis-à-vis des modèles de l'état de l'art. Ce manuscrit ouvre de nombreuses perspectives que nous synthétisons dans ce qui suit.
À moyen terme, nous proposons de généraliser nos contributions selon deux dimensions, liées aux scénarios de recherche médicale sur le web.

1. *Assistance à la recherche d'information médicale par suggestion de requêtes adaptées au profil des utilisateurs.* Nous avons proposé un modèle spécifique à la tâche de RI médicale experte. L'expertise du domaine est étroitement liée aux connaissances du domaine. Nous pensons qu'il pourrait être intéressant d'intégrer une modélisation des utilisateurs en prenant en compte leur niveau d'expertise. Nous proposerons alors : 1) des ressources permettant spécifiquement la réécriture de requêtes d'un lexique avec un niveau d'expertise vers un autre lexique d'un autre niveau d'expertise ; cela convient particulièrement à la traduction du niveau expert vers le niveau novice de manière à rendre l'information plus accessible à ces derniers, 2) des

techniques de suggestions de requêtes suite à une détection de sessions de recherche complexes caractérisées par une dérive du sujet.

2. *RI médicale sur le web.* Avec la quantité abondante des sources d'information médicales sur le web, l'utilisateur qui n'est pas forcément un professionnel de santé peut avoir des difficultés à analyser les réponses retournées par les moteurs de recherche et, plus spécifiquement, à s'assurer de la crédibilité de cette information lors de l'auto-diagnostic. En effet, plusieurs travaux ont montré que la recherche sur le web en lien avec la santé peut amener à l'anxiété chez l'utilisateur, et cela a des répercussions sur les décisions à prendre quant à la prise en charge de santé. Ce phénomène est connu sous le nom *Cyberchondria* (White et Horvitz, 2009b). Les auteurs ont montré que le web est une source potentielle d'inquiétude inopportune pour les utilisateurs. Pour faire face à ce phénomène, nous envisageons de proposer un cadre d'évaluation multicritère de la pertinence en RI médicale en exploitant différents facteurs de pertinence selon par exemple les profils des utilisateurs, la nature du besoin en information et la crédibilité des sources d'information.

Bibliographie

Adriani, M. et Rijsbergen, C. J. V. (1999). Term similarity-based query expansion for cross-language information retrieval. *In In Proceedings of the third European Conference on Research and Advanced Technology for Digital Libraries (ECDL '99*, pages 311–322.

Agirre, E., Arregi, X. et Otegi, A. (2010). Document expansion based on wordnet for robust ir. *In Proceedings of the 23rd International Conference on Computational Linguistics : Posters*, COLING '10, pages 9–17, Stroudsburg, PA, USA. Association for Computational Linguistics.

Akrivas, G., Wallace, M., Andreou, G., Stamou, G. et Kollias, S. (2002). Context-sensitive semantic query expansion. *In Artificial Intelligence Systems, 2002. (ICAIS 2002). 2002 IEEE International Conference on*, pages 109–114.

Al-Maskari, A. et Sanderson, M. (2011). The effect of user characteristics on search effectiveness in information retrieval. *Inf. Process. Manage.*, 47(5):719–729.

Allen, B. (1991). Topic knowledge and online catalog search formulation. *The Library Quarterly : Information, Community, Policy*, 61(2):pp. 188–213.

Alper, B. S., Stevermer, J. J., White, D. S. et Ewigman, B. G. (2001). Answering family physicians' clinical questions using electronic medical databases. *Journal of Family Practice*, 50(11):960–965.

Andreassen, H. K., Bujnowska-Fedak, M. M., Chronaki, C. E., Dumitru, R. C., Pudule, I., Santana, S., Voss, H. et Wynn, R. (2007). European citizens' use of e-health services : a study of seven countries. *BMC public health*, 7(1):1.

Andrews, J. E., Pearce, K. A., Ireson, C. et Love, M. M. (2005). Information-seeking behaviors of practitioners in a primary care practice-based research network (PBRN). *Journal of the Medical Library Association : JMLA*, 93(2):206–212.

Andualem, M., Kebede, G. et Kumie, A. (2013). Information needs and seeking behaviour among health professionals working at public hospital and health centres in bahir dar, ethiopia. *BMC Health Services Research*, 13(1):1–9.

Arnold, P. et Rahm, E. (2015). Semrep : A repository for semantic mapping. *In Datenbanksysteme für Business, Technologie und Web (BTW), 16. Fachtagung des GI-Fachbereichs "Datenbanken und Informationssysteme" (DBIS), 4.-6.3.2015 in Hamburg, Germany. Proceedings*, pages 177–194.

Aronson, A. R. (2001). Effective mapping of biomedical text to the umls metathesaurus : the metamap program. *Proc AMIA Symp*, pages 17–21.

Arora, N., Hesse, B., B.K, R., Viswanath, K., Clayman, M. et Croyle, R. (2007). Frustrated and confused : the american and public rates its cancer-related information-seeking experiences. *Journal of General Internal Medicine*, 23(3):223–228.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T. *et al.* (2000). Gene ontology : tool for the unification of biology. *Nature genetics*, 25(1):25–29.

Attia, A. (2013). Adaptation of international evidence based clinical practice guidelines : The adapte process. *Middle East Fertility Society Journal*, 18(2):123–126.

Baeza-Yates, R. et Ribeiro-Neto, B. (2011). *Modern Information Retrieval : The Concepts and Technology behind Search (2nd Edition) (ACM Press Books)*. Addison-Wesley Professional, 2 édition.

Baeza-Yates, R., Ribeiro-Neto, B. *et al.* (1999). *Modern information retrieval*, volume 463. ACM press New York.

Baker, N. C. et Hemminger, B. M. (2010). Mining connections between chemicals, proteins, and diseases extracted from medline annotations. *Journal of Biomedical Informatics*, 43(4):510 – 519.

Baziz, M., Aussenac-Gilles, N. et Boughanem, M. (2003). Désambiguisation et Expansion de Requêtes dans un SRI,Etude de l'apport des liens sémantiques. *Revue des Sciences et Technologies de l'Information (RSTI) série ISI*, 8(4/2003):113–136.

Bhavnani, S. K. (2001). Important Cognitive Components of Domain-Specific Search Knowledge. pages 571–578.

Bhavnani, S. K. (2002). Domain-specific search strategies for the effective retrieval of healthcare and shopping information. *In CHI '02 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '02, pages 610–611, New York, NY, USA. ACM.

Bigot, A. (2013). Adapter les moteurs de recherche aux besoins en information - Prise en compte de la difficulté du besoin (regular paper). pages 59–74, http ://www.univ-paris1.fr/. Université Paris 1.

Blair, D. C. et Maron, M. E. (1985). An evaluation of retrieval effectiveness for a full-text document-retrieval system. *Commun. ACM*, 28(3):289–299.

Boden, C. (2009). Overcoming the linguistic divide : a barrier to consumer health information. *Journal Canadian Health Libraries Association*, 30(3):75–80.

Borlund, P. (2003). The concept of relevance in ir. *Journal of the American Society for information Science and Technology*, 54(10):913–925.

Boudin, F., Nie, J., Bartlett, J., Grad, R., Pluye, P. et Dawes, M. (2010a). Combining classifiers for robust pico element detection. *BMC Medical Informatics and Decision Making*, 10(1):29.

Boudin, F., Nie, J. Y. et Dawes, M. (2010b). Clinical information retrieval using document and PICO structure. *In NAACL HLT*, pages 822–830.

Boudin, F., Nie, J.-Y. et Dawes, M. (2010c). Positional language models for clinical information retrieval. *In EMNLP*, pages 108–115.

Boudin, F., Nie, J.-Y. et Dawes, M. (2012). Using a medical thesaurus to predict query difficulty. *In* Baeza-Yates, R. A., de Vries, A. P., Zaragoza, H., Cambazoglu, B. B., Murdock, V., Lempel, R. et Silvestri, F., éditeurs : *ECIR*, volume 7224 de *Lecture Notes in Computer Science*, pages 480–484. Springer.

Boudin, F., Shi, L. et Nie, J.-Y. (2010d). Improving medical information retrieval with pico element detection. *In Proceedings of the 32Nd European Conference on Advances in Information Retrieval*, ECIR'2010, pages 50–61, Berlin, Heidelberg. Springer-Verlag.

Boughanem, M. et Savoy, J. (2008). *Recherche d'information : état des lieux et perspectives*. Collection Recherche d'information et web. Hermès science publ. Lavoisier, Paris.

Brashers, D. E., Neidig, J. L., Haas, S. M., Dobbs, L. K., Cardillo, L. W. et Russell, J. A. (2000). Communication in the management of uncertainty : The case of persons living with hiv or aids. *Communication Monographs*, 67(1):63–84.

Broder, A. (2002). A taxonomy of web search. *SIGIR Forum*, 36(2):3–10.

Buckley, C., Salton, G. et Allan, J. (1992). Automatic retrieval with locality information using smart. *In* Harman, D. K., éditeur : *TREC*, volume Special Publication 500-207, pages 59–72. National Institute of Standards and Technology (NIST).

Buckley, C. et Voorhees, E. M. (2000). Evaluating evaluation measure stability. *In Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '00, pages 33–40, New York, NY, USA. ACM.

Burstein, F., Fisher, J., McKemmish, S., Manaszewicz, R. et Malhotra, P. (2005). User centred quality health information provision : Benefits and challenges. *In 38th Hawaii International Conference on System Sciences (HICSS-38 2005), CD-ROM / Abstracts Proceedings, 3-6 January 2005, Big Island, HI, USA*.

Califf, R. M., McCall, J. et Harrington, R. A. (2013). Assessing research results in the medical literature : trust but verify. *JAMA internal medicine*, 173(12):1053–1055.

Carpineto, C., de Mori, R., Romano, G. et Bigi, B. (2001). An information-theoretic approach to automatic query expansion. *ACM Trans. Inf. Syst.*, 19(1):1–27.

Carpineto, C. et Romano, G. (2012). A survey of automatic query expansion in information retrieval. *ACM Comput. Surv.*, 44(1):1 :1–1 :50.

Carpineto, C., Romano, G. et Giannini, V. (2002). Improving retrieval feedback with multiple term-ranking function combination. *ACM Trans. Inf. Syst.*, 20(3):259–290.

Cartright, M.-A., White, R. W. et Horvitz, E. (2011). Intentions and attention in exploratory health search. *In Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, SIGIR '11, pages 65–74, New York, NY, USA. ACM.

Case, D. O. (2012). *Looking for information : A survey of research on information seeking, needs and behavior*. Emerald Group Publishing.

Chabou, S. et Iglewski, M. (2015). Pico extraction by combining the robustness of machine-learning methods with the rule-based methods. *In Information Technology and Computer Applications Congress (WCITCA), 2015 World Congress on*, pages 1–4. IEEE.

Chung, G. Y. (2009). Sentence retrieval for abstracts of randomized controlled trials. *BMC Med Inform Decis Mak*, 9:10.

Clarke, M. A., Belden, J. L., Koopman, R. J., Steege, L. M., Moore, J. L., Canfield, S. M. et Kim, M. S. (2013). Information needs and information-seeking behaviour analysis of primary care physicians and nurses : a literature review. *Health Information Libraries Journal*, 30(3):178–190.

Cleverdon, C. W. (1991). The significance of the cranfield tests on index languages. *In Proceedings of the 14th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '91, pages 3–12, New York, NY, USA. ACM.

Clough, P. D. et Sanderson, M. (2013). Evaluating the performance of information retrieval systems using test collections. *Inf. Res.*, 18(2).

Cook, D. J., Mulrow, C. D. et Haynes, R. B. (1997). Systematic reviews : Synthesis of best evidence for clinical decisions. *Annals of Internal Medicine*, 126(5):376–380.

Córdoba, J. M., López, M. J. M., Díaz, N. P. C., Vázquez, J. M., Aparicio, F., de Buenaga Rodríguez, M., Glez-Peña, D. et Fdez-Riverola, F. (2011). Medical-miner at trec 2011 medical records track. *In TREC*.

Cronen-Townsend, S. et Croft, W. B. (2002). Quantifying query ambiguity. HLT '02, pages 104–109, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Cui, H., Wen, J.-R., Nie, J.-Y. et Ma, W.-Y. (2003). Query expansion by mining user logs. *IEEE Trans. on Knowl. and Data Eng.*, 15(4):829–839.

Cullen, R. J. (2002). In search of evidence : family practitioners' use of the internet for clinical information. *JOURNAL-MEDICAL LIBRARY ASSOCIATION*, 90:370–379.

Curé, O., Maurer, H., Shah, N. et LePendu, P. (2013). Refining health outcomes of interest using formal concept analysis and semantic query expansion. *In Proceedings of the 7th International Workshop on Data and Text Mining in Biomedical Informatics*, DTMBIO '13, pages 5–6, New York, NY, USA. ACM.

Cutrell, E. et Guan, Z. (2007). What are you looking for ? : An eye-tracking study of information usage in web search. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '07, pages 407–416, New York, NY, USA. ACM.

Da Costa Pereira, C., Dragoni, M. et Pasi, G. (2009). Multidimensional Relevance : A New Aggregation Criterion. *In* Boughanem, M., Berrut, C., Mothe, J. et Soule-Dupuy, C., éditeurs : *Advances in Information Retrieval*, volume 5478 de *Lecture Notes in Computer Science*, chapitre 25, pages 264–275. Springer Berlin Heidelberg, Berlin, Heidelberg.

Daoud, M., Kasperowicz, D., Miao, J. et Huang, J. (2011). York university at trec 2011 : Medical records track. *In TREC*.

Darmoni, S. et Joubert, M. (2000). Cismef. *Methods of information in medicine*, 39(1):30–35.

Dawes, M., Pluye, P., Shea, L., Grad, R., Greenberg, A. et Nie, J.-Y. (2007). The identification of clinically important elements within medical journal abstracts : Patient–population–problem, exposure–intervention, comparison, outcome, duration and results (pecodr). *Informatics in Primary care*, 15:9–16.

Dawes, M. et Sampson, U. (2003). Knowledge management in clinical practice : a systematic review of information seeking behavior in physicians. *Int J Med Inform*, 71(1):9–15.

Demner-Fushman, D. et Lin, J. (2007). Answering clinical questions with knowledge-based and statistical techniques. *Comput. Linguist.*, 33(1):63–103.

Diem, L., Chevallet, J.-P. et Thuy, D. T. B. (2007). Thesaurus-based query and document expansion in conceptual indexing with umls.

Dinh, D. et Tamine, L. (2010). Sense-based biomedical indexing and retrieval. *In International Conference on Application of Natural Language to Information Systems*, pages 24–35. Springer.

Dinh, D. et Tamine, L. (2011a). Biomedical concept extraction based on combining the content-based and word order similarities. *In Proceedings of the 2011 ACM Symposium on Applied Computing*, SAC '11, pages 1159–1163, New York, NY, USA. ACM.

Dinh, D. et Tamine, L. (2011b). Combining global and local semantic contexts for improving biomedical information retrieval. *In Proceedings of the 33rd European Conference on Advances in Information Retrieval*, ECIR'11, pages 375–386, Berlin, Heidelberg. Springer-Verlag.

Dinh, D. et Tamine, L. (2011c). Voting techniques for a multi-terminology based biomedical information retrieval (regular paper). *In Conference on Artificial Intelligence in Medicine (AIME), Bled, Slovenia, 02/07/2011-06/07/2011*, volume 6747 de *LNAI*, pages 184–193, http ://www.springerlink.com. Springer. acceptance rate : 26

Dixon, B. E., Lai, P. et Grannis, S. J. (2013). Variation in information needs and quality : Implications for public health surveillance and biomedical informatics. *In AMIA 2013, American Medical Informatics Association Annual Symposium, Washington, DC, USA, November 16-20, 2013*.

Dogan, R. I., Murray, G. C., Névéol, A. et Lu, Z. (2009). Understanding pubmed® user search behavior through log analysis. *Database*, 2009.

Doszkocs, T. (1979). *AID : An Associative Interactive Dictionary for Online Bibliographic Searching*. University of Maryland.

Downing, R. E., Moore, J. L. et Brown, S. W. (2005). The effects and interaction of spatial visualization and domain expertise on information seeking. *Computers in Human Behavior*, 21(2):195–209.

Duan, L., Dong, S., Cui, S. et Ma, W. (2016). *Proceedings of ELM-2015 Volume 1 : Theory, Algorithms and Applications (I)*, chapitre Extreme Learning Machine with Gaussian Kernel Based Relevance Feedback Scheme for Image Retrieval, pages 397–408. Springer International Publishing, Cham.

Eastin, M. S. et Guinsler, N. M. (2006). Worried and wired : Effects of health anxiety on information-seeking and health care utilization behaviors. *Cyberpsy., Behavior, and Soc. Networking*, 9(4):494–498.

Efron, M., Organisciak, P. et Fenlon, K. (2012). Improving retrieval of short texts through document expansion. *In Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '12, pages 911–920, New York, NY, USA. ACM.

Efthimiadis, E. N. (1996). Query expansion. *Annual review of information science and technology*, 31:121–187.

Ely, J., Osheroff, J. A., Ebell, M. H., Chambliss, M. L., Vinson, D., Stevermer, J. J. et Pifer, E. A. (2002). Obstacles to answering doctors' questions about patient care with evidence : Qualitative study. *BMJ*, 324(7339):710.

Ely, J. W., Osheroff, J. A., Ebell, M. H., Bergus, G. R., Levy, B. T., Chambliss, M. L. et Evans, E. R. (1999). Analysis of questions asked by family doctors regarding patient care. *BMJ*, 319(7206):358–361.

Ely, J. W., Osheroff, J. A., Gorman, P. N., Ebell, M. H., Chambliss, M. L., Pifer, E. A. et Stavri, P. Z. (2000). A taxonomy of generic clinical questions : classification study. *BMJ*, 321(7258):429–432.

Eysenbach, G. et Köhler, C. (2002). How do consumers search for and appraise health information on the world wide web ? qualitative study using focus groups, usability tests, and in-depth interviews. *BMJ*, 324(7337):573–577.

Fernandez-Beltran, R. et Pla, F. (2016). Latent topics-based relevance feedback for video retrieval. *Pattern Recognition*, 51:72 – 84.

Fineout-Overholt, E. et Johnston, L. (2005). Teaching ebp : Asking searchable, answerable clinical questions. *Worldviews on Evidence-Based Nursing*, 2(3):157–160.

Fox, S. et Duggan, M. (2013). The social life of health information. Rapport technique, Pew Internet & American Life Project.

Fox, S. et Jones, S. (2009). The social life of health information. Rapport technique, Pew Internet & American Life Project.

Francke, A., Smit, M. et de Veer, A. (2008). Factors influencing the implementation of clinical guidelines for health care professionals : a systematic meta-review. *BMC Medical Information Decision Making*, 8:8–38.

Freund, L., Toms, E. G. et Waterhouse, J. (2005). Modeling the information behaviour of software engineers using a work-task framework. *Proceedings of the American Society for Information Science and Technology*, 42(1).

Furnas, G. W., Landauer, T. K., Gomez, L. M. et Dumais, S. T. (1987). The vocabulary problem in human-system communication. *Commun. ACM*, 30(11):964–971.

Ghezaiel, L. B., Latiri, C. C. et Ahmed, M. B. (2012). Conceptual indexing documents in IR based on ontology enrichment. *In Advances in Knowledge-Based and Intelligent Information and Engineering Systems - 16th Annual KES Conference, San Sebastian, Spain, 10-12 September 2012*, pages 1920–1931.

Gobeill, J., Ruch, P. et Zhou, X. (2009). Query and document expansion with medical subject headings terms at medical imageclef 2008. *In* Peters, C., Deselaers, T., Ferro, N., Gonzalo, J., Jones, G., Kurimo, M., Mandl, T., Peñas, A. et Petras, V., éditeurs : *Evaluating Systems for Multilingual and Multimodal Information Access*, volume 5706 de *Lecture Notes in Computer Science*, pages 736–743. Springer Berlin Heidelberg.

Goeuriot, L., Jones, G. J., Kelly, L., Leveling, J., Hanbury, A., Müller, H., Salantera, S., Suominen, H. et Zuccon, G. (2013). Share/clef ehealth evaluation lab 2013, task 3 : Information retrieval to address patients' questions when reading clinical reports. *CLEF 2013 Online Working Notes*, 8138.

Goeuriot, L., Kelly, L., Li, W., Palotti, J., Zuccon, G., Hanbury, A., Jones, G. J. et Müller, H. (2014). Share/clef ehealth evaluation lab 2014, task 3 : User-centred health information retrieval. *CLEF 2014 Online Working Notes*, 1180:43–61.

Goeuriot, L., Kelly, L., Suominen, H., Hanlen, L., Névéol, A., Grouin, C., Palotti, J. et Zuccon, G. (2015). *Experimental IR Meets Multilinguality, Multimodality, and Interaction : 6th International Conference of the CLEF Association, CLEF'15, Toulouse, France, September 8-11, 2015, Proceedings*, chapitre Overview of the CLEF eHealth Evaluation Lab 2015, pages 429–443. Springer International Publishing, Cham.

Gorman, P. N. et Helfand, M. (1995). Information seeking in primary care how physicians choose which clinical questions to pursue and which to leave unanswered. *Medical Decision Making*, 15(2):113–119.

Graber, M. A., Bergus, G. R. et York, C. (1999). Using the World Wide Web to answer clinical questions : how efficient are different methods of information retrieval ? *The Journal of family practice*, 48(7):520–524.

Greenhalgh, T. (2004). Effectiveness and efficiency : Random reflections on health services. *BMJ*, 328(7438):529.

Griffin, R. J., Neuwirth, K., Dunwoody, S. et Giese, J. (2004). Information sufficiency and risk communication. *Media Psychology*, 6(1):23–61.

Hammache, A., Boughanem, M. et Ahmed-Ouamer, R. (2013). Pseudo-réinjection de pertinence basée sur un modèle de langue mixte combinant les termes simples et composés. *In CORIA 2013*

- Conférence en Recherche d'Infomations et Applications - 10th French Information Retrieval Conference, Neuchâtel, Suisse, April 3-5, 2013., pages 175–190.

Hanauer, D. A. (2006). Emerse : the electronic medical record search engine. *In AMIA annual symposium proceedings*, volume 2006, page 1189. American Medical Informatics Association.

Hanbury, A. (2012). Medical information retrieval : An instance of domain-specific search. *In Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '12, pages 1191–1192, New York, NY, USA. ACM.

Haux, R., Grothe, W., Runkel, M., Schackert, H., Windeler, H.-J., Winter, A., Wirtz, R., Herfarth, C. et Kunze, S. (1996). Knowledge retrieval as one type of knowledge-based decision support in medicine : results of an evaluation study. *International Journal of Bio-Medical Computing*, 41(2):69 – 86.

Haynes, R. B., McKibbon, K. A., Wilczynski, N. L., Walter, S. D. et Werre, S. R. (2005). Optimal search strategies for retrieving scientifically strong studies of treatment from medline : analytical survey. *BMJ*, 330(7501):1179.

Hazimeh, H. et Zhai, C. (2015). Axiomatic analysis of smoothing methods in language models for pseudo-relevance feedback. *In Proceedings of the 2015 International Conference on The Theory of Information Retrieval*, ICTIR '15, pages 141–150, New York, NY, USA. ACM.

Hembrooke, H. A., Granka, L. A., Gay, G. K. et Liddy, E. D. (2005). The effects of expertise and feedback on search term selection and subsequent learning : Research articles. *J. Am. Soc. Inf. Sci. Technol.*, 56(8):861–871.

Hersh, W. (2009). *Information Retrieval : A Health and Biomedical Perspective*. Springer, 3 édition.

Hersh, W., Price, S. et Donohoe, L. (2000). Assessing thesaurus-based query expansion using the umls metathesaurus. *In AMIA Symposium*, pages 344–348.

Hersh, W. et Voorhees, E. (2009). Trec genomics special issue overview. *Inf. Retr.*, 12(1):1–15.

Hersh, W. R. (2006). Viewpoint paper : Who are the informaticians ? what we know and should know. *JAMIA*, 13(2):166–170.

Hersh, W. R. et Bhupatiraju, R. T. (2003). TREC GENOMICS track overview. *In Proceedings of The Twelfth Text REtrieval Conference, TREC 2003, Gaithersburg, Maryland, USA, November 18-21, 2003*, pages 14–23.

Hersh, W. R., Bhupatiraju, R. T., Greene, P. S., Smothers, V. et Cohen, C. (2006a). Adopting e-learning standards in health care : Competency-based learning in the medical informatics domain. *In AMIA 2006, American Medical Informatics Association Annual Symposium, Washington, DC, USA, November 11-15, 2006.*

Hersh, W. R., Bhupatiraju, R. T., Ross, L., Cohen, A. M., Kraemer, D. et Johnson, P. (2004). TREC 2004 genomics track overview. *In Proceedings of the Thirteenth Text REtrieval Conference, TREC 2004, Gaithersburg, Maryland, USA, November 16-19, 2004.*

Hersh, W. R., Cohen, A. M., Roberts, P. M. et Rekapalli, H. K. (2006b). TREC 2006 genomics track overview. *In Proceedings of the Fifteenth Text REtrieval Conference, TREC 2006, Gaithersburg, Maryland, USA, November 14-17, 2006.*

Hersh, W. R., Cohen, A. M., Ruslen, L. et Roberts, P. M. (2007). TREC 2007 genomics track overview. *In Proceedings of The Sixteenth Text REtrieval Conference, TREC 2007, Gaithersburg, Maryland, USA, November 5-9, 2007.*

Hersh, W. R., Cohen, A. M., Yang, J., Bhupatiraju, R. T., Roberts, P. M. et Hearst, M. A. (2005). TREC 2005 genomics track overview. *In Proceedings of the Fourteenth Text REtrieval Conference, TREC 2005, Gaithersburg, Maryland, USA, November 15-18, 2005*.

Hersh, W. R., Crabtree, M. K., Hickam, D. H., Sacherek, L., Friedman, C. P., Tidmarsh, P., Mosbaek, C. et Kraemer, D. (2002). Research paper : Factors associated with success in searching medline and applying evidence to answer clinical questions. *JAMIA*, 9(3):283–293.

Herskovic, J. R., Tanaka, L. Y., Hersh, W. et Bernstam, E. V. (2007a). A day in the life of pubmed : analysis of a typical day's query log. *Journal of the American Medical Informatics Association*, 14(2):212–220.

Herskovic, J. R., Tanaka, L. Y., Hersh, W. R. et Bernstam, E. V. (2007b). Research paper : A day in the life of pubmed : Analysis of a typical day's query log. *JAMIA*, 14(2):212–220.

Heuwing, B. et Mandl, T. (2007). Robust retrieval experiments at the university of hildesheim. *In Advances in Multilingual and Multimodal Information Retrieval, 8th Workshop of the Cross-Language Evaluation Forum, CLEF 2007, Budapest, Hungary, September 19-21, 2007, Revised Selected Papers*, pages 134–136.

HOBU, P. P. M., SCHMIDT, H. G., BOSHUIZEN, H. P. A. et PATEL, V. L. (1987). Contextual factors in the activation of first diagnostic hypotheses : expert-novice differences. *Medical Education*, 21(6):471–476.

Hölscher, C. et Strube, G. (2000). Web search behavior of internet experts and newbies. *Computer networks*, 33(1):337–346.

Hong, Y., Cruz, N., Marnas, G., Early, E. et Gillis, R. (2002). A query analysis of consumer health information retrieval. *In Proceedings of Annual Symposium for biomedical and health informatics*, pages 791–792.

Hsieh-Yee, I. (1993). Effects of search experience and subject knowledge on the search tactics of novice and experienced searchers. *Journal of the American Society for Information Science*, 44(3):161.

Huang, J. et Efthimiadis, E. N. (2009). Analyzing and evaluating query reformulation strategies in web search logs. *In Proceedings of the 18th ACM Conference on Information and Knowledge Management*, CIKM '09, pages 77–86, New York, NY, USA. ACM.

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.

Inthiran, A., Alhashmi, S. M. et Ahmed, P. K. (2012). The affects of task difficulty on medical searches. *In Proceedings of the 21st International Conference on World Wide Web*, WWW '12 Companion, pages 533–534, New York, NY, USA. ACM.

Jamal, A., McKenzie, K., Clark, M. J. *et al.* (2009). The impact of health information technology on the quality of medical and health care : a systematic review. *Health Information Management Journal*, 38(3):26.

Jiang, J. et Zhai, C. (2007). An empirical study of tokenization strategies for biomedical information retrieval. *Inf. Retr.*, 10(4-5):341–363.

Jones, S. (1972). A statistical interpretation of term specificity and its application to retrieval. *Journal of documentation*, 28(1):11–20.

Jonnalagadda, S., Cohen, T., Wu, S. et Gonzalez, G. (2012). Enhancing clinical concept extraction with distributional semantics. *Journal of biomedical informatics*, 45(1):129–140.

Kalpathy-Cramer, J., Müller, H., Bedrick, S., Eggel, I., García Seco de Herrera, A. et Tsikrika, T. (2011). The CLEF 2011 medical image retrieval and classification tasks. *In Working Notes of CLEF 2011 (Cross Language Evaluation Forum)*.

Karimi, S., Martinez, D., Ghodke, S., Cavedon, L., Suominen, H. et Zhang, L. (2011). Search for medical records : Nicta at trec 2011 medical track. *In TREC*.

Keenan, G., Yakel, E., Lopez, K. D., Tschannen, D. et Ford, Y. B. (2013). Challenges to nurses' efforts of retrieving, documenting, and communicating patient care information. *Journal of the American Medical Informatics Association*, 20(2):245–251.

Kelly, L., Goeuriot, L., Suominen, H., Schreck, T., Leroy, G., Mowery, D. L., Velupillai, S., Chapman, W. W., Martinez, D., Zuccon, G. et Palotti, J. (2014). *Information Access Evaluation. Multilinguality, Multimodality, and Interaction : 5th International Conference of the CLEF Initiative, CLEF 2014, Sheffield, UK, September 15-18, 2014. Proceedings*, chapitre Overview of the ShARe/CLEF eHealth Evaluation Lab 2014, pages 172–191. Springer International Publishing, Cham.

Keselman, A., Smith, C. A., Divita, G., Kim, H., Browne, A. C., Leroy, G. et Zeng-Treitler, Q. (2008). Consumer Health Concepts That Do Not Map to the UMLS : Where Do They Fit ? *J Am Med Inform Assoc*, 15(4):496–505.

Kim, S. N., Martinez, D., Cavedon, L. et Yencken, L. (2011). Automatic classification of sentences to support evidence based medicine. *BMC bioinformatics*, 12(2):1.

King, B., Wang, L., Provalov, I. et Zhou, J. (2011). Cengage learning at TREC 2011 medical track. *In Proceedings of The Twentieth Text REtrieval Conference, TREC 2011, Gaithersburg, Maryland, USA, November 15-18, 2011*.

Kostagiolas, P., Korfiatis, N., Kourouthanasis, P. et Alexias, G. (2014). Work-related factors influencing doctors search behaviors and trust toward medical information resources. *International Journal of Information Management*, 34(2):80–88.

Kules, B. et Xie, B. (2011). Older adults searching for health information in medlineplus – an exploratory study of faceted online search interfaces. *Proceedings of the American Society for Information Science and Technology*, 48(1):1–10.

Kwan, P. W., Welch, M. C., Foley, J. J., Kwan, P., Welch, M. et Foley, J. (2015). A knowledge-based decision support system for adaptive fingerprint identification that uses relevance feedback. *Knowledge-Based Systems*, 73(Complete):236–253.

Lambert, S. D. et Loiselle, C. G. (2007). Health information—seeking behavior. *Qualitative Health Research*, 17(8):1006–1019.

LaRock, T., Mathews, L., Roberts, M., Lim, D. et Small, S. G. (2014). Siena's twitter information retrieval system : The 2014 microblog track. *In Proceedings of The Twenty-Third Text REtrieval Conference, TREC 2014, Gaithersburg, Maryland, USA, November 19-21, 2014*.

Latiri, C., Haddad, H. et Hamrouni, T. (2012). Towards an effective automatic query expansion process using an association rule mining approach. *Journal of Intelligent Information Systems*, 39(1):209–247.

Lee, K. S., Croft, W. B. et Allan, J. (2008). A cluster-based resampling method for pseudo-relevance feedback. *In Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '08, pages 235–242, New York, NY, USA. ACM.

Li, C. et Wang, J. (2012). A clustering approach to improving pseudo-relevance feedback : Improving retrieval effetiveness by removing noisy documents. *2012 Fourth International Symposium on Information Science and Engineering*, 0:35–38.

Limsopatham, N., Macdonald, C. et Ounis, I. (2015). Modelling the usefulness of document collections for query expansion in patient search. *In Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, CIKM '15, pages 1739–1742, New York, NY, USA. ACM.

Liu, C., Liu, J., Cole, M., Belkin, N. J. et Zhang, X. (2012). Task difficulty and domain knowledge effects on information search behaviors. *Proceedings of the American Society for Information Science and Technology*, 49(1):1–10.

Liu, R.-L. et Huang, Y.-C. (2011). Medical query generation by term-category correlation. *Inf. Process. Manage.*, 47(1):68–79.

Liu, S., Liu, F., Yu, C. et Meng, W. (2004). An effective approach to document retrieval via utilizing wordnet and recognizing phrases. *In Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '04, pages 266–272, New York, NY, USA. ACM.

Lundahl, B., Moleni, T., Burke, B. L., Butters, R., Tollefson, D., Butler, C. et Rollnick, S. (2013). Motivational interviewing in medical care settings : A systematic review and meta-analysis of randomized controlled trials. *Patient Education and Counseling*, 93(2):157 – 168.

Luo, G. (2009). Lessons learned from building the imed intelligent medical search engine. *In Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*, pages 5138–5142. IEEE.

Luo, G. et Tang, C. (2008). On iterative intelligent medical search. *In Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '08, pages 3–10, New York, NY, USA. ACM.

Luo, G., Tang, C., Yang, H. et Wei, X. (2008). Medsearch : A specialized search engine for medical information retrieval. *In Proceedings of the 17th ACM Conference on Information and Knowledge Management*, CIKM '08, pages 143–152, New York, NY, USA. ACM.

Lykke, M., Price, S. et Delcambre, L. M. L. (2012). How doctors search : A study of query behaviour and the impact on search results. *Inf. Process. Manage.*, 48(6):1151–1170.

Magrabi, F., Coiera, E. W., Westbrook, J. I., Gosling, A. S. et Vickland, V. (2005). General practitioners' use of online evidence during consultations. *I. J. Medical Informatics*, 74(1):1–12.

Mandala, R., Tokunaga, T. et Tanaka, H. (1999). Combining multiple evidence from different types of thesaurus for query expansion. *In Proceedings of the 22Nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '99, pages 191–197, New York, NY, USA. ACM.

Manning, C. D., Raghavan, P. et Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press, New York, NY, USA.

Mao, J., Lu, K., Mu, X. et Li, G. (2015). Mining document, concept, and term associations for effective biomedical retrieval : Introducing mesh-enhanced retrieval models. *Inf. Retr.*, 18(5):413–444.

Marchionini, G. (1995). *Information Seeking in Electronic Environments*. Cambridge University Press, New York, NY, USA.

Marchionini, G., Dwiggins, S., Katz, A. et Lin, X. (1993). Information seeking in full-text end-user-oriented search systems : The roles of domain and search expertise. *Library & information science research*, 15(1):35–69.

Maron, M. E. et Kuhns, J. L. (1960). On relevance, probabilistic indexing and information retrieval. *J. ACM*, 7(3):216–244.

McCray, A. T. et Tse, T. (2003). Understanding search failures in consumer health information systems. *AMIA*, pages 430–434.

McKnight, L. et Srinivasan, P. (2003). Categorization of sentence types in medical abstracts. *In AMIA*.

Merry, D. E. et Korsmeyer, S. J. (1997). Bcl-2 gene family in the nervous system. *Annual Review of Neuroscience*, 20(1):245–267.

Miller, S. A. et Forrest, J. L. (2001). Enhancing your practice through evidence-based decision making : Pico, learning how to ask good questions. *Journal of Evidence Based Dental Practice*, 1(2):136 – 141.

Min, J., Leveling, J., Zhou, D. et Jones, G. J. F. (2010). Document expansion for image retrieval. *In Adaptivity, Personalization and Fusion of Heterogeneous Information*, RIAO '10, pages 65–71, Paris, France, France. LE CENTRE DE HAUTES ETUDES INTERNATIONALES D'INFORMATIQUE DOCUMENTAIRE.

Mitra, M., Singhal, A. et Buckley, C. (1998). Improving automatic query expansion. *In Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '98, pages 206–214, New York, NY, USA. ACM.

Mizzaro, S. (1997). Relevance : The whole history. *J. Am. Soc. Inf. Sci.*, 48(9):810–832.

Moturu, S., Liu, H. et Johnson, W. G. (2008). Trust evaluation in health information on the world wide web. *In Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*, pages 1525–1528.

Mu, X., Ryu, H. et Lu, K. (2011). Supporting effective health and biomedical information retrieval and navigation : A novel facet view interface evaluation. *J. of Biomedical Informatics*, 44(4): 576–586.

Müller, H., Deselaers, T., Deserno, T. M., Kalpathy-Cramer, J., Kim, E. et Hersh, W. (2007). Overview of the imageclefmed 2007 medical retrieval and medical annotation tasks. *In Advances in Multilingual and Multimodal Information Retrieval*, pages 472–491. Springer.

Murad, M. H., Montori, V. M., Ioannidis, J. P., Jaeschke, R., Devereaux, P., Prasad, K., Neumann, I., Carrasco-Labra, A., Agoritsas, T., Hatala, R. *et al.* (2014). How to read a systematic review and meta-analysis and apply the results to patient care : users' guides to the medical literature. *Jama*, 312(2):171–179.

Nabarette, H. (2002). L'internet médical et la consommation d'information par les patients. *Réseaux*, (4):249–286.

Natarajan, K., Stein, D., Jain, S. et Elhadad, N. (2010). An analysis of clinical queries in an electronic health record search utility. *International journal of medical informatics*, 79(7):515–522.

Nawab, R., Stevenson, M. et Clough, P. (2016). An ir-based approach utilising query expansion for plagiarism detection in medline. *International Journal of Computational Biology and Drug Design*.

Nicholas, D., Huntington, P., Gunter, B., Withey, R. et Russell, C. (2003). The british and their use of the web for health information and advice : a survey. *In Aslib Proceedings*, volume 55, pages 261–276. MCB UP Ltd.

Nourbakhsh, E., Nugent, R., Wang, H., Cevik, C. et Nugent, K. (2012). Medical literature searches : a comparison of pubmed and google scholar. *Health Information Libraries Journal*, 29(3):214–222.

Oh, S. (2012). The characteristics and motivations of health answerers for sharing information, knowledge, and experiences in online environments. *JASIST*, 63(3):543–557.

Ounis, I., Macdonald, C., de Rijke, M., Mishne, G. et Soboroff, I. (2006). Overview of the TREC 2006 blog track. *In Proceedings of the Fifteenth Text REtrieval Conference, TREC 2006, Gaithersburg, Maryland, USA, November 14-17, 2006*.

Packer, H. S., Samangooei, S., Hare, J. S., Gibbins, N. et Lewis, P. H. (2012). Event detection using twitter and structured semantic query expansion. *In Proceedings of the 1st International Workshop on Multimodal Crowd Sensing*, CrowdSens '12, pages 7–14, New York, NY, USA. ACM.

Pandolfini, C. et Bonati, M. (2002). Follow up of quality of public oriented health information on the world wide web : systematic re-evaluation. *BMJ*, 324(7337):582–583.

Pang, P. C.-I., Verspoor, K., Chang, S. et Pearce, J. M. (2014). Designing for health exploratory seeking behaviour.

Pao, M., Grefsheim, S., Barclay, M., Woolliscroft, J., Shipman, B. et M., M. (1994). Effect of search experience on sustained medline usage by students. *Academic Medicine*.

Patel, S. (1994). *Effectiveness of Expert Semantic Knowledge as a Navigational Aid Within the Hypertext*. State University of New York at Buffalo.

Paton, C. R. (1999). Evidence-based medicine ; evidence-based health care : How to make health policy and management decisions. *BMJ*, 318(7177):201.

Pereira, C. D. C., Dragoni, M. et Pasi, G. (2012). Multidimensional relevance : Prioritized aggregation in a personalized information retrieval setting. *Information Processing Management*, 48(2):340 – 357.

Petticrew, M. et Roberts, H. (2006). *Systematic Reviews in the Social Sciences : A Practical Guide*. Blackwell Publishing.

Pirkola, A. et Järvelin, K. (2001). Employing the resolution power of search keys. *JASIST*, 52(7):575–583.

Plovnick, M. R. et Zeng, T. Q. (2004). Reformulation of consumer health queries with professional terminology : A pilot study. *J Med Internet Res*, 6(3):27.

Ponte, J. M. et Croft, W. B. (1998). A language modeling approach to information retrieval. *In Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '98, pages 275–281, New York, NY, USA. ACM.

Pratt, W. et Wasserman, H. (2000). QueryCat : automatic categorization of MEDLINE queries. *Proceedings / AMIA ... Annual Symposium. AMIA Symposium*, pages 655–659.

Radhouani, S., Kalpathy-Cramer, J., Bedrick, S., Bakke, B. et Hersh, W. R. (2009). Using media fusion and domain dimensions to improve precision in medical image retrieval. *In* Peters, C., Caputo, B., Gonzalo, J., Jones, G. J. F., Kalpathy-Cramer, J., Müller, H. et Tsikrika, T., éditeurs : *CLEF (2)*, volume 6242 de *Lecture Notes in Computer Science*, pages 223–230. Springer.

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):A12–3.

Richesson, R. L., Shereff, D., Spisla, C., Albarracin, N., Konicek, D. J. et Andrews, J. E. (2010). The use of snomed ct to support retrieval and re-use of question and answer sets for patient registries. *I. J. Functional Informatics and Personalised Medicine*, 3(4):342–365.

Robertson, S., Walker, S., Jones, S., Hancock-Beaulieu, M. et Gatford, M. (1996). Okapi at trec-3. pages 109–126.

Robertson, S. E. (1991). On term selection for query expansion. *J. Doc.*, 46(4):359–364.

Robertson, S. E. (1997). Readings in information retrieval. chapitre The Probability Ranking Principle in IR, pages 281–286. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

Robertson, S. E. et Sparck Jones, K. (1988). Document retrieval systems. chapitre Relevance Weighting of Search Terms, pages 143–160. Taylor Graham Publishing, London, UK, UK.

Rocchio, J. J. (1971). Relevance feedback in information retrieval. *In* Salton, G., éditeur : *The Smart retrieval system - experiments in automatic document processing*, pages 313–323. Englewood Cliffs, NJ : Prentice-Hall.

Ruch, P. (2006). Automatic assignment of biomedical categories : toward a generic approach. *Bioinformatics*, 22(6):658–664.

Sackett, D. L., Rosenberg, W. M. C., Gray, J. A. M., Haynes, R. B. et Richardson, W. S. (1996). Evidence based medicine : what it is and what it isn't. *BMJ*, 312(7023):71–72.

Sadasivam, R. S., Kinney, R. L., Lemon, S. C., Shimada, S. L., Allison, J. J. et Houston, T. K. (2013). Internet health information seeking is a team sport : Analysis of the pew internet survey. *International Journal of Medical Informatics*, 82(3):193 – 200.

Salton, G. (1969). A comparison between manual and automatic indexing methods. *American Documentation*, 20(1):61–71.

Salton, G. (1970). Evaluation problems in interactive information retrieval. *Information Storage and Retrieval*, 6(1):29–44.

Salton, G. et Buckley, C. (1997). Readings in information retrieval. chapitre Improving Retrieval Performance by Relevance Feedback, pages 355–364. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

Salton, G. et McGill, M. J. (1986). *Introduction to Modern Information Retrieval*. McGraw-Hill, Inc., New York, NY, USA.

Salton, G., Wong, A. et Yang, C. S. (1975). A vector space model for automatic indexing. *Commun. ACM*, 18(11):613–620.

Sanderson, M. (2010). Test collection based evaluation of information retrieval systems. *Foundations and Trends in Information Retrieval*, 4(4):247–375.

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.

Schlosser, R. W. et O'Neil-Pirozzi, T. (2006). Problem formulation in evidence-based practice and systematic reviews. *Contemp Issues Commun Sci Disord*, 33:5–10.

Simpson, M. S., Voorhees, E. M. et Hersh, W. (2014). Overview of the trec 2014 clinical decision support track. Rapport technique, DTIC Document.

Singhal, A. et Pereira, F. (1999). Document expansion for speech retrieval. *In Proceedings of the 22Nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '99, pages 34–41, New York, NY, USA. ACM.

Smeaton, A. F., Over, P. et Kraaij, W. (2006). Evaluation campaigns and trecvid. *In Proceedings of the 8th ACM International Workshop on Multimedia Information Retrieval*, MIR '06, pages 321–330, New York, NY, USA. ACM.

Smith, R. (1996). What clinical information do doctors need ? *BMJ*, 313(7064):1062–1068.

Soldaini, L., Cohan, A., Yates, A., Goharian, N. et Frieder, O. (2015a). Query reformulation for clinical decision support search. *The Twenty-Third Text REtrieval Conference Proceedings (TREC 2014)*.

Soldaini, L., Cohan, A., Yates, A., Goharian, N. et Frieder, O. (2015b). Retrieving medical literature for clinical decision support. *In* Hanbury, A., Kazai, G., Rauber, A. et Fuhr, N., éditeurs : *Advances in Information Retrieval*, volume 9022 de *Lecture Notes in Computer Science*, pages 538–549. Springer International Publishing.

Soldaini, L., Yates, A., Yom-Tov, E., Frieder, O. et Goharian, N. (2015c). Enhancing web search in the medical domain via query clarification. *Information Retrieval Journal*, 19(1):149–173.

Song, F. et Croft, W. B. (1999). A general language model for information retrieval. *In Proceedings of the Eighth International Conference on Information and Knowledge Management*, CIKM '99, pages 316–321, New York, NY, USA. ACM.

Song, M., Spallek, H., Polk, D., Schleyer, T. et Wali, T. (2010). How information systems should support the information needs of general dentists in clinical settings : suggestions from a qualitative study. *BMC Medical Informatics and Decision Making*, 10(1):1–9.

Spink, A. et Jansen, B. J. (2006). *Web search : Public searching of the Web*, volume 6. Springer Science & Business Media.

Spink, A., Yang, Y., Jansen, J., Nykanen, P., Lorence, D. P., Ozmutlu, S. et Ozmutlu, H. C. (2004). A study of medical and health queries to web search engines. *Health Information & Libraries Journal*, 21(1):44–51.

Srinivasan, P. (1996). Query expansion and medline. *Inf. Process. Manage.*, 32(4):431–443.

Steve, C. R. et Croft, W. (2002). Quantifying query ambiguity. *In Proceedings of the second international conference on Human Language Technology Research*, HLT '02, pages 104–109, San Francisco, CA, USA.

Stokes, N., Li, Y., Cavedon, L. et Zobel, J. (2009). Exploring criteria for successful query expansion in the genomic domain. *Inf. Retr.*, 12(1):17–50.

Suominen, H., Salanter, S., Velupillai, S., Chapman, W., Savova, G., Elhadad, N., Pradhan, S., South, B., Mowery, D., Jones, G., Leveling, J., Kelly, L., Goeuriot, L., Martinez, D. et Zuccon, G. (2013). Overview of the share/clef ehealth evaluation lab 2013. *In* Forner, P., Muller, H., Paredes, R., Rosso, P. et Stein, B., éditeurs : *Information Access Evaluation. Multilinguality, Multimodality, and Visualization*, volume 8138 de *Lecture Notes in Computer Science*, pages 212–231. Springer Berlin Heidelberg.

Tamine, L., Chouquet, C. et Palmer, T. (2015). Analysis of biomedical and health queries : Lessons learned from trec and clef evaluation benchmarks. *Journal of the Association for Information Science and Technology*, 66(12):2626–2642.

Thain, A. et Wales, A. (2005). Information needs of specialist healthcare professionals : a preliminary study based on the west of scotland colorectal cancer managed clinical network. *Health Information & Libraries Journal*, 22(2):133–142.

Thesprasith, O. et Jaruskulchai, C. (2014). Query expansion using medical subject headings terms in the biomedical documents. *In* Nguyen, N., Attachoo, B., Trawiński, B. et Somboonviwat, K., éditeurs : *Intelligent Information and Database Systems*, volume 8397 de *Lecture Notes in Computer Science*, pages 93–102. Springer International Publishing.

Timpka, T. et Arborelius, E. (1990). The gp's dilemmas : A study of knowledge need and use during health care consultations. Rapport technique R-90-30. Also published in Methods of Information in Medicine 1990 ;29 :23-29.

Toms, E. G. et Latter, C. (2007). How consumers search for health information. *Health informatics journal*, 13(3):223–235.

Tracy Edinger, N., Cohen, A. M., Bedrick, S., Ambert, K. et Hersh, W. (2012). Barriers to retrieving patient information from electronic health record data : failure analysis from the trec medical records track.

Trieschnigg, D. (2010). *Proof of concept : concept-based biomedical information retrieval*. Thèse de doctorat, University of Twente.

Trieschnigg, R., Kraaij, W. et Schuemie, M. (2006). Concept based document retrieval for genomics literature. *In* Voorhees, E. et Buckland, L., éditeurs : *The Fifteenth Text REtrieval Conference (TREC 2006) Proceedings*, pages 453–460.

Vibert, N., Rouet, J.-F., Ros, C., Ramond, M. et Deshoullieres, B. (2007). The use of online electronic information resources in scientific research : The case of neuroscience. *Library Information Science Research*, 29(4):508 – 532.

Voorhees, E. M. (1994). Query expansion using lexical-semantic relations. *In Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '94, pages 61–69, New York, NY, USA. Springer-Verlag New York, Inc.

Voorhees, E. M. (2002). The philosophy of information retrieval evaluation. *In Revised Papers from the Second Workshop of the Cross-Language Evaluation Forum on Evaluation of Cross-Language Information Retrieval Systems*, CLEF '01, pages 355–370, London, UK, UK. Springer-Verlag.

Voorhees, E. M. (2005). The TREC robust retrieval track. *SIGIR Forum*, 39(1):11–20.

Voorhees, E. M. et Ellis, A., éditeurs (2014). *Proceedings of The Twenty-Third Text REtrieval Conference, TREC 2014, Gaithersburg, Maryland, USA, November 19-21, 2014*, volume Special Publication 500-308. National Institute of Standards and Technology (NIST).

Wang, J., Xiao, N. et Rao, H. R. (2012). An exploration of risk information search via a search engine : Queries and clicks in healthcare and information security. *Decision Support Systems*, 52(2):395 – 405.

Weifield, J. et Finkelstein, K. (1996). How to answer your clinical questions more efficiently. *Family practice management*, 12(7):37.

White, R. et Moris, D. (2008). How medical expertise influences web search behaviour. *In Proceedings of the 31st international ACM SIGIR conference on Research and development in Information Retrieval*, SIGIR '08, pages 791–792.

White, R. W., Dumais, S. et Teevan, J. (2008). How medical expertise influences web search interaction. *In Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '08, pages 791–792, New York, NY, USA. ACM.

White, R. W., Dumais, S. T. et Teevan, J. (2009). Characterizing the influence of domain expertise on web search behavior. *In Proceedings of the Second ACM International Conference on Web Search and Data Mining*, WSDM '09, pages 132–141, New York, NY, USA. ACM.

White, R. W. et Horvitz, E. (2009a). Cyberchondria : Studies of the escalation of medical concerns in web search. *ACM Trans. Inf. Syst.*, 27(4):23 :1–23 :37.

White, R. W. et Horvitz, E. (2009b). Cyberchondria : Studies of the escalation of medical concerns in web search. *ACM Trans. Inf. Syst.*, 27(4):23 :1–23 :37.

White, R. W. et Horvitz, E. (2012). Studies of the onset and persistence of medical concerns in search logs. *In Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '12, pages 265–274, New York, NY, USA. ACM.

White, R. W. et Morris, D. (2007). Investigating the querying and browsing behavior of advanced search engine users. *In Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '07, pages 255–262, New York, NY, USA. ACM.

Wildemuth, B. (2004). The effects of domain-knowledge on search tactic formulation. volume 55, pages 246–258.

Wildemuth, B. M., de Bliek, R., Friedman, C. P. et File, D. D. (1995). Medical students' personal knowledge, searching proficiency, and database use in problem solving. *Journal of the American Society for Information Science*, 46(8):590–607.

Wilson, T. (2006). Revisiting user studies and information needs. *Journal of Documentation*, 62(6):680–684.

Wilson, T. D. (1999). Models in information behaviour research. *Journal of documentation*, 55(3):249–270.

Xiao, N., Sharman, R., Rao, H. et Upadhyaya, S. (2014). Factors influencing online health information search : An empirical analysis of a national cancer-related survey. *Decision Support Systems*, 57:417 – 427.

Xu, J. et Croft, B. (2006). Query expansion using local and global document analysis. *In ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 4–11.

Xu, J. et Croft, W. B. (1996). Query expansion using local and global document analysis. *In Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '96, pages 4–11, New York, NY, USA. ACM.

Yang, L., Mei, Q., Zheng, K. et Hanauer, D. A. (2011). Query log analysis of an electronic health record search engine. *AMIA*, 2011:915–924.

Yoo, I. et Mosa, A. S. M. (2015). Analysis of pubmed user sessions using a full-day pubmed query log : A comparison of experienced and nonexperienced pubmed users. *Journal of the American Society for Information Science*, 3(3).

Yu, S., Cai, D., Wen, J.-R. et Ma, W.-Y. (2003). Improving pseudo-relevance feedback in web information retrieval using web page segmentation. *In Proceedings of the 12th International Conference on World Wide Web*, WWW '03, pages 11–18.

Zeng, Q., Cimino, J. J. et Zou, K. H. (2002a). Providing concept-oriented views for clinical data using a knowledge-based system. *Journal of the American Medical Informatics Association*, 9(3):294–305.

Zeng, Q., Crowell, J., Plovnick, R., Kim, E., Ngo, L. et Dibble, E. (2006). Research paper : Assisting consumer health information retrieval with query recommendations. *Journal of American Medical Informatics Associations*, 13(1):80–90.

Zeng, Q., Kogan, S., Ash, N., Greenes, R. A. et Boxwala, A. A. (2002b). Characteristics of consumer terminology for health information retrieval. *Methods of information in medicine*, 41(4):289–298.

Zeng, Q. T., Kogan, S., Plovnick, R. M., Crowell, J., Lacroix, E.-M. et Greenes, R. A. (2004). Positive attitudes and failed queries : an exploration of the conundrums of consumer health information retrieval. *I. J. Medical Informatics*, 73(1):45–55.

Zhang, X., Anghelescu, H. G. B. et Yuan, X.-J. (2005). Domain knowledge, search behaviour, and search effectiveness of engineering and science students : an exploratory study. *Inf. Res.*, 10(2).

Zhang, Y. (2010). Contextualizing consumer health information searching : an analysis of questions in a social q/a community. IHI '10, pages 210–219, New York, NY, USA. ACM.

Zhang, Y. (2012). Consumer health information searching process in real life settings. *Proceedings of the American Society for Information Science and Technology*, 49(1):1–10.

Zhang, Y. (2014). Searching for specific health-related information in medlineplus : Behavioral patterns and user experience. *Journal of the Association for Information Science and Technology*, 65(1):53–68.

Zhang, Y. et Fu, W.-T. (2011). Designing consumer health information systems : What do user-generated questions tell us ? *In* Schmorrow, D. et Fidopiastis, C., éditeurs : *Foundations of Augmented Cognition. Directing the Future of Adaptive Systems*, volume 6780 de *Lecture Notes in Computer Science*, pages 536–545. Springer Berlin Heidelberg.

Zhao, J., yen Kan, M., Procter, P. M., Zubaidah, S., Yip, W. K. et Li, G. M. (2010). Improving search for evidence-based practice using information extraction. *BMC Medical Informatics and Decision Making*, 10(29).

Zhou, W., Yu, C., Smalheiser, N., Torvik, V. et Hong, J. (2007). Knowledge-intensive conceptual retrieval and passage extraction of biomedical literature. *In Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '07, pages 655–662, New York, NY, USA. ACM.

Zhu, D., Wu, S. T., Carterette, B. et Liu, H. (2014). Using large clinical corpora for query expansion in text-based cohort identification. *Journal of Biomedical Informatics*, 49:275–281.

Zhu, D., Wu, S. T.-I., Masanz, J. J., Carterette, B. et Liu, H. (2013). Using discharge summaries to improve information retrieval in clinical domain. *In CLEF (Working Notes).*

Zickuhr, K. (2006). Generations 2010. Rapport technique, Pew Internet & American Life Project.

Znaidi, E., Tamine, L., Chouquet, C. et Latiri, C. (2013a). Analyse exploratoire des requêtes d'experts médicaux : cas des campagnes d'évaluation TREC et CLEF (regular paper). *In* Soualmia, L. et Tamine, L., éditeurs : *Symposium sur l'Ingénierie de l'Information Médicale, Lille, 01/07/2013-05/07/2013*, page (en ligne), http ://univ-lille1.fr, France. Université de Lille.

Znaidi, E., Tamine, L., Chouquet, C. et Latiri, C. C. (2013b). Characterizing health-related information needs of domain experts. *In Artificial Intelligence in Medicine - 14th Conference on Artificial Intelligence in Medicine, AIME 2013, Murcia, Spain, May 29 - June 1, 2013. Proceedings*, pages 48–57.

Znaidi, E., Tamine, L. et Latiri, C. (2015). Answering PICO clinical questions : A semantic graph-based approach. *In Artificial Intelligence in Medicine - 15th Conference on Artificial Intelligence in Medicine, AIME 2015, Pavia, Italy, June 17-20, 2015. Proceedings*, pages 232–237.

Znaidi, E., Tamine, L. et Latiri, C. (2016a). Aggregating semantic information nuggets for answering clinical queries. *In ACM SAC- 2016. 31st Symposium on Applied Computing. Pisa, Italy, April 4-8, 2016.*

Znaidi, E., Tamine, L. et Latiri, C. (2016b). Répondre à des requêtes cliniques PICO. *In CORIA 2016 - Conférence en Recherche d'Informations et Applications- 13th French Information Retrieval Conference. CIFED 2016 Colloque International Francophone sur l'Ecrit et le Document, Toulouse, France, March 9-11, 2016, Toulouse, France, March 9-11, 2016.*, pages 447–462.