République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université Yahia Fares de MEDEA

Faculté des Sciences

Département de Mathématiques et d'Informatique



Mémoire de Fin d'Etudes

Pour l'obtention du Diplôme de Master en Informatique

Spécialité : Ingénierie des Systèmes et Technologies du Web

Thème :

Reformulation Interactive Personnalisée de la Requête Utilisateur pour améliorer la Pertinence des Résultats de Recherche d'Information

Réalisé par :

DOUMA Fatma

RAMDANI Fathia

Encadré par :

Mr.RAFA Tahar

Année Universitaire : **2023/2024**

ملخص

في البحث عن المعلومات، غالباً ما تعيق صياغة الاستعلامات غير الدقيقة من قبل المستخدمين الحصول على نتائج ذات صلة. في هذا البحث، نتناول هذه المشكلة من خلال اقتراح نهج لتخصيص البحث عن المعلومات لإعادة صياغة الاستعلامات بطريقة تفاعلية وشخصية، مع إشراك المستخدم.

يعتمد هذا النهج على حساب تعريف المستخدم الذي يجمع بين الجانب الموضوعي، المستمد من المصطلحات ذات الصلة المستخرجة من المقالات و المؤلفات السابقة، الجانب الاجتماعي، المتمثل في التعليقات التوضيحية الأكثر استخداماً من قبل المستخدم لتوصيف المؤلفات ذات الصلة. يتم استخدام حساب تعريف المستخدم، لاقتراح مجموعة من عمليات اعادة الصياغة الممكنة لجملة البجث الاولية ، ثم يعطى القرار النهائي للمستخدم لاختيار اعادة الصياغة الملائمة حسب حاجاته .

الكلمات الرئيسية: البحث عن المعلومات ، الحاجة إلى المعلومات، الاستعلام، التعليقات التوضيحية، الصلة، ترجيح المصطلح، إعادة صياغة الاستعلام، البحث عن المعلومات المخصصة، حساب تعريف المستخدم.

Résumé

Dans la recherche d'information, l'obtention de résultats pertinents est souvent entravée par la formulation imprécise des requêtes par les utilisateurs. Dans ce mémoire, nous abordons ce problème en proposant une approche de personnalisation de la recherche d'information pour reformuler les requêtes de manière interactive, impliquant l'intervention de l'utilisateur, et personnalisée.

Cette approche s'appuie sur un profil utilisateur combinant à la fois le contexte thématique, issu des termes pertinents de l'historique de recherche de l'utilisateur, et le contexte social, basé sur les tags les plus utilisés par l'utilisateur pour annoter les documents pertinents. Le profil utilisateur est exploité pour proposer un ensemble de reformulations possibles de la requête initiale de l'utilisateur, puis l'utilisateur lui-même est impliqué pour choisir la variante à envoyer à la recherche.

Mots clés : Recherche d'Information (RI), Besoin en information, Requête, tags, Pertinence, Pondération des termes, Reformulation de la requête, Recherche d'information Personnalisée, profil d'utilisateur.

Abstract

In information retrieval, obtaining relevant results is often hindered by the imprecise formulation of queries by users. In this thesis, we address this problem by proposing an approach for personalizing information retrieval to interactively and personalized reformulate queries, involving user intervention.

This approach relies on a user profile that combines both the thematic context, derived from the relevant terms in the user's search history, and the social context, based on the tags most frequently used by the user to annotate relevant documents. The user profile is leveraged to propose a set of possible reformulations of the user's initial query, and then the user himself is involved in choosing which variant to send to the search.

Keywords: Information retrieval, Information need, Query, Tags, Term weighting, Relevance, Query reformulation, Personalized Information Retrieval, User profile.

.

Remerciements

Avant toute chose, nous souhaitons exprimer notre gratitude à **Dieu** Tout-Puissant pour nous avoir donné la force, la patience et la persévérance nécessaires pour mener à bien ce mémoire. Sans Sa bénédiction et Sa guidance, ce travail n'aurait pas été possible.

Nous tenons également à remercier chaleureusement notre promoteur de mémoire, **Mr. Rafa Tahar**, pour ses conseils avisés, son soutien constant et ses encouragements tout au long de ce travail. Sa disponibilité et ses précieuses suggestions ont été déterminantes pour la réussite de ce projet.

Nous remercions vivement les membres du jury qui ont eu l'amabilité de porter une appréciation sur ce travail et de participer au jury de soutenance.

Nous sommes profondément reconnaissants envers nos collègues et amis pour leurs encouragements et leur soutien moral. Leur camaraderie a été une source inestimable de motivation et de réconfort.

Enfin, nous souhaitons adresser nos plus sincères remerciements à nos familles, en particulier nos parents, pour leur amour inconditionnel, leur patience et leur soutien indéfectible tout au long de nos études. Leur confiance en nous a été une véritable inspiration.

À tous, un grand merci.

Dédicaces

Rien n'aurait été plus beau à offrir ce modeste travail à mes très chers parents qui sont toujours présents à mes cotés, durant mes succès et mes échecs, que Dieu puisse m'aider à les honorer, les servir et les combler.

J'espère qu'ils trouveront dans ce travail toute ma reconnaissance et ma gratitude.

A mes très chers grands parents.

A mon unique frère **Mohamed**.

A mes très chers sœurs **Faiza**, **Wissam**, **Ikram** et ma petite adorable **Rihab**.

A mes nièces **Taline**, **Aya** et mon neveu **Amir**.

A tous mes camarades de la section pour les moments qu'on a passé ensembles et leurs aides aussi petites qu'elle soit.

A mon binôme **Fathia**.

A tous ceux qui m'on aide de prés ou loin afin d'accomplir ce travail, et a tous ceux dont l'oublie du nom n'est pas celui du cœur.

Fatma

Dédicaces

Après un long chemin d'études plein d'effort et plein de joie. Aujourd'hui j'arrive à la fin de mon parcours.

Merci à mon dieu pour l'opportunité des commencements et l'accomplissement des fins.

Je dédie cette réussite à moi-même en premier lieu, puis à toute personne ayant participé de près ou de loin à cette réussite.

À celle qui m'a porté, protégée, donnée la vie, elle m'a facilité les choses grâce à sa prière jour et nuit, mon modèle de vie, ma chère mère (Naima) qui s'est efforcée avec patience et sacrifices pour ma réussite, que dieu la protège.

À celui dont je porte le nom avec tout fierté qui n'a pas manqué de me motiver avec tous les moyens, à celui qui m'a enveloppé de son amour, de ses soins et de son soutien continuel, mon père (Ahmed).

À ceux qui ont toujours été heureux et très fiers de moi, mon frère et mes sœurs (Saleh, Achouak, Hiba, Amina), sans oublier ma cousine Amira, âme sœur.

À ceux qui ont partagé avec moi des instants de joie et défi, à ceux qui ont été pour moi un soutien mes amis proches (Amina, Ibtissam, Kaouther).

À celui qui m'a accompagné, de son précieux temps, ses conseils et ses orientations de valeur (Abderrahmane).

À mes amis de spécialités ISTW, à mon binôme (Fatma), qui a partagé ce travail avec moi.

Fathia

Tables de matières

Introduction Générale 1
Chapitre 1 : Recherche d'Information Classique 4
1 Introduction 4
2 Notions de base de la recherche d'information 4
3 Processus de recherche d'information 6
3.1 Expression du besoin en information 7
3.2 Processus d'indexation des documents 8
3.2.1 Analyse lexicale (extraction des mots simples) 8
3.2.2 Elimination des mots vides 8
3.2.3 Normalisation 9
3.2.4 Pondération des termes 9
3.3 Reformulation de la requête 11
3.3.1 Reformulation automatique 11
3.3.2 Reformulation manuelle 11
3.4 Appariement document requête 12
4 Les modèles de la Recherche d'information 12
4.1 Le modèle booléen 12
4.2 Le modèle vectoriel 13
4.3 Le modèle probabiliste 14
5 Evaluation des systèmes de recherche d'information 15
5.1 Collection de test 16
5.2 Les mesures d'évaluation 16
5.2.1 Les mesures les plus utilisés 16
5.2.2 Les mesures alternatives 19
6 Limites de la RI classique 20
7 Conclusion 20
Chapitre02 : La Recherche d'Information Personnalisée 22
1 Introduction 22

2 Emergence de la recherche d'information personnalisée 22
3 Modélisation de l'utilisateur 26
3.1 Pertinence Utilisateur 26
3.2 Profil d'utilisateur 27
3.3 Sources des données de profil 28
3.4 Acquisition des données de profil 28
3.4.1 L'acquisition explicite 28
3.4.2 L'acquisition implicite 29
3.5 Techniques de construction de centres d'intérêts 29
3.6 Représentation des données de profil d'utilisateur 30
3.6.1 Représentation basée sur l'historique de recherche 30
3.6.2 Représentation ensembliste 30
3.6.3 Représentation connexionniste 31
3.6.4 Représentation multidimensionnelle 31
3.6.5 Représentation conceptuelle 32
3.7 Evolution de profil utilisateur 33
4 Les dimensions du contexte de recherche 33
4.1 Dimension sociale 33
4.1.1 Les annotations sociales (Tags) 34
4.1.2 Les relations sociales 34
4.2 Dimension localisable 34
4.3 Dimension temporelle 35
4.4 Dimension sémantique 35
5 Intégration de profil utilisateur dans le processus de recherche 36
5.1 Reformulation de requête 37
5.2 Reclassement de résultats 38
6 L'évaluation des systèmes de recherche d'information personnalisés 39
6.1 Les limites de l'évaluation classique 39
6.2 La problématique de l'évaluation personnalisée 41
6.3 Protocole de l'évaluation personnalisée 42
6.4 Les mesures de l'évaluation des systèmes personnalisée 44
7 Conclusion 46
Chapitre 03 : Reformulation Interactive Personnalisée 48
1 Introduction 48
2 Revue sur la reformulation de requête dans les approches de RIP : 49

3 Profil d'utilisateur 57
3.1 Données personnelles 57
3.2 Historique de recherche pertinent 58
3.2.1 Les termes pertinents 58
3.2.2 Les tags pertinents 59
3.3 Base de liens de cooccurrence terme-terme 60
3.4 La base de liens de cooccurrence terme-tag 61
4 Processus de Recherche 62
4.1 La reformulation de requête 63
4.1.1 Proposition des variantes de reformulation 64
4.1.2 Pondération et classement des variantes 65
4.1.3 Intervention de l'utilisateur 66
4.1.4 Mis à jour des poids des variantes 66
5 Evaluation de l'approche 66
5.1 Stratégie d'évaluation 67
5.2 Expérimentation 67
5.2.1 Construction des profils et requêtes : 68
5.2.2 Résultats et discussion 71
5.3 Présentation d'application 83
5.3.1 Modélisation de la base de données 83
5.3.2 Présentation de la base de données 86
5.3.2.1 Présentation de MySQL 86
5.3.2.2 Création des tables de base de données 87
5.3.2.3 Liste des tables de la base de données 87
5.3.3 Environnement de développement de l'application 90
5.3.4 Présentation des interfaces 92
6 Conclusion 98
Bibliographie 111

Table des figures

Figure 1.1:*Processus général de RI.* *7*
Figure 1.2:*Modèle vectoriel.* *13*
Figure 1.3:*Evaluation de Recherche d'Information.* *16*
Figure 1.4:*Courbe de Précision-Rappel.* *18*
Figure 2.1:*De 1ère page et 2ème résultats Google sur "Mars".* *23*
Figure 2.2:*De 3ème page au 4ème résultat Google sur "Mars".* *24*
Figure 2.3:*De 5ème au 9ème résultat Google sur "Mars".* *25*
Figure 2.4:*Triangle sémantique.* *36*
Figure 3.1:*Modèle de données personnelles.* *58*
Figure 3.2:*Modèle de l'historique de recherche pertinent.* *60*
Figure 3.3:*Modèle de la base des relations de cooccurrence de terme-terme.* *61*
Figure 3.4:*Modèle de la base des relations de cooccurrence de terme-tag.* *63*
Figure 3.5:*Processus de Recherche.* *64*
Figure 3.6:*Représentation des résultats de requête « Thérapie »sans reformulation.* *77*
Figure 3.7:*Représentation des résultats de requête « Thérapie » après la reformulation automatique.* *78*
Figure 3.8:*Représentation de résultats de requête « Thérapie » après la reformulation interactive.* . *79*
Figure 3.9:*Représentation des résultats de requête « visiter les monuments » sans reformulation.* ... *80*
Figure 3.10:*Représentation des résultats de requête "visiter les monuments" après la reformulation automatique.* *80*
Figure 3.11: *Représentation des résultats de requête "visiter les monuments" après la reformulation interactive* *81*
Figure 3.12:*Digramme à barres de comparaison des résultats des 3 scénarios.* *82*
Figure 3.13:*Cercle relatif de comparaison de taux d'amélioration des 3 scénarios.* *83*
Figure 3.14:*MCD du profil utilisateur.* *84*
Figure 3.15:*Tables de la base de données.* *85*
Figure 3.16:*Exemple de la table "tags".* *88*
Figure 3.17:*Exemple de la table "sim_coccur_them".* *89*
Figure 3.18:*Exemple de la table "pertinence_terme".* *90*
Figure 3.19:*Exemple de la table "profil".* *91*
Figure 3.20:*Exemple de la table "reformulation_interactive".* *92*
Figure 3.21:*Page d'identification.* *93*
Figure 3.22:*Page de recherche utilisateur.* *96*
Figure 3.23:*Résultats de recherche en mode classique.* *97*
Figure 3.24:*Reformulation automatique de la requête.* *98*
Figure 3.25:*Résultats de recherche après la reformulation automatique.* *99*
Figure 3.26:*Classement des variantes.* *100*
Figure 3.27:*Choix d'une variante.* *101*

Figure 3.28:*Résultats de recherche après la reformulation interactive.* *102*
Figure 3.29:*Poids avant mis à jour.* *103*
Figure 3.30:*Mis à jour des poids.* *104*

Liste des tableaux

Tableau 3.1: *Profils d'utilisateurs.* *71*
Tableau 3.2:*Résultats expérimentaux détaillés des précisions.* *73*
Tableau 3.3:*Cas d'égalité des performances.* *74*
Tableau3. 4:*Quelques exemples de déroulement de reformulation automatique et interactive.* *75*
Tableau 3.5:*La moyenne des précisions des résultats pour toutes les requêtes soumises par les utilisateurs.* *82*
Tableau 3.6:*La moyenne des précisions des résultats pour toutes les requêtes soumises par les utilisateurs.* *82*
Tableau 3.7:*Entités du Modèle Conceptuel de Données (MCD).* *85*
Tableau 3.8:*Relations d'association dans le Modèle Conceptuel de Données (MCD).* *86*

Introduction générale

Introduction Générale :

Introduction Générale

L'objectif sain des systèmes de recherche d'information personnalisée est d'inférer le besoin en information de l'utilisateur caché derrière sa requête, et de lui retourner les résultats les mieux correspondants à ce besoin. La requête initiale de l'utilisateur est souvent insuffisante seule pour permettre la sélection des documents répondant à son besoin en information. De ce fait, des techniques sont proposées pour reformuler cette requête dans l'objectif de bien refléter le besoin en information qu'elle exprime.

La reformulation de la requête de l'utilisateur est un processus permettant de générer une requête plus adéquate à la recherche d'information que celle initialement formulée par l'utilisateur. La reformulation consiste à ajouter de nouveaux termes à une requête initiale ou alors re-pondérer les poids de ses termes dans le but de cibler plus de documents pertinents. La génération de la nouvelle requête s'effectue soit par reformulation automatique en se basant directement sur l'historique de recherche pertinent, ou par reformulation interactive par l'intervention directe de l'utilisateur.

Chacune des deux approches de reformulation automatique ou interactive, a ses inconvénients. La reformulation confiée complément au système soufre du problème de désorientation de la recherche vu l'instabilité et la subjectivité de la pertinence utilisateur. Et la reformulation interactive se confronte souvent avec le problème que l'utilisateur ne réagit pas positivement avec les opérations interactives occupantes telles que répondre à un questionnaire, évaluer un service, donner un avis, …etc.

Dans ce travail, nous visons à combiner les deux principes de reformulation automatique et interactive dans un cadre de personnalisation de la recherche. L'objectif est de combler les lacunes de l'un par les avantages de l'autre.

La contribution de notre approche de reformulation interactive personnalisée de la requête utilisateur se résume à :

- La conception et la construction d'un profil d'utilisateur qui sert à identifier, organiser et sauvegarder les intérêts et les préférences de l'utilisateur. Principalement, ce profil est composé des termes issus des documents jugés pertinents par l'utilisateur, ainsi que les tags utilisés pour annoter ces documents pertinents. Les termes et les tags sont pondérés et des scores de similarités basés sur les cooccurrences sont définis entre eux.

- Les informations fournies par le profil utilisateur sont utilisées pour proposer plusieurs reformulations possibles à la requête initiale de l'utilisateur, tout en assurant que les éléments ajoutés à la requête, dans chaque variante de reformulation, sont à la fois pertinents pour l'utilisateur et en relation avec le sujet de la requête. Le classement des variantes de reformulation est élaboré en fonction d'un score dynamique qui mesure le poids de chacune.
- L'implication de l'utilisateur en lui donnant la possibilité de participer à la reformulation de la requête en choisissant l'une des variantes proposées. L'objectif là, est de donner la décision finale à l'utilisateur pour ne pas négliger l'aspect subjectif de la pertinence utilisateur. L'intervention de l'utilisateur n'est pas vraiment occupante de fait qu'il va juste cliquersur une variante de reformulation. Un *Timer* est programmé pour lancer en recherche la variante classée en premier en cas de non réaction de l'utilisateur.
- Le choix d'une variante par l'utilisateur, et ces réactions (jugements de pertinence) avec les 5 tops documents retournés par le système, permettent de mettre à jours les poids, et par conséquent, le classement des variantes de reformulations dans les futures recherches.

Ce mémoire se compose de trois chapitres. Ce qui suit est une brève présentation de leur contenu respectif :

Le chapitre 1 introduit les notions de base de la recherche d'information, puis le processus de recherche d'information et les principaux modèles de systèmes d'information. Ensuite, nous abordons la définition et l'évaluation des systèmes d'information. Enfin, nous concluons par les limites de la recherche d'information.

Le chapitre 2 est consacré à la recherche d'information personnalisée. Nous y mettons en lumière l'émergence de la recherche d'information personnalisée, la modélisation de l'utilisateur et sa pertinence, ainsi que la conception et la construction du profil utilisateur jusqu'à son intégration dans le processus d'accès à l'information. Enfin, nous terminons par l'évaluation des systèmes de recherche personnalisés.

Le dernier chapitre présente notre contribution et est divisé en deux parties. La première partie est consacrée à la conception du profil utilisateur et aux résultats expérimentaux obtenus pour évaluer notre approche de personnalisation, selon une

démarche d'évaluations par simulation. La deuxième partie décrit l'application que nous avons développée à cet effet.

Chapitre 1

La Recherche d'Information Classique

Chapitre 1 : Recherche d'Information Classique

1 Introduction

La recherche d'information (RI) désigne la branche de l'informatique qui porte sur l'acquisition, l'organisation, le stockage et la recherche de l'information (Salton, 1968). Ce domaine s'est développé en réponse aux besoins des utilisateurs exprimés à travers des requêtes, marquant ses débuts avec l'avènement des premiers ordinateurs dans les années 1940 et la gestion émergente des bibliothèques numériques.

Les décennies suivantes ont été témoins des premières expérimentations sur des collections tests, jetant ainsi les bases d'une évaluation systématique des performances. Les années 1960-1970 ont vu l'émergence des premiers modèles de RI, notamment le modèle vectoriel. L'avènement d'Internet dans les années 1990 a constitué un tournant majeur, confrontant la RI à des volumes de données sans précédent(Rafa, 2023).

L'objectif fondamental d'un système de RI demeure la fourniture d'informations pertinentes en réponse à la requête de l'utilisateur, tout en minimisant le bruit (documents non pertinents retrouvés) et le silence (documents pertinents non retrouvés). Cependant, la pertinence demeure une notion subjective, variante selon l'utilisateur et le système.

Dans ce chapitre, nous explorerons en détail les concepts clés de la RI tels que le besoin en information, la requête, le document, l'appariement, etc. Nous examinerons également les principales étapes du processus de RI ainsi que les modèles et techniques permettant la représentation, évaluation des systèmes de recherche d'informations. En outre, nous aborderons les limitations de la RI classique face à l'énormité des données actuelles et à la diversité des contextes des utilisateurs, tout en explorant comment la RI adaptative et sémantique cherche à relever ces défis.

2 Notions de base de la recherche d'information

Dans le domaine de la recherche d'information, divers concepts fondamentaux sont traités et explorés.

a) Document : Le document est l'unité d'information de base dans un système de recherche d'information (SRI). Il constitue l'élément atomique pouvant être indexé, stocké, retrouvé et restitué à l'utilisateur. Un document peut prendre diverses formes : texte, image, vidéo, enregistrement sonore, etc. Il possède généralement un contenu

sémantique et des métadonnées décrivant son origine, sa structure, son format, etc. contenu informationnel du document est extrait et représenté lors de la phase d'indexation du SRI, via des descripteurs sémantiques ou des signatures physiques. Le document devient alors interrogeable par le système.

b) Collection de documents : La collection de documents, aussi appelée corpus ou fonds documentaire, désigne l'ensemble des documents disponibles et accessibles dans le SRI. Elle constitue la source d'information dans laquelle la recherche de documents pertinents est effectuée. Plus la collection est vaste en taille (nombre de documents) et hétérogène en contenu (formats, langues, disciplines), plus le défi de la recherche d'information est complexe. Les enjeux sont alors l'exhaustivité de l'indexation et l'efficacité de la recherche.

c) Besoin en information : Le besoin en information correspond à l'objectif de l'utilisateur, ce qu'il cherche à obtenir comme information en interrogeant le SRI. On distingue classiquement :

- Le besoin vérificatif : l'utilisateur cherche à vérifier ou retrouver une information précise déjà connue.

- Le besoin thématique connu : l'utilisateur souhaite à obtenir ses connaissances dans un domaine est déjà connu.

- Le besoin thématique inconnu : l'utilisateur explore un sujet nouveau, inconnu.

Le besoin évolue au cours du processus de recherche : de vague à précis, de général a ciblé. L'utilisateur affine sa compréhension de l'information recherchée.

d) Requête : La requête constitue la traduction formelle du besoin d'information de l'utilisateur sous une forme compréhensible par le SRI, souvent exprimée à l'aide de mots-clés ou de descripteurs. Elle représente l'interface entre l'utilisateur et le système. Différents types de requêtes existent : mots-clés, phrases en langage naturel, expression booléenne, graphe sémantique, etc. L'enjeu est l'adéquation entre le besoin réel et sa formulation, qui conditionne la pertinence des résultats. La requête initiale est souvent imparfaite et nécessite des reformulations.

e) Pertinence : La pertinence évalue le degré d'adéquation entre un document retrouvé et le besoin en information exprimé dans la requête. On distingue :

- La pertinence système : score objectif mesuré automatiquement par le SRI via la fonction d'appariement RSV (Q/D).

- La pertinence utilisateur : jugement subjectif de l'utilisateur sur la pertinence d'un document après son examen. La pertinence utilisateur est l'évaluation finale et déterminante, mais difficile à anticiper automatiquement.

f) Système de Recherche d'Information : Le SRI permet de retrouver et restituer à l'utilisateur les documents les plus pertinents vis-à-vis du besoin exprimé dans la requête. Il repose sur la comparaison entre la représentation des documents (index) et la requête via une fonction d'appariement, puis le classement des documents par pertinence système. L'enjeu est l'efficacité de la recherche (rappel/précision) et la satisfaction utilisateur.

3 Processus de recherche d'information

L'objectif essentiel d'un processus de recherche d'information (RI) consiste à identifier les documents qui correspondent le mieux aux besoins informationnels de l'utilisateur tels qu'exprimés dans une requête. Pour atteindre cet objectif, le système de recherche d'information (SRI) opère sur deux fronts : d'une part, il manipule un ensemble de documents qu'il transforme en une collection de documents indexés grâce à un processus d'indexation, et d'autre part, il gère les besoins informationnels de l'utilisateur exprimés sous forme de requête, qui est elle-même indexée. Ensuite, le système effectue un appariement entre ces deux représentations (appariement document-requête), aboutissant à la production d'une liste de documents estimés comme pertinents, constituant ce que l'on appelle la pertinence du système. La (Figure 1.1) offre une illustration des deux phases principales du processus de la recherche d'information.

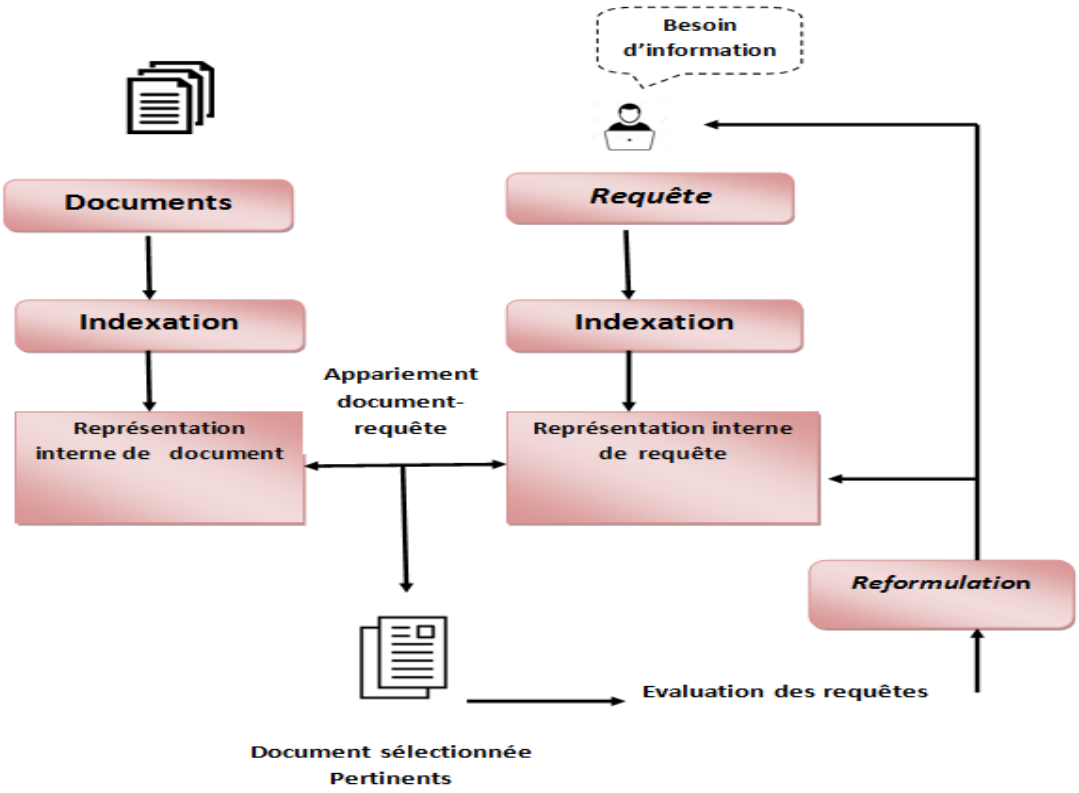


Figure 1.1:*Processus général de RI.*

3.1 Expression du besoin en information

Le besoin en information de l'utilisateur est l'expression mentale de ce qu'il recherche. Ce besoin est IO représenté à travers une requête, qui sera traitée par le SRI. C'est la première phase dans le processus de recherche d'information. Cette phase est liée au facteur cognitif humain selon lequel l'utilisateur définit son besoin en information pour combler une lacune dans ses connaissances (Baziz, 2005).

L'utilisateur est le déclencheur de l'opération de recherche d'information, et le validateur du résultat de cette recherche. Mieux comprendre et satisfaire les besoins cognitifs de l'utilisateur permettrait d'améliorer les performances d'un SRI. Toutefois, il ne faut pas perdre de vue que l'utilisateur est plus concerné par retrouver l'information sur le sujet qui l'intéresse plutôt que par retrouver des données qui satisfont une requête donnée, le plus souvent mal formulée(El ghali, 2016).

3.2 Processus d'indexation des documents

L'indexation vise à construire une représentation simplifiée décrivant le contenu informationnel des documents en vue de faciliter la recherche. Elle consiste à détecter les termes les plus représentatifs du contenu du document qui serviront de clés pour retrouver ce document au sein d'une collection (Zargayouna, 2005). L'indexation recouvre un ensemble de techniques visant à transformer les documents (ou les requêtes) en descripteurs (mots clés) capables de représenter leur contenu (Daoud, 2009).

La recherche d'information propose diverses méthodes d'indexation :

– Indexation manuelle : Un spécialiste du domaine sélectionne les termes qu'il estime pertinents pour décrire le contenu sémantique du document. Cette méthode garantit un vocabulaire d'indexation sous contrôle, renforçant ainsi la cohérence et la qualité de la représentation obtenue. Cependant, cette approche demeure subjective et impraticable pour une collection volumineuse.

– Indexation automatique : Elle s'appuie sur des algorithmes pour attribuer automatiquement des descripteurs à différentes parties du document, sans nécessiter l'intervention d'un expert. Dans le cas des documents textuels, chaque mot peut être considéré comme un potentiel index du document qui le contient.

– Indexation semi-automatique : Cette méthode résulte de la fusion des deux approches précédentes ; un processus initial automatisé extrait les termes du document, tandis que le choix final des descripteurs est confié au spécialiste du domaine. Ce dernier utilise un vocabulaire contrôlé, tel qu'un thésaurus ou une base terminologique, pour effectuer cette sélection.

3.2.1 Analyse lexicale (extraction des mots simples)

L'analyse lexicale constitue la phase visant à convertir le texte d'un document en un ensemble de termes, également connus sous le nom de "token" en anglais. Cet analyseur lexical détecte et élimine les espaces entre les mots, les chiffres, la ponctuation, et d'autres éléments, simplifiant ainsi la représentation du contenu.

3.2.2 Elimination des mots vides

Cette opération n'est pas toujours obligatoire. La liste des termes extrait précédemment contient des mots qui ont un rôle syntaxique, non signifiant appelés 'mots vides ' (appelés aussi : Stop-List ou : anti-lexique) (Fox, 1992), tel que : les pronoms

personnels, les prépositions, Conjonctions, etc. l'élimination des mots vides peut être réalisée à l'aide d'un anti-dictionnaire dédié ou en excluant les termes qui apparaissent fréquemment dans la collection. Cette approche simplifie la représentation du texte en se concentrant sur les termes plus informatifs, améliorant ainsi l'efficacité du traitement des données textuelles.

3.2.3 Normalisation

La normalisation en traitement du langage naturel, et plus spécifiquement en recherche d'information, fait référence à un processus visant à rendre les termes ou les expressions linguistiques plus uniformes et comparables. Elle sert à éliminer les variations linguistiques qui peuvent exister entre différentes formes d'un même mot, telles que les pluriels, les conjugaisons verbales, les synonymes, etc. Dans le contexte de la normalisation, on utilise souvent deux techniques principales :

a)Lemmatisation : Est un processus qui consiste à représenter les différentes formes d'un mot appartenant à la même catégorie grammaticale par leur forme canonique, également appelée lemme. Par exemple, dans le cas des verbes tels que "a, aurai, eurent,...", la lemmatisation les ramène à leur forme infinitive "avoir ".

b) Racinisation : Est un processus visant à représenter un ensemble de mots par leur racine commune, également appelée "stem", en éliminant les divers suffixes qui les accompagnent. Par exemple, les termes "joueur, joueuse, jouabilité" seraient tous réduits à la racine "jou".

3.2.4 Pondération des termes

La pondération permet d'évaluer l'importance des termes dans les textes. Cette importance est exprimée par des poids affectés aux termes (Baziz, 2005) ;(Daoud, 2009). Les poids des termes sont calculés par des formules de pondération basées généralement sur les fréquences des termes dans le texte.

Certaines méthodes de pondération reposent sur la fréquence simple d'un terme TF (Terme Frequency). Cette mesure est directement liée à la fréquence du terme dans le document. L'idée fondamentale est que la fréquence d'un terme dans un document est proportionnelle à son importance dans la description de ce document. Le TF est généralement formulé de l'une des manières suivantes :

A. Tf : utilisation brute

B.$0.5 + 0.5\ \frac{TF}{MAX(TF)}$

Où : MAX(TF) représente le nombre maximum d'occurrences d'un terme dans un document.

D'autres méthodes de pondération reposent sur la fréquence inverse de document IDF (Inverse Document Frequency). L'IDF (Inverse de Fréquence des Documents) évalue la pertinence d'un terme dans l'ensemble d'une collection. Le concept fondamental est que les termes présents dans un nombre limité de documents de la collection sont plus révélateurs du contenu de ces documents que ceux qui apparaissent dans tous les documents de la collection. Cette mesure peut être formulée de différentes manières, notamment :

A.$IDF = \log\ (N/df)$

B.$IDF = \log\ (N/(1 + df))$

Où df est le nombre de documents contenant le terme t et N est le nombre de documents dans toute la collection.

On peut explorer d'autres formulations permettant de fusionner les deux mesures (TF-IDF) :

$$TF * IDF = \log(1 + TF) * \log\ (N/df)$$

La fonction de pondération de la forme Tf.Idf consiste à multiplier les deux mesures Tf et Idf. Une formule largement utilisée est la suivante (Zemirli, 2004) :

$$TF.IDF = \left(0.5 + 0.5\ \frac{TF}{MAX(TF)}\right) * \log\ (\frac{N}{df})$$

Une normalisation de la mesure du Tf.Idf par rapport à la longueur des documents a été proposée comme suit (Daoud, 2009) :

$$TF.IDF = \frac{TF + \log\ (\frac{N-df+0.5}{df+0.5})}{2 * (0.25 + 0.75 * \frac{\mathrm{dl}}{\Delta d})}$$

Idf est la longueur du document en nombre de termes et Δd la longueur moyenne des documents de la collection.

3.3 Reformulation de la requête

Parfois l'utilisateur est confronté à une situation difficile, il est incapable de trouver les mots précis pour exprimer son besoin en information. Alors, certains pour ne pas dire la majorité des documents qui lui sont retournés l'intéressent moins que d'autres. Outre cela, certaines requêtes sont courtes, du coup, elles ne sont pas sémantiquement riches, pour que le SRI puisse retourner des documents pertinents. Afin de pallier ces problèmes, les chercheurs en RI se sont orientés vers l'intégration d'une étape supplémentaire dans le processus de recherche : la reformulation ou expansion de requêtes. Elle consiste à modifier la requête initiale de l'utilisateur par l'ajout de termes significatifs et/ou la ré-estimation de leur poids. Si les termes rajoutés proviennent des documents de la collection, on parle de réinjection de pertinence (relevance feedback). En revanche, s'ils sont issus d'une ressource conceptuelle externe (ontologie, thésaurus ou dictionnaire), on parle dans ce cas de reformulation de requêtes (Saidl'haj, 2007). On peut distinguer deux types de reformulation :

3.3.1 Reformulation automatique

Egalement dite pseudo réinjection de pertinence. Dans ce cas, l'utilisateur n'intervient pas, l'extension de la requête est effectuée en exploitant les premiers documents retournés par le SRI en réponse à la requête initiale (blind feedback)(Salton, 1968) (Boughanem & Chrisment, 1999)ouà partir d'une ressource externe qui peut être un thesaurus, une ontologie, etc. (Khan, 2000),(Andreasen, 2003).

3.3.2 Reformulation manuelle

Dans ces approches, l'utilisateur est amené à choisir les documents qu'il juge pertinents parmi ceux obtenus à partir de sa requête initiale. Ce jugement de pertinence est ensuite utilisé pour ajuster la requête initiale en modifiant l'importance attribuée aux termes qu'elle contient et/ou en introduisant de nouveaux termes considérés comme utiles pour trouver des documents pertinents. On la nomme communément réinjection de la pertinence ou « relevance feedback ». En résumé, ces méthodes reposent sur l'interaction de l'utilisateur pour affiner la recherche en tenant compte de ses préférences et de ses jugements de pertinence (Dounas & Ould Fella, 2015).

La technique de réinjection de pertinence a été mise en place à l'origine dans le modèle vectoriel (Rocchio, 1971).

3.4 Appariement document requête

Le processus d'appariement implique la comparaison entre la représentation de la requête et celles des documents. Il évalue, pour chaque paire requête-document, une mesure appelée pertinence système, qui indique le degré de similarité entre la requête et le document en question. Ce processus repose sur l'utilisation d'une fonction de similarité, notée RSV (RetrievalStatus Value). Cette fonction varie d'un modèle de recherche d'information à un autre, et d'ailleurs, un modèle de RI est distingué par sa fonction de similarité ainsi que son modèle d'indexation. (Rafa, 2023)

4 Les modèles de la Recherche d'information

Les modèles RI sont des outils consiste à trouver des documents pertinents parmi une collection des documents, en fonction des besoins d'information d'un utilisateur. Il existe plusieurs modèles on présentant les trois catégories principalesde modèles : modèle booléen, modèle vectoriel et le modèle probabilistes.

Un modèle de RI est défini par un quadruplet (D, Q, F, R (q, d)) :

- D : représente l'ensemble de documents.
- Q : représente l'ensemble de requêtes.
- F : représente le schéma du modèle théorique de représentation des documents et des requêtes. (décrit comment les documents et les requêtes sont représentés dans le modèle, c'est comme un schéma qui dit comment les informations sont structurées).
- R (q, d) est la fonction de pertinence du document d à la requête q (Yates & Neto, 1999).

Il existe de nombreux modèles de recherche d'information (RI), parmi lesquels nous mettons en avant les plus couramment utilisés : le modèle booléen, le modèle vectoriel et le modèle probabiliste.

4.1 Le modèle booléen

Le modèle booléen, introduit par (Salton G. , 1971), est le premier modèle de recherche d'information (RI). Il repose sur la théorie des ensembles, où un document est représenté par un ensemble de termes, tandis qu'une requête est formulée comme un ensemble de mots associés à des opérateurs booléens : AND (∧), OR (∨) et NOT (¬). La correspondance exacte s'effectue en vérifiant la présence ou l'absence des termes de la requête dans les documents.

Appariement (q, d) = RSV (q, d)=1 ou 0

4.2 Le modèle vectoriel

Proposé par Salton dans le système SMART (Salton G. , 1971), ce modèle repose sur des principes algébriques, en particulier le calcul vectoriel. Les documents et la requête sont représentés comme des vecteurs dans l'espace vectoriel (Figure 1.2).

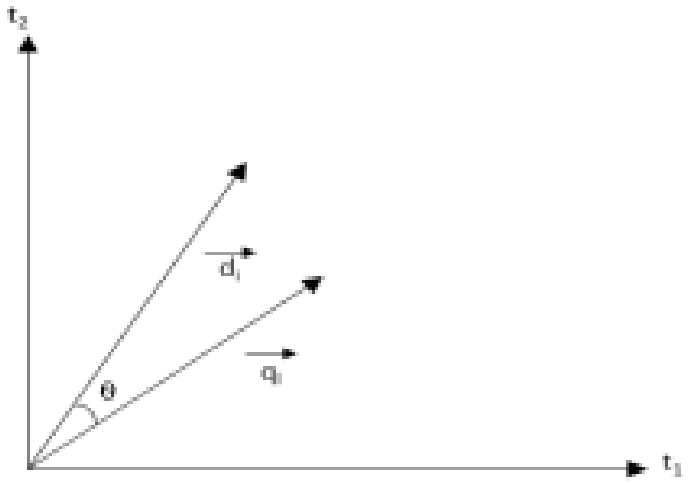


Figure 1.2:*Modèle vectoriel.*

Il inclut le modèle vectoriel, le modèle vectoriel généralisé, LSI (Latent Semantic Indexing) et le modèle connexionniste. La similarité entre chaque document 'Dj' et la requête 'Q' est déterminée à l'aide de l'une des fonctions suivantes :

- La fonction TF.IDF :

$$W_{i,j} = TF_{i,j} * IDF_i$$

Ils existent d'autres méthodes de pondération qui combinent les deux mesures TF et IDF(Allan & Salton, 1994)(McGill & Salton, 1983)(Baeza-Yates & Ribeiro-Neto, 1999).

Par exemple :

$$T F * IDF = \log (1 + T F) * \log (N/ df_t)$$

- **Fonctions de similarités :**

La pertinence du document corresponde au degré de similarité entre le vecteur de la requête et celui du document

- Le produit scalaire :

$$RSV(q_i, d_j) = \sum_{i=1}^{N} q_i * d_{ij}$$

- Mesure de cosinus :

$$RSV(q_i, d_j) = \frac{\sum_{i=1}^{N} q_i * d_{ij}}{\sqrt{\sum_{i=1}^{N} q_i^2} * \sqrt{\sum_{i=1}^{N} d_{ij}^2}}$$

- Mesure de Jaccard :

$$RSV(q_i, d_j) = \frac{\sum_{i=1}^{N} q_i . d_{ij}}{\sum_{i=1}^{N} q_i^2 + \sum_{i=1}^{N} d_{ij}^2 - \sum_{i=1}^{N} q_i . d_{ij}}$$

Où :

d_i: Poids du terme t_i dans le document D_j,

q_i: Poids du terme t_i dans la requête Q (Zemirli W. , 2004).

4.3 Le modèle probabiliste

Le modèle probabiliste a été développé dans les années 70, et a connu des développements récents, car les approches basées sur ce modèle ont obtenu de très bons résultats dans TREC. Il se base sur un calcul de similarité probabiliste (Robertson & Sparck Jones, 1976) ;(McGill & Salton, 1983)(Maron & Kuhns, 1960).

La fonction d'appariement entre un document d_C et une requête Q repose sur l'évaluation de la probabilité de pertinence entre eux. Pour déterminer cette probabilité, il est nécessaire de calculer d'abord deux probabilités distinctes :

- P (R,d_j): La probabilité que le document d_C soit pertinent pour la requête Q.

- P ($\bar{R}|d_j$) : La probabilité que le document d_C ne soit pas pertinent pour la requête Q.

La similarité entre le document d_C et la requête Q, est calculée par la formule suivante :

$$RSV(q,d) = \frac{P(R/q)}{p(\bar{R}/d)}$$

Où : R est l'évènement de pertinence, $\bar{R}$ est l'évènement de non pertinence.

5 Evaluation des systèmes de recherche d'information

L'évaluation du SRI permet de vérifier l'efficacité de retrouver des documents pertinents (Moulahi, 2015), et d'évaluer la performance du système. Le but d'un système de recherche d'information est de satisfaite le besoin recherché par l'utilisateur et retrouvée les documents pertinents et rejeter les documents non pertinents. On a deux types de pertinence (Bouhini, 2014) **:**

- **La pertinence utilisateur :** c'est l'utilisateur lui-même qui détermine la pertinence d'un document en fonction de ses besoins d'information.
- **La pertinence système :** le SRI évaluée la pertinence d'un document pour une requête s'embase sur la fonction de pertinence RSV (q, d).

La pertinence utilisateur doit être la plus proche possible de la pertinence système.

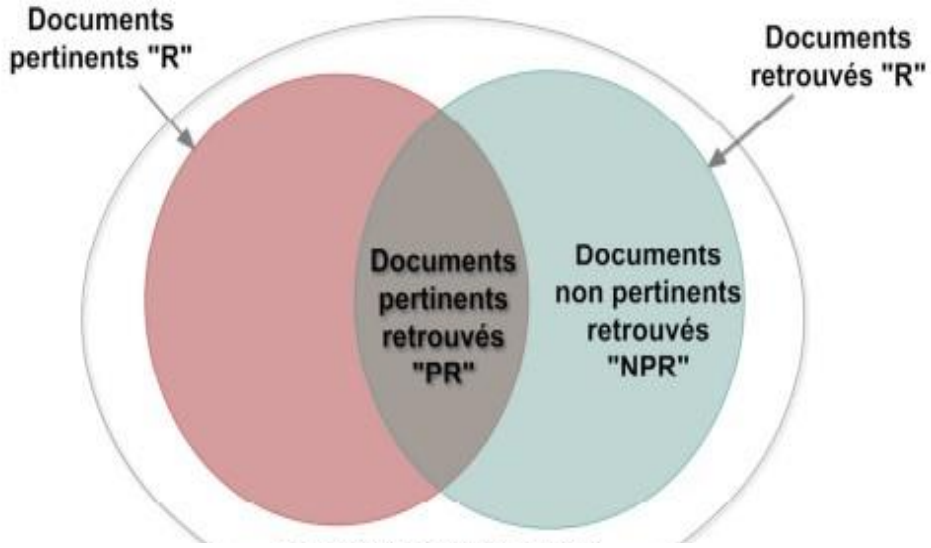


Figure 1.3:*Evaluation de recherche d'information.*

5.1 Collection de test

Une collection de test comprend généralement des corpus de documents, une liste de requêtes prédéfinies et des jugements de pertinence.

- Un corpus de documents : c'est un ensemble de documents à indexer, et utiliser pour évaluer le système.
- Collection des requêtes : Une liste de requêtes prédéfinies.
- Les jugements de pertinence : ils sont établis de manière manuelle (par l'utilisateur lui-même), ils contiennent la liste des documents pertinents pour chaque requête (Saoud, 2017).

Pour juger l'efficacité et la performance d'un système de RI, il faut comparer les résultats retournés par ce dernier avec les jugements de pertinence dans la collection pour chaque requête (Saoud, 2017). Il existe plusieurs collections de test dans le domaine de recherche d'information : la collection CACM, la collection CISI, la collection CLEF et la collection TREC.

5.2 Les mesures d'évaluation

L'objectif des SRI est de trouvée les documents les plus pertinents à une requête donnée, pour évaluer la performance d'un système de recherche d'information on utilise les mesures d'évaluation les plus utilisé : la précision, le rappelle et les mesures alternatives.

5.2.1 Les mesures les plus utilisés

Parmi les mesures proposées dans la littérature de la recherche d'information (RI) pour évaluer l'efficacité d'un système de recherche d'information (SRI), la précision et le rappel sont les plus couramment utilisées. Ci-dessous, nous présentons quelques-unes des mesures d'évaluation les plus courantes :

a) La précision :

Représente la capacité du système RI de retrouver que les documents pertinents et rejeter tous les documents non pertinents pour une requête donnée. On calcule la précision avec la formule suivante :

$$\text{Précision (P)} = \frac{Nombre\ des\ documents\ pertinents\ retrouvé}{Nombre\ des\ documents\ retrouvé}$$

$0 < Précision < 1$

P=1si tous les documents retrouvés sont pertinents.

P=0 si aucun document retrouvé n'est pertinent.

b) Le Rappel :

Représente la capacité du système RI de retrouver les documents pertinents de la collection pour une requête. Il définit comme suite :

$$Rappel\ (R) = \frac{Nombre\ des\ documents\ pertinents\ retrouvés}{Nombre\ totale\ des\ documents\ pertinent}$$

$0 < Rappel < 1$

R=1 alors le système retrouver tous les documents pertinents.

R=0 alors aucun document pertinent n'a été retrouvé.

Voici un exemple courbe de Précision-Rappel dans la Figure 1.4 :

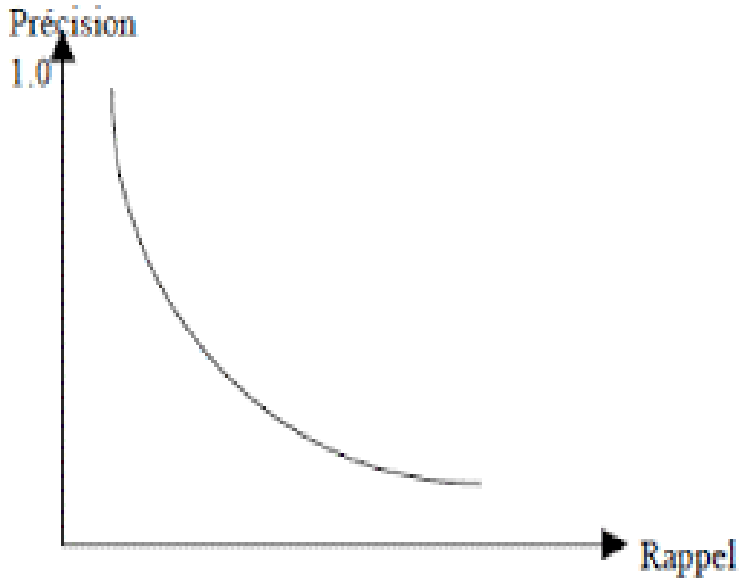


Figure 1.4*: Courbe de Précision-Rappel.*

Il existe deux mesures complémentaires au Rappel et précision :

- *Bruit :* c'est une mesure complémentaire à la Précision.
- *Silence :* c'est une mesure complémentaire du Rappel.

c) F-mesure :

Une mesure qui combine la Précision et le Rappel nommée F-score ou F-mesure défini par :

Si $\alpha = 1$

$$F = \frac{2 * Précision * Rappel}{Précision + Rappel}$$

Si $\alpha > 1$

$$F = \frac{(1 + \alpha^2) * Précision * Rappel}{Précision * \alpha^2 + Rappel}$$

Si $\alpha < 1$

$$F = \frac{(1 + \alpha^2) * Précision * Rappel}{\alpha^2 . Précision + Rappel}$$

d) La moyenne des Précisions moyennes MAP :

Est une mesure de performance globale du système. Il représente la moyenne des précisions moyennes pour l'ensemble des requêtes. Il est calculé en deux étapes :

- Calcule la précision moyenne (Average précision) pour une requête donnée AP_q :

- Calculer pour chaque document pertinent sa précisionP_d :

$$P_d = \frac{Nombre\ document\ pertinent\ retrouvé\ aurang}{rang\ du\ document\ pertinent\ retrouvé}$$

- Calculer la moyenne des précisions des documents pertinents :

$$AP_q = \frac{\sum_{i=1}^{N} P_d}{N}$$

D'où :

- N c'est le nombre total de document pertinent pour la requête q.

- Calcule la moyenne des précisions moyenne de chaque requête AP_q :

$$MAP = \frac{\sum_{j=1}^{M} AP_{qj}}{M}$$

D'où :

- **M :** c'est le nombre total de requêtes et $\mathbf{AP_{qj}}$: la précision moyenne pour la requête j.

5.2.2 Les mesures alternatives

Est une métrique différente des mesures classiques, conçue pour fournir une perspective complémentaire ou pour mieux s'adapter à des contextes spécifiques, en tenant compte de divers aspects de la performance du système.

a) Précision @ X : L'efficacité d'un système de recherche d'information est évaluée aussi par la précision au rang X. Cette précision mesurer la proportion des documents pertinents sélectionnés parmi les X premiers documents retournée par le système. La formule de calcul est comme suite :

$$Précision\ @\ X \quad = \frac{Nombre\ de\ documents\ pertinents\ parmi\ les\ X\ premiers\ docs\ retourner}{X}$$

b) R-Précision : La précision exact ou R-précision, cette précision mesure la proportion des documents pertinents retrouves après que R documents ont été retrouvé, où R est le nombre de documents pertinents pour la requête considérée. Par exemple, 2-précision est la précision juste après trouver le deuxième document pertinent.

c) La moyenne des réciproques des rangs MRR : La mesure MRR permet d'évaluer le nombre de documents qu'il faut considérer avant de retrouver le premier document pertinent. Elle est égale à la moyenne du rang du premier document pertinent(Rafa, 2023).La formule est :

$$MRR = \frac{1}{|q|}\sum_{i=1}^{|q|}\frac{1}{rank\ (i)}$$

Où *Rank(i) :* est le rang du premier document pertinent pour la i^{eme} requête.

6 Limites de la RI classique

La finalité fondamentale des systèmes de recherche d'information consiste toujours à repérer, au sein d'un ensemble étendu d'informations diverses, celles qui sont pertinentes et répondent aux besoins spécifiques de l'utilisateur en matière d'information.

La limite majeure de la RI classique est qu'elle est basée sur une approche généraliste qui évalue invariablement les requêtes des utilisateurs et délivrent des résultats sans tenir compte des besoins en information de l'utilisateur qui a émis la requête ou du contexte de recherche (Daoud, 2009).

Les nouveaux défis auxquels doit faire face la recherche d'information traditionnelle, tels que l'expansion rapide du volume d'informations disponibles, la diversité et l'évolution constante des besoins en information des utilisateurs, la complexité des requêtes exprimées par ces derniers, ainsi que la subjectivité associée à la pertinence utilisateur, ont engendré l'émergence de la recherche d'information personnalisée. Cette approche novatrice vise à prendre en considération de manière proactive l'utilisateur et ses intérêts tout au long du processus de recherche, avec pour objectif d'améliorer significativement la pertinence des résultats obtenus(Rafa, 2023).

7 Conclusion

Dans ce chapitre, nous avons présenté les concepts et les notions fondamentaux de la recherche d'information, à savoir le document et la collection de document, la requête, les mots-clés et l'analyse lexicale, aussi la pertinence et le besoin d'information. Nous avons abordé les processus de la RI tel que la correspondance document-requête et la reformulation des requêtes. Puis nous avons étudié les différents modèles de recherche et les mesures d'évaluation de système.

En raison des limites de la recherche d'information classique, les efforts ont été dirigés vers la recherche d'information personnalisée, qui vise à répondre aux besoins d'information des utilisateurs en améliorant la pertinence de ces résultats de recherche.

Chapitre 02

La Recherche d'Information Personnalisée

Chapitre02 : La Recherche d'Information Personnalisée

1 Introduction

Face aux limites de plus en plus évidentes des systèmes de recherche d'information (SRI) classiques pour répondre aux attentes ou besoins des utilisateurs, les approches en matière de recherche d'information évoluent vers une nouvelle génération de SRI centrée sur la personnalisation de l'accès à l'information. L'objectif principal de ces nouvelles approches est d'améliorer la satisfaction des besoins en information des utilisateurs en exploitant le contexte de l'utilisateur ainsi que les connaissances associées à la requête. Le composant utilisateur est décrit à travers son profil, englobant ses intérêts, ses compétences et ses objectifs de recherche, ce qui constitue l'élément primordial permettant de fournir des informations pertinentes qui répondent pleinement aux attentes de l'utilisateur.

Dans ce chapitre, nous mettons en lumière l'émergence de la recherche d'information personnalisée, la modélisation de l'utilisateur ainsi que son importance, la création et l'intégration du profil utilisateur dans le processus de recherche d'information, et nous concluons par l'évaluation des systèmes de recherche personnalisée.

2 Emergence de la recherche d'information personnalisée

Ces dernières années, il est devenu évident que les Systèmes de Recherche d'Information (SRI) doivent évoluer pour mieux répondre aux besoins individuels des utilisateurs. Cette évolution a donné naissance à un nouveau paradigme de recherche : la Recherche d'Information personnalisée (RIP). Cette approche révolutionnaire reconnaît que chaque utilisateur a des besoins uniques et des préférences spécifiques lorsqu'il recherche de l'information en ligne. En conséquence, la RIP vise à adapter les résultats de recherche en fonction des intérêts et des exigences de chaque individu. En plaçant l'utilisateur au centre du processus de recherche, la RIP offre une expérience plus pertinente et plus satisfaisante, ouvrant ainsi la voie à une exploration plus efficace et plus enrichissante de l'information disponible.

Pour exemplifier la difficulté liée à la personnalisation, prenons en compte le scénario suivant

Si on formule sur le moteur de recherche *Google* la requête « Mars », on remarque, comme indiqué sur les figures (N° 2.1, 2.2 et 2.3), que les 10 premiers résultats sont comme suit :

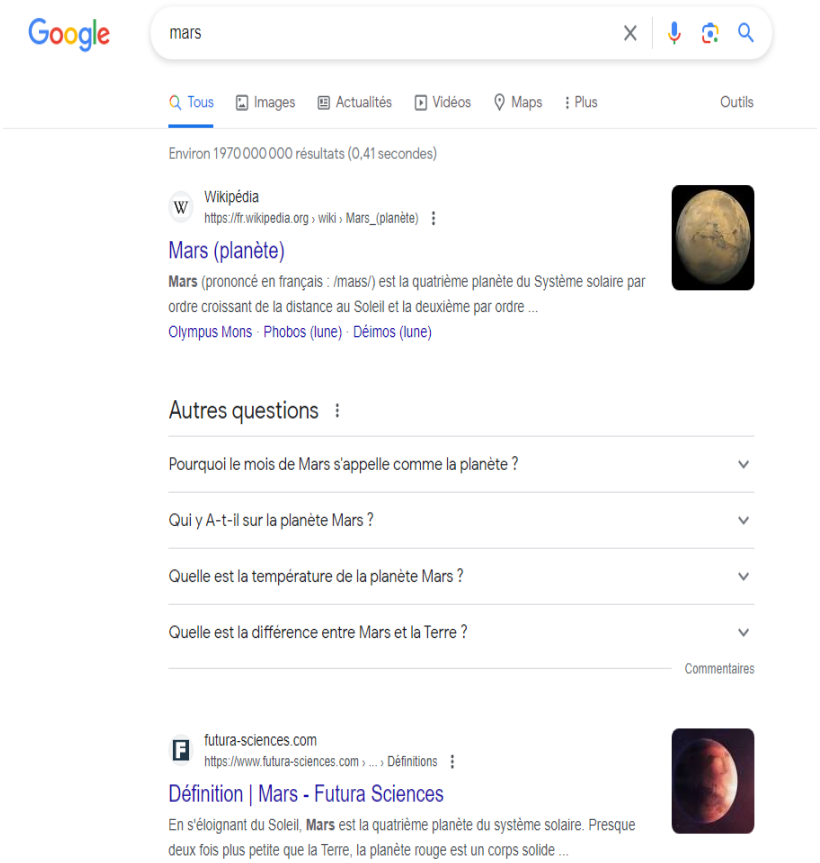


Figure 2.1:*De 1ère page et 2ème résultats Google sur "Mars".*

Le premier et deuxième résultat retournée des articles, des images et des informations sur la planète Mars, y compris des données scientifiques, des actualités sur les missions d'exploration martienne, et des descriptions de sa surface et de son atmosphère. Ainsi que le troisième et quatrième résultat. (Figure 2.2)

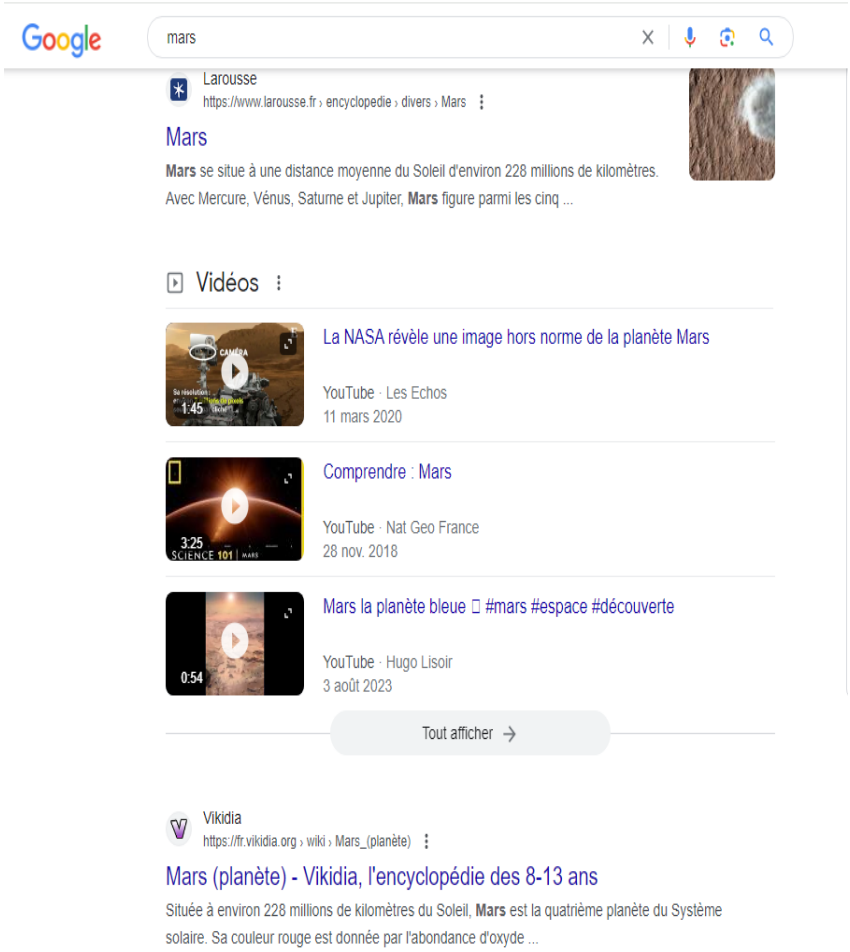


Figure 2.2:*De 3ème page au 4ème résultat Google sur "Mars".*

Le cinquième résultat suivant (Figure 2.3) concerne le mois de Mars, comprenant des informations sur son calendrier, ses événements historiques notables survenus en ce mois, des célébrations ou festivals spécifiques associés à cette période, ou même des prévisions météorologiques pour Mars, le mois. Et les autres résultats retournés.

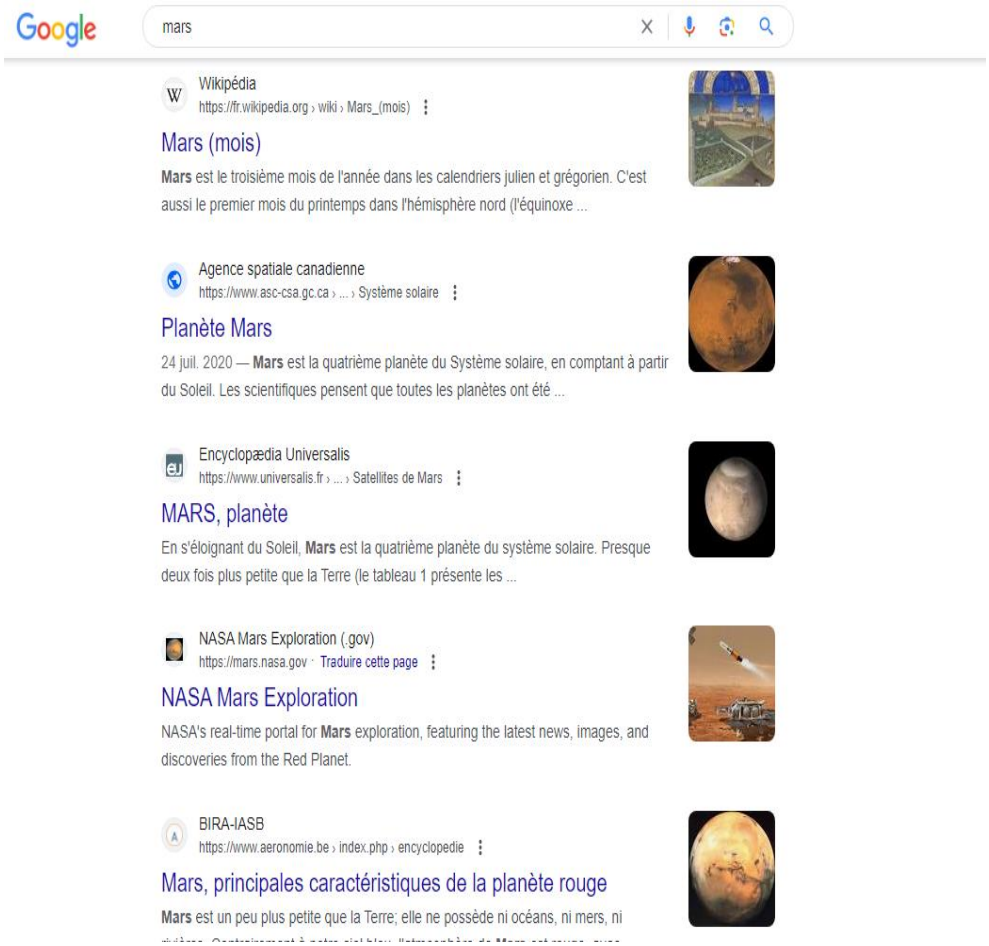


Figure 2.3:*De 5ème au 9ème résultat Google sur "Mars".*

Dans ce cas, si l'utilisateur intéressé par l'astronomie ou la science, les résultats sur la planète Mars sera satisfait à ces résultats (sera pertinents). En revanche, s'il recherchera des informations sur les événements se déroulant en Mars, le mois, comme des anniversaires, des fêtes nationales ou des événements historiques, les résultats sur le mois de Mars seraient mauvais pertinents dans quatre premiers résultats car il ne sont pas pertinents selon les besoins de l'utilisateur.

En résumé, l'émergence de la recherche d'informations personnalisées vise à satisfaire les exigences particulières des utilisateurs dans un contexte où l'abondance et la complexité de l'information continuent de s'accroître.

3 Modélisation de l'utilisateur

La modélisation de l'utilisateur est une discipline de recherche qui se concentre sur l'amélioration des interactions entre les humains et les machines. Son objectif principal est de comprendre et de prédire les intentions, les préférences et le contexte des utilisateurs en se basant sur les informations collectées lors des interactions. En d'autres termes, il s'agit d'analyser les actions et les comportements des utilisateurs pour mieux adapter les réponses et les fonctionnalités des systèmes informatiques afin de répondre à leurs besoins de manière plus efficace et intuitive(Haddad, 2015).

Pour créer un modèle de l'utilisateur, il est essentiel de comprendre comment structurer et utiliser les informations le concernant de manière efficace. Dans le domaine de la recherche d'information personnalisée, cette structure est connue sous le nom de "profil utilisateur". Ce profil est une représentation des caractéristiques d'un utilisateur spécifique, telles que ses intérêts, ses préférences et ses besoins en informations. Il est utilisé pour fournir un contenu adapté à chaque utilisateur, et il contient généralement une variété d'informations telles que des mots-clés, des concepts, des annotations, des dates, des lieux, des URL, etc., qui reflètent les centres d'intérêt de l'utilisateur (Ould Amer, 2020)(Krir, 2012).

On peut simplifier la modélisation de l'utilisateur en se demandant : quelles sont les informations significatives sur l'utilisateur pouvant représenter des capteurs de pertinence pour lui, que doit contenir son profil pour guider ses nouvelles recherches?

3.1 Pertinence Utilisateur

La pertinence est sans aucun doute la préoccupation principale lorsqu'il s'agit d'accéder à de l'information. Cependant, il convient de noter que la pertinence est un concept subjectif qui varie d'un utilisateur à l'autre et selon le type d'information recherchée. Elle est influencée par plusieurs facteurs tels que les intérêts de l'utilisateur, le contexte temporel et spatial, ainsi que le support utilisé pour accéder à l'information et le système qui la fournit.

Cette notion subjective, dépendant essentiellement du point de vue de l'utilisateur, dans le cadre de l'accès personnalisé à l'information. Dans ce contexte, la pertinence est

spécifiée comme étant un concept multidimensionnel (Borlund, 2003), dont on distingue principalement quatre types :

- **Pertinence algorithmique** : La pertinence algorithmique est déterminée par une mesure qui tient compte des caractéristiques distinctes des requêtes ainsi que des documents, calculée au moyen d'un algorithme. Cette forme de pertinence se distingue des autres par son caractère indépendant du contexte.
- **pertinence thématique :** La pertinence thématique mesure dans quelle mesure l'information répond, du moins partiellement, au thème abordé par la requête. Ce type de pertinence est celui qui est évalué par les évaluateurs lors de la campagne d'évaluation TREC.
- **Pertinence cognitive :** La pertinence cognitive se réfère à la pertinence qui est influencée par la perception ou les connaissances de l'utilisateur concernant le thème de la requête. C'est une pertinence "pondérée" par la compréhension préalable ou les expériences de l'utilisateur sur ce sujet spécifique.
- **Pertinence situationnelle** : La pertinence situationnelle se rapporte à la pertinence qui est directement liée à la tâche de recherche. Ce type de pertinence exprime principalement l'utilité de l'information par rapport à l'objectif de recherche spécifique de l'utilisateur.

La recherche d'information personnalisée se concentre principalement sur l'évaluation de la pertinence cognitive et de la pertinence situationnelle.

3.2 Profil d'utilisateur

Le profil utilisateur désigne une organisation permettant de représenter et de conserver les informations décrivant l'utilisateur au sein d'un système. Ce profil offre au système une compréhension des centres d'intérêt, des préférences et des besoins en information de l'utilisateur ou d'un groupe d'utilisateurs, lui permettant ainsi de fournir des réponses mieux adaptées.

Plusieurs définitions sont données à la notion du profil, on cite les deux suivantes :

Un profil utilisateur est une source de connaissance qui contient des acquisitions sur tous les aspects de l'utilisateur qui peuvent être utiles pour le comportement du système (Wahlster & Kobsa, 1986).

Le profil utilisateur est une représentation explicite des caractéristiques d'un utilisateur particulier qui peuvent être pertinentes pour une interaction personnalisée du système (Razmerita, 2009).

Pour modéliser le profil il faut décrire le « quoi » c'est-à-dire ce qui doit être représenté et le « comment » c'est-à-dire comment représenter le « quoi » au sein du profil (Amato & Straccia, 1999).

- Le « Quoi » : identifier le contenu à inclure dans le profil.
- Le « Comment » : définir la structure du contenu du profil et les techniques à utiliser pour le construire.

3.3 Sources des données de profil

Diverses sources peuvent fournir des informations caractérisant l'utilisateur, notamment l'historique des interactions, les documents visités, ainsi qu'environnement (temps, localisation) etc. (Rafa, 2023).

Une revue de la littérature montre que les sources les plus utilisées sont les suivantes (Bouidghaghen, 2011):

- Le comportement de l'utilisateur peut être appréhendé à travers des indicateurs d'évaluation implicite tels que l'historique des clics, les données de navigation, etc.
- Les pages et sites web préférés .
- Des données locales et contextuelles telles que les sources accédées, comme les journaux, les blogs et les sites de commerce en ligne.
- Les premières pages ou résumés de pages web renvoyées par un moteur de recherche.

3.4 Acquisition des données de profil

La première phase consiste à collecter les informations pertinentes pour instancier le profil de l'utilisateur. Ce processus peut collecter ces informations soit directement à partir de la machine de l'utilisateur (côté client) ou à partir de l'application (côté serveur). Ce processus d'acquisition peut être explicite et/ou implicite (Zemirli N. W., 2008):

3.4.1 L'acquisition explicite

Implique de recueillir directement des informations auprès de l'utilisateur en lui posant des questions ou en lui demandant de remplir des formulaires. Cela permet de collecter

ses préférences sur différents aspects, ainsi que des détails sur son environnement, tels que la taille de son écran, la puissance de son processeur ou la capacité de sa mémoire. En exprimant ses préférences, l'utilisateur fournit une indication directe de son niveau d'intérêt en attribuant une valeur de pertinence sur une échelle graduée, allant du moins intéressant au plus intéressant.

3.4.2 L'acquisition implicite

Dans le cadre de l'acquisition implicite, également appelée "feedback implicite", les informations sur l'utilisateur sont collectées en observant les dimensions et les membres qu'il sollicite fréquemment, ainsi qu'en scrutant les caractéristiques de l'environnement dans lequel il interagit, telles que les capacités et les limites du dispositif utilisé. Cette collecte se base sur l'historique des interactions de l'utilisateur avec le système.

L'avantage principal de cette approche réside dans le fait qu'elle ne requiert aucune participation directe de l'utilisateur ni d'effort de sa part pour fournir des informations. De plus, elle préserve la sécurité et la confidentialité des données. Cependant, l'inconvénient réside dans la complexité des algorithmes utilisés, qui peuvent nécessiter beaucoup de temps pour être mis en place.

3.5 Techniques de construction de centres d'intérêts

Les modèles des centres d'intérêts des utilisateurs sont construits en utilisant une variété de techniques d'extraction d'informations à partir des sources d'information utilisateur (Krir, 2012):

- *Extraction d'ensemble de termes :* La technique d'extraction d'ensemble de termes se fonde sur des approches statistiques pour analyser les mots clés. En effet, elle consiste à analyser le contenu des documents que l'utilisateur consulte afin de repérer les termes significatifs. Ces termes sont ensuite attribués des poids pour refléter leur importance, formant ainsi des vecteurs de termes qui représentent les centres d'intérêt de l'utilisateur.
- *Extraction de réseaux de termes :* De manière analogue à l'extraction d'un ensemble de termes, cette approche implique l'extraction de termes à partir des documents que l'utilisateur juge pertinents. Cependant, elle se distingue par sa représentation des termes sous forme de réseau de nœuds. Cette méthode se concentre principalement sur les représentations sémantiques du modèle utilisateur. Ainsi, pour construire ce

modèle, il est primordial d'exploiter les relations préexistantes entre les termes et les concepts, qui peuvent être puisées dans des sources sémantiques telles que WordNet.

- *Extraction de concepts :* Cette méthode de construction est principalement dédiée à élaborer des modèles d'utilisateurs qui sont représentés par une hiérarchie de concepts pondérés. Son principe de base réside dans l'utilisation d'une taxonomie de concepts comme référence fondamentale. En général, cette approche comprend deux grandes étapes : (1) l'identification des concepts et des niveaux pertinents de l'ontologie à exploiter, et (2) l'extraction des centres d'intérêt de l'utilisateur en les associant aux concepts de l'ontologie par analogie.

3.6 Représentation des données de profil d'utilisateur

Le profil utilisateur peut être décrit à l'aide de divers modèles de représentation. Dans ce qui suit, nous examinons en détail les principaux types de modèles de représentation du profil utilisateur.

3.6.1 Représentation basée sur l'historique de recherche

Cette représentation simple et basique englobe toutes les requêtes et les pages web visitées ou cliquées par l'utilisateur, ainsi que toutes les requêtes et les résumés textuels de leurs résultats associés, accumulés au cours des sessions de recherche de l'utilisateur (Rafa, Combinaison d'un Profil Social et un Profil Situationnel pour la Recherche d'Information, 2023).

Nous citons dans cette catégorie, les travaux des auteurs (Raghavan & Sever, 1995)qui représentent le profil utilisateur par son historique de recherche en se servant comme une base de données des requêtes soumises précédemment ainsi que leurs résultats associés. Cette base de données est utilisée pour sélectionner les requêtes les plus similaires à une requête en cours d'évaluation. Aussi l'approche proposée par l'article (Tan, Shen, & Zhai, 2006), où le profil utilisateur est représenté par l'historique des requêtes passées et l'historique des clicks sur les résultats sélectionnés par l'utilisateur.

3.6.2 Représentation ensembliste

La représentation ensembliste du profil utilisateur repose sur une collection de mots clés (ou vecteurs de mots clés) pondérés, généralement représentés selon le modèle vectoriel (Yang & Salton, 1973) approche constitue la première méthode conçue pour caractériser le profil utilisateur.

Les ensembles de termes reflètent les centres d'intérêt de l'utilisateur. Le poids de chaque terme est généralement déterminé par le schéma TF*IDF, largement utilisé en recherche d'information, et représente l'intensité de l'intérêt de l'utilisateur pour ce terme dans son profil. Dans (Daoud, 2009), l'auteur décompose la représentation ensembliste en trois sous-modèles :

- Une collection de termes pondérés, où chaque terme représente un potentiel centre d'intérêt de l'utilisateur.
- Un vecteur de termes pondérés qui représente un centre d'intérêt individuel (Lieberman, 1997).
- Une collection de vecteurs de termes pondérés, chacun représentant un centre d'intérêt distinct (Gowan, 2003).

3.6.3 Représentation connexionniste

Dans (Rafa, 2023) la représentation connexionniste, les centres d'intérêts de l'utilisateur sont représentés par un réseau de nœuds pondérés, où chaque nœud représente un concept correspondant à un centre d'intérêt. Les relations de dépendance sémantique entre les centres d'intérêt du modèle de l'utilisateur, établies dans cette représentation, permettent de résoudre les lacunes de la représentation ensembliste, telles que la polysémie des termes et les éventuelles incohérences entre les centres d'intérêts. Cette représentation apporte de la sémantique au modèle de l'utilisateur, facilitant ainsi une meilleure compréhension et interprétation des données.

Cette représentation présente néanmoins certaines limitations. En effet, la source de données du réseau sémantique représentant le modèle de l'utilisateur n'est autre que l'historique de recherche de l'utilisateur qui est souvent limité, d'où la nécessité d'utiliser des sources sémantiques externes tels que les ontologies (Daoud, 2009).

3.6.4 Représentation multidimensionnelle

La représentation multidimensionnelle du profil s'inscrit dans une approche globale de la personnalisation de l'information. En effet, le profil est organisé en fonction d'un ensemble de dimensions, représentées à travers différents formalismes.

Différentes propositions ont été abordés pour cet aspect, on a le standards P3P pour la sécurisation des profils ont défini des classes distinguant les attributs démographiques des utilisateurs (identité, données personnelles), les attributs professionnels (employeur, Adresse, type) et les attributs de comportement (trace de navigation). Une

autre proposition faite par Amato (Amato & Straccia, 1999)consiste à représenter le contenu du profil utilisateur par un modèle structuré de dimensions (ou catégories) prédéfinis. C'est la première approche où les informations sont structurées et qui offre un modèle général.

Le modèle de profil contient cinq catégories : catégorie de données personnelles, catégorie de données collectées, catégorie de données de livraison, catégorie de données de comportement, catégorie de données de sécurité.

Amato et kostadinov (Amato & Straccia, 1999)(Kostadinov, 2003), ils ont proposé un ensemble de dimensions ouvertes, capables d'accueillir la plupart des informations caractérisant un profil. Il distingue principalement huit dimensions : les données personnelles, le centre d'intérêt, l'ontologie du domaine, la qualité attendue, la customisation, la sécurité, le retour de préférences, Les informations diverses

3.6.5 Représentation conceptuelle

Pour surmonter les limitations de la représentation connexionniste, plusieurs études adoptent une approche basée sur une représentation conceptuelle du profil utilisateur. Cette méthode exploite les ontologies de domaines ou les hiérarchies de concepts préalablement établies. La représentation conceptuelle commence par définir les niveaux des concepts de l'ontologie à prendre en compte. Ensuite, elle applique un processus de pondération des données pour ces concepts. En fin de compte, le profil utilisateur est représenté par un réseau de nœuds conceptuels interconnectés, respectant la structure des liens définis dans les hiérarchies ou ontologies utilisées. On peut identifier différentes formes de représentations conceptuelles, telles que les hiérarchies de concepts, des parties d'ontologies de référence, des matrices de concepts, ou encore des graphes de concepts (Bouidghaghen, 2011).

De l'association des centres d'intérêts de l'utilisateur aux concepts des domaines de l'ontologie, on obtient un profil représenté sous forme d'une hiérarchie de concepts. Les données implicites issues des interactions de l'utilisateur sont classifiées dans ces concepts et l'intérêt de l'utilisateur pour de tels concepts est enregistré. Plusieurs mécanismes peuvent être appliqués pour exprimer le degré d'intérêt de l'utilisateur pour chaque domaine. Néanmoins la technique la plus utilisée est l'affectation d'un poids d'une valeur numérique pour chaque domaine (Gauch, Chaffe, & Pretschner, 2003).

3.7 Evolution de profil utilisateur

L'évolution du profil utilisateur représente un processus complémentaire à sa création initiale, impliquant l'adaptation du profil aux évolutions des centres d'intérêt des utilisateurs au fil du temps.

L'évolution du profil utilisateur implique principalement la saisie des évolutions des centres d'intérêt de l'utilisateur dans un premier temps, suivie de la propagation de ces changements au niveau de la structure du profil.

Les méthodes de collecte d'informations employées dans le processus d'évolution du profil utilisateur sont étroitement liées à la portée temporelle du profil. On différencie ainsi le profil à long terme et le profil à court terme.

L'évolution du profil à court terme(Gowan, 2003) représente les centres d'intérêts liés aux activités de recherches courantes de l'utilisateur. Il sert à mieux cibler la recherche puisqu'il contient des informations considérées spécifiques et pertinentes au besoin en information de l'utilisateur. Le but fondamental de l'évolution du profil à court terme est d'améliorer la précision de recherche en utilisant le profil le plus utile et approprié à la requête.

L'évolution du profil à long terme représente les centres d'intérêts persistants de l'utilisateur et issus de son historique de recherche tout entier. Il consiste à ajouter ou modifier un profil préalablement appris selon des changements éventuels des centres d'intérêts de l'utilisateur au cours des sessions de recherche (Sieg, Mobasher, Burke, Prabu, & Ly, 2004).

4 Les dimensions du contexte de recherche

Le contexte de recherche en Recherche d'Information (RI) englobe divers facteurs influençant la perception de pertinence des résultats, comme les caractéristiques de l'utilisateur et le domaine de la recherche (Borlund, 2003); (Taylor, Cool, Belkin, & Amad, 2007); (Saracevic, 2007). Les capteurs de pertinence exploitent des données thématiques et contextuelles (sociales, temporelles, etc.) pour améliorer la précision des systèmes de recherche personnalisée (Moulahi, 2015) ;(Bouidghaghen, 2011).

4.1 Dimension sociale

Les informations sociales sont utilisées pour créer des profils d'utilisateurs dans la recherche d'information personnalisée, comprenant annotations, commentaires, notes, et relations sociales. Deux types principaux sont identifiés :

4.1.1 Les annotations sociales (Tags)

Les annotations sociales ou tags, reflètent les centres d'intérêt des utilisateurs et leur perception du contenu des documents (Kumar, Lee, & Kim, 2014); (Ould Amer, 2020). Différentes approches évaluent le poids des tags selon leur fréquence d'utilisation (TF), TF-IDF, ou modèles combinés (Noll & Meinel, 2007), (Vallet, Cantador, & Jose, 2010). Tags utilisés décrivent souvent les documents, facilitant la recherche et partage (Heymann, Ramage, & Garcia-Moli, 2008) ;(Bischoff & al, 2008).

4.1.2 Les relations sociales

L'intégration des relations sociales améliore la personnalisation des résultats en enrichissant le profil utilisateur avec les interactions sociales (tweets, commentaires, tags) et les comportements des utilisateurs proches (Smyth, 2011). Diverses méthodes exploitent les relations sociales pour la recherche personnalisée, en analysant par exemple les tweets ou les bibliographies des abonnés (Ould Amer, 2020); (Badache & al, 2019); (Bouhini, Gery, & Largeron, 2016); (Saoud & Kechid, 2016); (Shafiq & al, 2015).

4.2 Dimension localisable

L'intégration de la localisation dans la recherche d'information est motivée par les technologies mobiles (PDAs, téléphones, tablettes), qui permettent un accès ubiquitaire au contenu en ligne. Cette dimension se focalise sur le lieu de recherche, prenant en compte la position géographique de l'utilisateur comme élément crucial (Bouidghaghen, 2009). La pertinence des résultats de recherche varie selon la localisation de l'utilisateur, par exemple, la recherche de "cinémas à proximité" montrerait des résultats différents à Paris et à Tokyo.

Des études montrent que 25% des besoins informationnels des utilisateurs mobiles sont liés à la localisation. Les données situationnelles (GPS, noms et types de lieux) sont utilisées pour filtrer les informations en fonction de la proximité géographique (Bouidghaghen, 2011). Une approche de personnalisation de la recherche est basée sur un profil situationnel de l'utilisateur, intégrant les différentes situations de recherche et leurs centres d'intérêt (Bouidghaghen, 2011).

(Akermi, Boughanem, & Faiz, 2015)Proposent une recommandation personnalisée guidée par un profil situationnel basé sur la localisation et les préférences de l'utilisateur. (Bila, Cao, Dino, Ho, & Hull, 2008)Présente un profil où les situations de l'utilisateur sont représentées par les lieux les plus visités, et les relations entre ces lieux sont modélisées par un arbre. (Hattori, Tezuka, & Tanaka, 2007)Traitent l'ambiguïté des requêtes par une méthode d'expansion basée sur la localisation, transformant les coordonnées spatiales en mots contextuels via des systèmes d'information géographique et des techniques d'apprentissage des weblogs.

4.3 Dimension temporelle

L'aspect temporel est crucial pour comprendre les intérêts des utilisateurs en tenant compte du moment de la recherche, sa durée, et l'évolution des préférences. Intégrer cette dimension améliore les modèles de Recherche d'Information (RI) traditionnels, en considérant les événements marquants et les informations temporelles (Moulahi, 2015).

Les informations temporelles concernent le moment des requêtes et des publications des documents. (Bouidghaghen, 2011)Propose de représenter le temps en périodes spécifiques (matin, weekend, été) et utilise des ontologies temporelles pour mieux manipuler l'information. (Vandenbussche & Teissèdre, 2011)Combinent les informations temporelles des documents et de leur contenu, tandis que (Kanhabua & Nørvag, 2012) et (Basile, Caputo, & Semera, 2014)développent des systèmes pour l'extraction et l'indexation des événements à partir des textes.

4.4 Dimension sémantique

La sémantique dans la recherche d'information (RI) personnalise l'expérience utilisateur en dépassant les limitations des mots-clés traditionnels, souvent imprécis à cause de l'ambiguïté et de la disparité des mots (Boubekeur, 2008). En comprenant la signification des termes dans leur contexte et en tenant compte des synonymes et des concepts associés, la sémantique améliore la pertinence des résultats de recherche.

Les approches de RI sémantique utilisent des ressources externes comme les ontologies, thesaurus, et dictionnaires pour établir des liens entre les termes (Zargayouna, Roussey, & Chevallet, 2015).

La figure ci-dessous représente le triangle sémiotique proposé par les auteurs (Ogden & Richards, 1923) qui illustre la relation entre terme et concept

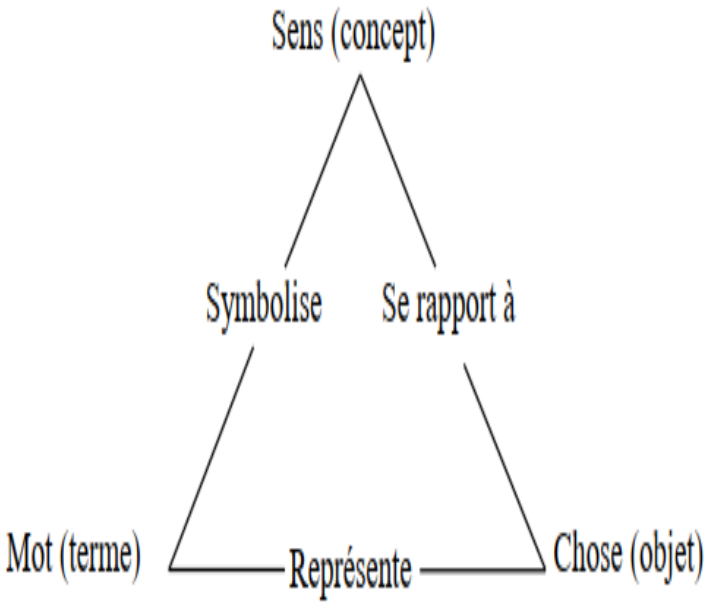


Figure 2.4: *Triangle sémantique.*

Récemment, de nombreuses études ont intégré la sémantique dans les profils utilisateurs pour améliorer la précision des résultats. (Daoud, Tamine, & Boughanem, 2011)Proposent de modéliser les profils et documents en graphes de concepts à partir de l'ontologie ODP et utilisent une mesure de distance sémantique pour le ré-appariement. (Duong, Uddin, & Nguyen, 2013) emploient des concepts de l'ODP pour annoter les documents pertinents et créer un profil sémantique.

(Ould Amer, 2020) met en avant l'utilisation des concepts de bases de connaissances comme DBpedia pour représenter les centres d'intérêts des utilisateurs, avec des méthodes de représentation par entités, par catégories, ou une combinaison des deux, afin de capturer des intérêts non identifiables par les mots-clés seuls.

5 Intégration de profil utilisateur dans le processus de recherche

Une fois que le profil de l'utilisateur est construit, il est intégré dans le processus de recherche pour tirer parti des indicateurs de pertinence, permettant d'anticiper les nouveaux besoins en information.

Le SRIP utilise les informations continues du profil pour cibler un ensemble de documents, en les comparants avec les descripteurs de la collection. Les documents similaires sont alors sélectionnés pour constituer le sous-espace de recherche. Cette intégration se produit lors de :(1) la reformulation de la requête initiale ou du (2) reclassement des résultats de la recherche.

5.1 Reformulation de requête

Généralement, les utilisateurs expriment leurs besoins en information à travers des requêtes courtes comprenant peu de mots en moyenne (2,35 mots/requête)(Rafa, 2023). Ces requêtes succinctes peuvent être imprécises et ambiguës quant aux besoins réels de l'utilisateur. Par exemple, la requête "Amazon" pourrait être associée à la forêt amazonienne, relevant ainsi du domaine de la géographie, tout comme elle pourrait faire référence à l'entreprise multinationale américaine qui propose une variété de services en ligne (vente en ligne ...etc.), relevant alors du domaine de la technologie de l'information et de la communication. C'est pourquoi les moteurs de recherche personnalisés sont amenés à reformuler la requête initiale de l'utilisateur afin de mieux comprendre ses besoins en information, en s'appuyant sur les informations disponibles dans son profil. Cette reformulation de requête intègre les composantes informationnelles du profil de l'utilisateur pour identifier, enrichir et cibler précisément son intention de recherche.

Le principal objectif de la reformulation de requêtes en utilisant le profil est de cibler la recherche de documents pertinents en enrichissant la requête avec des termes issus du profil de l'utilisateur, afin de mieux répondre à ses besoins en information. Dans la plupart des approches de reformulation ou d'expansion de requêtes utilisant les informations du profil utilisateur, la question est d'identifier les meilleurs termes à ajouter à la requête initiale de l'utilisateur. La requête initiale Q est alors transformée en une nouvelle requête Q' censée mieux représenter le besoin en information de l'utilisateur. Cette transformation de Q en Q' peut prendre la forme d'un raffinement ou d'une expansion. Le raffinement consiste à réduire la requête pour éliminer les informations superflues, tandis que l'expansion ajoute de nouveaux termes à la requête initiale pour la rendre moins ambiguë et élargir son champ de recherche (Badache, 2016).

La problématique de reformulation de requête est traitée différemment dans les approches de recherche d'information personnalisée, par exemple :

L'approche dans (Tamine, Boughanem, & Zemirli, 2008)intègre le profil utilisateur dans la fonction d'appariement du modèle de recherche bayésien. Ainsi, la pertinence d'un document par rapport à une requête ne dépend plus uniquement de la requête elle-même, mais également du centre d'intérêt de l'utilisateur qui la soumet.

Dans (Wen, Lao, & Ma, 2004), l'expansion des termes de la requête repose sur les corrélations entre les termes de la requête, les éléments contextuels et les termes des documents. Lorsqu'une nouvelle requête est soumise avec un contexte spécifique, une liste de termes issus des documents corrélés est identifiée, puis triée en fonction des probabilités conditionnelles calculées à l'aide de l'un des modèles probabilistes proposés. Ensuite, les termes les plus pertinents sont sélectionnés et utilisés pour reformuler la requête de l'utilisateur.

Certaines autres approches proposent de sélectionner les termes du profil par des méthodes de sélection sémantique. Donc, pour chaque terme de la requête, les top-k termes similaires sont sélectionnés. Les méthodes peuvent s'appuyer sur des approches exploitant des matrices de cooccurrences (Biancalana & al, 2013), des plongements de mots (word2vec) (Ould Amer, Mulhem, & Géry, 2016) ou des modèles thématiques (Zhou, Lawless, & Wade, 2012). D'autres approches proposent d'inclure le cercle social de l'utilisateur (Bouhini, Géry, & Large, 2014)(Ould Amer, Mulhem, & Géry, 2016) pour enrichir le profil de l'utilisateur ou étendre la requête avec des termes des profils du cercle social.

5.2 Reclassement de résultats

Étant donné que la plupart des utilisateurs ne consultent que les premiers résultats renvoyés par les moteurs de recherche, il devient crucial de privilégier le classement des quelques documents les plus pertinents en tête de la liste des résultats, plutôt que de simplement fournir un grand nombre de documents pertinents mais mal classés. Ainsi, il est nécessaire de réorganiser les résultats de recherche afin de mettre en avant les documents les plus pertinents, en tenant compte du profil de l'utilisateur (Rafa, 2023).

Les approches basées sur le ré ordonnancement des résultats de recherche (Challam, Gauch, & Chandramouli, 2007)(Mobasher, Sieg, & Burke, 2007) consistent souvent à combiner le score original du document et son score de similarité avec le profil d'utilisateur. Des variantes des approches de ré ordonnancement des résultats consiste en une catégorisation personnalisée (Ma, Pant, & Sheng, 2007) basée sur la classification des résultats de recherche dans des catégories représentatives des centres

d'intérêts du profil utilisateur. La personnalisation dans (Liu, Yu, & Meng, 2004)consiste à créer plusieurs listes de résultats associées aux catégories associées à la requête, ensuite à les combiner selon une méthode de ré ordonnancement par vote majoritaire.

6 L'évaluation des systèmes de recherche d'information personnalisés

Divers travaux ont tenté de mettre en place un cadre d'évaluation approprié aux SRI personnalisés. Il en ressort que l'objectif d'un tel protocole d'évaluation est de mesurer l'efficacité de la méthode d'apprentissage (construction et évolution) du profil utilisateur, et évaluer l'impact de l'intégration du profil utilisateur dans le processus d'accès sur les performances de recherche. De ce fait, tout protocole d'évaluation doit répondre à deux exigences (Zemirli N. W., 2008):

- Valider l'approche de personnalisation en mesurant l'adéquation du profil utilisateur ainsi que l'efficacité de la méthode de construction du profil utilisateur.
- Tester les paramètres de l'approche de personnalisation à travers la comparaison des performances du SRIP obtenus avec l'intégration du profil de l'utilisateur et ceux obtenus sans son intégration.

6.1 Les limites de l'évaluation classique

L'évaluation est une étape permettant de mesurer la capacité des systèmes de recherche d'information à satisfaire les besoins d'information des utilisateurs. Un système de recherche d'information classique est évalué à travers d'une collection de test comprenant trois éléments essentiels qui sont (Bouhini C. , 2014) :

- **Un ensemble de requêtes** qui représentent le besoin en information des utilisateurs. Cet ensemble est fourni afin d'évaluer pour chaque requête la capacité du système retrouver les documents pertinents.
- **Un ensemble de documents** qui contiennent les informations pertinentes qui répondent à l'ensemble des requêtes des utilisateurs.
- **Les jugements de pertinence** qui permettent d'identifier les documents pertinents pour chaque requête. Ces jugements sont fournis généralement par des utilisateurs.

Malgré la popularité et la reconnaissance de ces deux campagnes d'évaluation « TREC et CLEF »qui est généralement exprimées par les métriques précision/rappel, néanmoins ces approches pour l'évaluation des systèmes RI présentent certaines limites liées au sujet de la prise en compte de l'utilisateur.

Les tests effectués ne prennent pas en considération ni le contexte dans lequel se fait la recherche, ni la perception de la pertinence de l'utilisateur dans ce contexte, ni de la diversité des centres d'intérêts des utilisateurs.

Pour mieux comprendre les limites de l'évaluation traditionnelle des SRI, Nous avons examiné les recherches menées par (Tamine, Boughanem, & Daoud, 2010), (Chaudiron & Ihadjadenem, 2002),(Menegon, Mizzaro, Nazzi, & Vassena, 2009). À partir de cette synthèse, nous avons identifié trois catégories de problèmes, ces problèmes comprennent l'absence de l'utilisateur dans l'évaluation, les jugements de pertinence, ainsi que les limitations des corpus de documents et de requêtes.

- **Limites par rapport à l'utilisateur :** Nous pouvons reprocher à ces approches d'évaluation d'être artificielles et arbitraires. Si le TREC a effectivement amélioré l'efficacité du système, la notion d'utilisateur final implique des connaissances personnelles, une expérience et des capacités de recherche différentes, dont l'évaluation du système ne soucie pas. En effet, ces évaluations ne tiennent pas compte du contexte dans lequel se fait la recherche puisqu'elles ne sont pas effectuées en situation d'utilisation réelle. Dans ce contexte (Chaudiron & Ihadjadenem, 2002)affirme que l'absence de l'usager dans le processus d'évaluation est l'une des premières et sans doute la plus importante critique des approches classiques qui utilisent des critères autres que le rappel et la précision.
- **Limites par rapport aux jugements de pertinence :** La pertinence est une notion subjective et il paraît impensable de pouvoir la mesurer sans être arbitraire. Nous pouvons noter également que les jugements de pertinence dans le TREC opèrent de façon binaire : un document est jugé pertinent ou ne l'est pas. Pourtant, ce n'est évidemment pas toujours le cas, certains documents sont plus pertinents que d'autres qui le sont quand même. Ces degrés de pertinence dépendent encore de la disposition d'esprit de la personne ayant réellement besoin de ces documents. Cette constatation est validée par les travaux de(Tamine, Boughanem, & Daoud, 2010)montrant que la pertinence considérée dans l'évaluation classique des SRI est thématique, indépendante du contexte, de la situation de recherche et des centres d'intérêt des utilisateurs. De même les travaux de (Chaudiron & Ihadjadenem, 2002)ont montré que les jugements de pertinence doivent être nuancés dans le sens où ils sont stables et ne varient pas dans le temps et qu'ils sont attribués indépendamment les uns désastres.

- **Limites par rapport au corpus de documents et de requêtes :** Dans les corpus traditionnels, un document est un texte à part entière, et l'évaluation se fait par rapport au nombre de documents retrouvés, or en général, un utilisateur ne cherche pas des documents mais de l'information, et les documents ne contiennent jamais la même quantité d'information. De même pour les corpus de requêtes où la requête est un besoin d'information exprimé en langage naturel. Or, c'est tout le problème de la représentation du besoin d'information de l'usager. La tâche de RI se transforme en tâche de savoir poser les questions à ces systèmes car les écarts sont grands entre ce que nous pensons et ce qui est interprété.(Tamine, Boughanem, & Daoud, 2010)constate que le protocole d'évaluation étant en mode batch, les requêtes sont supposées ainsi représenter à elles seules l'utilisateur. Par conséquence les utilisateurs directs ayant émis ces requêtes, leurs centres d'intérêt et interactions avec le SRI ne font pas partie intégrante de la collection.

6.2 La problématique de l'évaluation personnalisée

Selon (Tamine & Daoud, 2018) (Toms, Agosti, Fuhr, & Pe, 2013), l'amélioration de la pertinence dans le contexte de la personnalisation de la recherche d'information est très subjective, car elle dépend principalement de l'utilisateur, qui est le seul à pouvoir évaluer la pertinence en fonction de ses propres critères. Cette réalité rend l'évaluation des approches de recherche d'information personnalisée extrêmement difficile, vue l'absence d'une collection de tests avec des requêtes et des pré- jugements faites par les utilisateurs concernés par l'étude eux-mêmes, on ne peut pas évaluer les améliorations apportées par une nouvelle approche pour l'utilisateur.

L'évaluation d'un système de recherche d'information personnalisée est un véritable défi, du fait qu'il est nécessaire de fournir les composantes principales suivantes (Bouhini, Gery, & Largeron, 2016) :

- **Une collection des informations sur les utilisateurs**, par convention appelée profil d'utilisateur, servant à l'organisation, la sauvegarde et l'exploitation des données sur l'utilisateur et son historique de recherches et d'interactions avec le système.
- **Une Dataset (collection de documents + ensemble de requêtes)**, pour laquelle doit être associé un ensemble de pré-jugements de pertinence fait par les mêmes utilisateurs concernés. Les pré-jugements des utilisateurs servent

pour référence d'évaluation des performances des nouvelles approches de personnalisation de la recherche.

La difficulté est qu'il n'existe pas de collections de tests standards pour l'évaluation de l'efficacité des systèmes de recherche d'information personnalisée. Fournir de telles collections est très complexe par le fait de la difficulté de la mise en place d'un système qui permet de collecter les informations des utilisateurs, leurs requêtes et ensuite les jugements de pertinence pour tous les documents pour chaque requête pour chaque utilisateur (Ould Amer, 2020).

6.3 Protocole de l'évaluation personnalisée

Vue l'absence d'un cadre d'évaluation standard ou consensuel dédié à la recherche d'information personnalisée, les approches de ce domaine ont eu recours à développer leurs propres cadres d'évaluation.

Ci-dessous, nous présentons quelques exemples de solutions :

Pour la collecte des données utilisateur en vue de construire les profils, différentes approches ont été adoptées.

- **Collection des informations sur l'utilisateur**

- (Bouhini, Gery, & Largeron, 2016),(Saoud & Kechid, 2016)ont exploité l'historique de bookmarks du site social Delicious.
- (Daoud, Tamine, & Boughanem, 2011) ont utilisé les fichiers logs de recherche du site commercial Exalead.
- (Mohamed & Abdelmoty, 2017) ont tiré parti de l'historique des check-ins des utilisateurs de la base de données géographique Four square sur une période définie.
- (Duong, Uddin, & Nguyen, 2013) ont utilisé Google Scholar Citation et Facebook pour collecter les données des utilisateurs.
- (Bouidghaghen, 2011)a proposé un algorithme basé sur les interactions utilisateur-requête pour simuler le profil utilisateur.
- (Bouramoul, Kholladi, & Doan, 2011) ont directement interrogé les utilisateurs au moyen d'un ensemble de questions pour définir leur profil.

- **Dataset**

  Concernant les Dataset utilisées pour évaluer les performances des approches proposées, un défi majeur réside dans l'absence de pré-jugements des utilisateurs qui permettraient de mesurer la pertinence des résultats de recherche. Pour résoudre ce problème, différentes approches de recherche d'information personnalisée proposent des solutions adaptées

- (Daoud, Tamine, & Boughanem, 2011) ont utilisé comme ensemble de données les 50 premiers documents retournés par le moteur de recherche Yahoo pour chacune des 50 requêtes envoyées.
- (Bouidghaghen, 2011)a retenu les 150 premiers documents pour chacune des 25 requêtes.
- (Saoud & Kechid, 2016), intéressés par la recherche distribuée, ont utilisé 10 moteurs de recherche différents et ont sélectionné comme ensemble de données les 20 meilleurs documents retournés par chaque moteur pour 90 requêtes.
- (Bouramoul, Kholladi, & Doan, 2011) ont considéré comme ensemble de données les 20 meilleurs documents retournés par les 3 moteurs de recherche Google, Yahoo et Bing pour chacune des 30 requêtes envoyées.

- **Evaluation des performances**

  Comme ces ensembles de données ne disposent pas de pré-jugements de pertinence des utilisateurs, sur lesquels les données sont collectées et les profils sont construits, les approches de recherche d'information personnalisée les remplacent par des post-jugements réalisés par des individus appelés « Participants ».

  Dans les études de [Daoud et al 2011 et Duang et al 2013], cinq participants ont été invités à évaluer la pertinence des résultats obtenus. Et respectivement, 6,15 et 24 participants sont invités dans(Bouidghaghen, 2011), (Saoud & Kechid, 2016)et(Bouramoul, Kholladi, & Doan, 2011).Les améliorations des approches proposées sont évaluées en fonction des post-jugements de pertinence fournis par ces participants.

6.4 Les mesures de l'évaluation des systèmes personnalisée

Pour l'évaluation d'un système d'accès personnalisé à l'information, on distingue en plus des mesures standards de précision et de rappel (voir section 5.2 chapitre 1) quelques autres mesures :

- **La mesure RR (Relative Relevance) :**

  L'objectif de cette mesure (Borlund & Ingwersen, 1998)est de prendre en compte les différents types de pertinence lors de l'évaluation d'un Système de Recherche d'Information Contextuelle (SRIC). Elle mesure le degré de correspondance entre les types de jugements de pertinence émis pour deux ensembles de jugements. Soit deux valeurs de jugements de pertinenceR_1 etR_2. Supposons queR_1 représente les scores de pertinence algorithmique renvoyés par un SRI, et R_2représente les scores de pertinence contextuelle correspondant à un type de pertinence « situationnelle ou thématique » La concordance entre R_1 et R_2 est habituellement évaluée à l'aide d'une mesure de Cosinus ou Jaccard comme suite :

$$RR(R_1R_2) = \frac{\sum(R_1 * R_2)}{(\sum R_1^2)^{\frac{1}{2}} * (\sum R_2^2)^{\frac{1}{2}}}$$

-Formule de Jaccard :

$$RR(R_1R_2) = \frac{\sum(R_1 * R_2)}{\sum(R_1) + \sum(R_2) - \sum(R_1 * R_2)}$$

Une valeur plus petite de RR indique une correspondance moindre entre la pertinence du système et la perception de pertinence de l'utilisateur dans un contexte de recherche. Cette mesure permet de comprendre comment les performances d'un SRI se comportent face à la subjectivité de l'utilisateur.

- **Les mesure CG (Cumulative Gain) et DCG (Discount Cumulative Gain) :** Les mesures CG et DCG (Jarvelin & Kekalainen, 2002) , sont des mesures orientées position définies dans le contexte d'une pertinence graduelle et dont l'objectif est d'estimer le gain de l'utilisateur en termes de pertinence cumulée en observant les documents situés jusqu'à un rang donné. Ces mesures sont définies comme suit :

$$\boldsymbol{CG[i] = \begin{cases} G[1], sii = 1 \\ CG[i-1] + G[i], sinon \end{cases}}$$

Où :

- G[i] est la valeur de pertinence associée au document de rang i.
- $\boldsymbol{CG[i] = \begin{cases} G[1], sii = 1 \\ CG[i-1] + G[i]/log_q, sinon \end{cases}}$

Comparativement à la mesure CG, la mesure DCG permet d'atténuer le gain de pertinence apporté par un document en fonction du rang associé. Ceci rejoint en effet l'hypothèse évidente que plus le rang d'un document est élevé, moins il est probable que l'utilisateur l'examine et donc moins il est à l'origine d'un gain effectif de pertinence. Les mesures CG et DCG présentent l'avantage d'estimer le gain de pertinence à n'importe quel rang, sans dépendre de la base de rappel, qui doit être fixée préalablement.

- **La mesure GRP (Generalized Recall and Precision)**

  La mesure GRP, proposée par (Jarvelin & Kekalainen, 2000),est une mesure axée sur la position qui étend les mesures classiques de rappel et de précision en tenant compte d'une pertinence graduelle. Le rappel généralisé (GR) et la précision généralisée (GP) sont calculés comme suit :

$$GP = \sum_{d \in R} r(d)/|R|$$

$$GR = \sum_{d \in R} r(d) / \sum_{d \in D} r(d)$$

Où :

- R est l'ensemble des documents retournés par le SRI.
- D est l'ensemble des documents de la collection.
- r(d) est la valeur de pertinence graduelle associée au document d.

De manière similaire aux mesures classiques de rappel/précision, ces mesures peuvent être agrégées pour plusieurs requêtes ou différents niveaux de rappel, permettant ainsi de tracer des courbes de performances.

7 Conclusion

En conclusion, ce chapitre a exploré en détail le domaine de la recherche d'information personnalisée. Nous avons examiné les différentes facettes de la modélisation de l'utilisateur, en mettant en lumière la pertinence de l'utilisateur, la construction et l'évolution du profil utilisateur, ainsi que les diverses dimensions du contexte de recherche. Nous avons également abordé l'intégration du profil utilisateur dans le processus de recherche, notamment à travers la reformulation de requête et le reclassement des résultats. Enfin, nous avons étudié les défis et les méthodes d'évaluation des systèmes de recherche d'information personnalisés, en soulignant les limites de l'évaluation classique et les complexités associées à l'évaluation personnalisée.

Chapitre 03

Reformulation Interactive Personnalisée

Chapitre 03 : Reformulation Interactive Personnalisée.

1 Introduction

Il est évident qu'une requête basée sur un ou deux mots-clés ne capture qu'une fraction des besoins de l'utilisateur. Lorsque la formulation de ces besoins est déjà une tâche complexe pour l'utilisateur du système de recherche d'information (SRI), elle devient encore plus difficile en cas d'imprécision ou de manque de familiarité avec le domaine de recherche. Par conséquent, il est important d'approfondir la compréhension du contexte des besoins de l'utilisateur pour améliorer les performances des SRI (Saint Requier, Dupont, & Sébastien, 2010).

Pour améliorer l'expression des besoins en information de l'utilisateur, il est essentiel d'enrichir la requête en ajoutant des termes supplémentaires en lien avec son sujet (Raza, Mokhtar, & Ahmad, 2018); (Bendella & Quafafou, 2019).Pour atteindre cet objectif, les systèmes de recherche d'information personnalisée mettent en œuvre une structure, désignée conventionnellement sous le nom de profil d'utilisateur (Uddin, Duong, Sean, & Jo, 2012), qui offre une description détaillée de l'utilisateur et de ses centres d'intérêt. Ce profil utilisateur est ensuite intégré dans les différentes phases du processus de recherche, notamment la reformulation de requêtes afin d'améliorer la pertinence des résultats de recherche.

Dans le but d'aider le SRI à répondre de manière plus précise aux besoins de l'utilisateur, la reformulation automatique de requêtes peut être bénéfique si les requêtes initiales permettent de trouver des documents pertinents. Cependant, lorsque ce n'est pas le cas, elle peut entraîner une dégradation des performances du système.

L'écart souvent remarqué entre la pertinence système et la pertinence utilisateur peut être dû à la subjectivité de la pertinence utilisateur. Pour cela nous voulons proposer une approche personnalisée et interactive de reformulation de requête qui repose sur deux volets :

- La construction et l'exploitation d'un profil d'utilisateur, à partir duquel on peut suggérer une liste de reformulations de la requête initiale de l'utilisateur. L'objectif ici est de ne pas négliger les intérêts et les préférences de l'utilisateur, et d'assurer que les éléments (termes ou tags) ajoutés à la requête

sont à la fois pertinents pour l'utilisateur et en relation avec le sujet de la requête.

- L'implication de l'utilisateur en lui donnant la possibilité de participer à la reformulation de la requête en choisissant l'une des variantes proposées. L'objectif là, est de donner la décision finale à l'utilisateur pour ne pas négliger l'aspect subjectif de la pertinence utilisateur.

Ce chapitre est structuré de la manière suivante : nous commençons tout d'abord dans la section 2, par une revue sur la reformulation de requête dans les approches de RIP**.** Ensuite, nous présentons le profil utilisateur que nous avons proposé ainsi que ses composantes. Puis, nous détaillons le processus de recherche dans la section 4. Nous poursuivons avec l'évaluation de notre approche dans la section 5, et enfin, nous concluons avec la section 6.

2 Revue sur la reformulation de requête dans les approches de RIP :

2.1. Approche RAFA et KECHID : Dans (Rafa & Kechid, 2018), ils ont construit un profil utilisateur géo-social sémantique qui se compose d'une base de situations de recherche (B_RS) qui enregistre l'historique de recherche pertinent de l'utilisateur, En regroupant les recherches effectuées dans différents lieux. Chaque situation de recherche (RS) comprend trois composantes : situationnelle, thématique et sociale. Les poids des termes et des tags sont calculés avec la formule tf * idf du modèle vectoriel (Salton & Buckley, 1988) comme suite :

$$W(t) = \left(0.5 + 0.5 * \frac{tf}{MAXtf}\right) * log(\frac{N}{n})$$

Avec :

- *tf* : fréquence totale de terme *t* dans les documents de l'historique.

- *Max tf* : la plus grande fréquence des termes dans les documents de l'historique.

- *N* : nombre de documents de l'historique.

- *n* : nombre de documents contenant le terme *t* de l'historique.

De même formule ont calculé les tags.

Et **(ii) l'espace de représentation sémantique de l'historique de recherche pertinent**, ils ont classé les termes en trois bases : la sémantique thématique, la sémantique sociale.

La sémantique thématique (lien terme-terme) établit des liens entre chaque terme pertinent dans l'historique utilisateur, basée sur leur cooccurrence fréquente, et est enregistrée dans une base (B_Sem_Them) et est exprimé par une similarité notée sim_them (ti, tj) et calculée comme suit :

$$SimThem(ti, tj) = \frac{|d(ti, tj)|}{|d(t_i)|^2 + |d(t_j) - d(t_{i,}t_j)|^2}$$

Avec :

– |d (ti, tj) | : nombre de documents contenant ti et tj.

– |d(ti)| : nombre de documents contenant ti.

– $|d(t_j) - d(t_{i,}t_j)|^2$: nombre de documents contenant tj et ne contenant pas ti.

La sémantique sociale (lien terme-tag) établit des liens entre chaque terme pertinent de l'utilisateur et les tags les plus utilisés pour annoter les documents contenant ce terme, et est enregistrée dans une base (B_Sem_Scl), et est exprimé par une similitude notée sim_scl (ti, gj), et est calculé comme suit :

$$SimThem(ti, gj) = \frac{|d(ti, gj)|}{|d(t_i)|^2 + |d(g_j) - d(t_{i,}g_j)|^2}$$

Avec :

- |d (ti, gj) | : nombre de documents contenant ti et annotés par le tag gj.
- |d (ti)| : nombre de documents contenant ti.
- $|d(gj) - d\ (ti, gj)|^2$: nombre de documents annotés par le tag gj et ne contenant pas ti.

Pour la reformulation de la requête, la requête est reformulée selon la base sélectionnée, ils ont choisi la base de profil la plus appropriée à la requête de l'utilisateur parmi les deux options (B-Sem-Them, B-Sem-Scl), le système est calculé un score de similarité entre la requête de l'utilisateur Q et chacune de ces bases respectivement comme suit :

$$Sim(B - Sem - Them, Q) = \sum_{ti \in Q} \sum_{t_j B-Sem-Them} sim - them(t_i, t_j)$$

$$Sim(B - Sem - Scl, Q) = \sum_{ti \in Q} \sum_{g_j B-Sem-Scl} sim - scl(t_i, g_j)$$

La reformulation de la requête utilisateur vise à élargir le champ de recherche tout en conservant la pertinence en ajoutant à la requête initiale Q des termes, tags selon la

base de profil sélectionnée. Ainsi, la nouvelle requête reformulée Q' est obtenue sur la base de la requête initiale Q de la manière suivante :

$$Q' = Q + P$$

Avec :

B-Sem-Them : P est le terme qui maximise la similarité ($SimThem(ti, tj)$) avec les termes de Q.

B-Sem-Scl :P est le tag qui maximise la similarité ($SimScl(ti, gj)$) avec les termes de Q.

2 .2*Approche KECHID* **:** Dans (Kechid, 2009), le profil d'utilisateur est composé de :**(i) données personnelles** telles que le nom et l'adresse, qui sont utilisées pour définir son identité. **(ii)L'historique des sources sélectionnées** permet de quantifier la pertinence des sources en fonction du nombre de fois qu'elles ont été sélectionnées. De plus, **(iii) l'historique des recherches** comprend les documents jugés pertinents par l'utilisateur, utilisés pour extraire ses centres d'intérêt, pondérés selon la formule de TF*IDF du modèle vectoriel (Salton & Allan, 1994). La collection de calcul est constituée par l'ensemble des k premiers documents retournés par chaque source dans l'historique. Chaque terme ti est indexé et pondéré suivant la formule en excluant les mots vides :

$$Poids_{id} = \frac{Freq_{id}}{\sqrt{\sum_{j=1}^{N} Freq_{ij} * log^2(\frac{N}{n_i})}} * \log(\frac{N}{n_i})$$

Avec, $Freq_{id}$ est la fréquence d'apparition du terme ti dans le document d, $Freq_{ij}$ est lafréquence d'apparition du terme ti dans le document j, N est le nombre de documents de l'historique,n_i est le nombre de documents de l'historique contenant le termet_i.

Ces **(iv) centres d'intérêt**, dérivés des documents consultés, servent à évaluer la pertinence des sources pour l'utilisateur.

Dans le processus de reformulation de la requête, le centre d'intérêts de l'utilisateur, représenté par un ensemble de termes pondérés, est crucial. Les X termes les plus élevés sont extraites de cet ensemble pour former le vecteur Pu. Ensuite, la requête utilisateur est enrichie en ajoutant les termes deP_u. Cette

nouvelle requête est pondérée par la formule suivante inspirée de (Rocchio, 1971) :

$$q^{nouvelle} = a.q^{ancienne} + \frac{b}{|P_u|}\sum_{t_i \epsilon P_u} t_i$$

Avec :$q^{nouvelle}$ est la requête reformulée, $q^{ancienne}$ est l'ancienne requête formulée par l'utilisateur, a et b sont des constantes empiriques, a, b ∈ [0,1].

***2 .3 Approche BOUHINI* : Dans** (Bouhini C. , 2014), Le contexte informationnel social de l'utilisateur est composé du profil social de l'utilisateur et du profil de son voisinage social.

Le profil social de l'utilisateur est construit à partir de ses activités d'annotations et le profil du voisinage social est construit à partir des annotations des amis de l'utilisateur « *les utilisateurs qui sont en relation social avec lui* ».

Le profil social "PS(u)" de l'utilisateur u est représenté par les termes de ses annotations :

$$\boldsymbol{PS(u) = \{t \in T_z, a_z, =< T_z, d, u >, a_z \in A(u)\}}$$

Où :

$\boldsymbol{T_z}$: est le sous-ensemble de termes qui apparaissent dans une annotation spécifique $\boldsymbol{a_z}$faite par l'utilisateur u et **A(u)** : est l'ensemble de toutes les annotations faites par l'utilisateur u.

Le profil du voisinage social "*PV* (*u*)"de l'utilisateur *u* peut être généré à partir des annotations des utilisateurs qui sont en relation sociale avec lui (amis).

$$\boldsymbol{PV(u) = \cup_{u' \in U/(u,u') \in Rel} PS(u')}$$

Où : *Rel* représente l'ensemble des liens sociaux entre les utilisateurs.

Pour la reformulation de la requête, les auteurs trouvent intéressant de tenir compte des niveaux de saturation, et proposent donc dans leur approche différentes variantes des modèles de RSPI combinant la requête et le contexte social de l'utilisateur

($BM25S_{FreqComb}$et$BM25S_{ScoreComb}$). Correspondant aux trois principaux niveaux de saturation des termes.

- Trois principaux niveaux de saturation des termes :
  - *k*3 =0 : cas ***binaire (bin)***. Il permet de négliger la valeur des *tf* des termes dans la partie requête. On parle dans ce cas d'une **saturation maximale** des termes au sein de la requête

$$QTF_{q,t} = \frac{(k_3 + 1) \times tf_{q,t}}{k_3 + tf_{q,t}}$$

*K*3=1000 cas ***fréquentiel (tf)***. Fixer la valeur du *k*3 à 1000 ou à une grande valeur revient quasiment à prendre en compte directement la valeur du nombre d'occurrence du terme *t* dans la requête *q* (*tfq*, *t*). Dans ce cas la saturation est nulle.

$$QTF_{q,t} \approx tf_{q,t}$$

*K*3 normalisé (ou fixé à 8) cas ***normalisé (w)*** car l'ajustement des valeurs du *k*3 pour trouver une valeur optimale de celui-ci ou le fait de le fixer à 8, permet d'ajuster l'effet de saturation des termes de la requête.

Afin de tenir compte de ces trois façons de contrôler la saturation de la requête via le paramètre *k*3, nous proposons trois modèles de recherche sociale personnalisée d'information,

- **Modèle de RSPI** : ***BM25S (d, u)* :** Dans ce modèle, ils considèrent que le contexte social de l'utilisateur peut être utilisé pour remplacer complètement la requête de l'utilisateur. Ce modèle produit une liste classée de documents pertinents pour un utilisateur u en prenant en compte son contexte informationnel social.

$$\boldsymbol{BM25S(d,u)} = \sum_{\boldsymbol{t \in d \cap CIS(u)}} \boldsymbol{w_{d,t} \times w_{u,t}}$$

Où :

$$\mathrm{w_{d,t}} = \mathrm{TF_{d,t}} \times \mathrm{IDF_t}$$

$$\mathrm{w_{u,t}} = \mathrm{UTF_{u,t}}$$

$$\mathrm{UTF_{u,t}} = \frac{(\mathrm{k_3} + 1) \times \left[\mathrm{w_u} \times \mathrm{tf_{u,t}} + \mathrm{w_v} \times \mathrm{tf_{v_u,t}}\right]}{\mathrm{k_3} + \left[\mathrm{w_u} \times \mathrm{tf_{u,t}} + \mathrm{w_v} \times \mathrm{tf_{v_u,t}}\right]}$$

$$\boldsymbol{BM25S(d,u) = \sum_{t \in d \cap u} TF_{d,t} \times IDF_t \times UTF_{u,t}}$$

$\boldsymbol{TF_{d,t}}$ et $\boldsymbol{IDF_t}$représentent respectivement $\boldsymbol{TF_{d,t}}$et $\boldsymbol{IDF_t}$classiques dans la formule BM25

$UTF_{u,t}$ C'est la fréquence ajustée du terme (t) dans un document (ou un contexte) spécifique (u). Cette fréquence ajustée tient compte des poids attribués aux termes provenant de différents contextes

K3 C'est un paramètre de contrôle qui ajuste l'impact de l'ajustement de fréquence sur le résultat final. Plus la valeur de k3 est élevée, plus l'ajustement aura d'effet.

- **Modèle de RSPI $\boldsymbol{BM25S_{FreqComb}}$ (*d, q, u*) :** ce modèle retourne une liste classée de documents pertinents pour un utilisateur *u* en considérant sa requête combinée à son contexte informationnel social au niveau des fréquences d'occurrence des termes.

$$\boldsymbol{BM25S_{FreqCombn}(\sum_{t \in d \cap q} TF_{d,t} \times IDF_t \times QTF_{S_{q,u,t}})}$$

Où :

$$\boldsymbol{QTF_{S_{q,u,t}} = \frac{(k_3 + 1) \times [tf_{q,t} + w_u \times tf_{u,t} + w_v \times tf_{v_{u,t}}]}{k_3 + [tf_{q,t} + w_u \times tf_{u,t} + w_v \times tf_{v_{u,t}}]}}$$

$\boldsymbol{QTF_{S_{q,u,t}}}$C'est la fréquence du terme (t) dans la requête (q) après ajustement. Cela signifie que chaque terme dans la requête peut être ajusté en fonction de sa fréquence dans la requête elle-même ainsi que d'autres facteurs.

- **ModèledeRSPI$\boldsymbol{BM25S_{ScoreComb}}$(*d, q, u*)** : ce modèle retourne une liste classée de documents pertinents pour un utilisateur *u* en combinant la requête à son contexte informationnel social au niveau des scores

$$\boldsymbol{BM25S_{ScoreComb}(d,q,u) = RSV(d,q) + w_u \times BM25S(d,q)}$$

2 .4*Approche SAOUD et KECHID* **:** Dans (Saoud & Kechid, 2016) deux types de profils sont abordés : le profil de l'utilisateur et le profil du document.

Profil de l'utilisateur : Le profil de l'utilisateur est défini par l'ensemble des tags utilisés par l'utilisateur comme (Noll & Meinel, 2007).les tags les plus fréquemment utilisées sont plus pertinentes et significatives pour décrire les intérêts de l'utilisateur.

$$\text{Profil}(\text{U}_\text{m}) = \left\{\left(t_i, tf_{u_m}(t_i)\right) \middle| tf_{u_m}(t_i) \geq avgtags_{u_m} and i \in [1..L]\right\}$$

Où :

- L : est le nombre de tags utilisés par l'utilisateuru_m.
- $t_i, tf_{u_m}(t_i)$: est la fréquence des tags basée sur l'utilisateur, ce qui signifie combien de fois l'utilisateur $u_m$utilise tagt_i.

$$avgtags_{u_m} = \frac{\sum_{i=1}^{L} tf_{u_m}(t_i)}{tags_number_{u_m}}$$

Où :

$tags_number_{u_m}$: est le nombre de tags utilisées par l'utilisateuru_m.

Profil du document : Afin d'utiliser tous les types d'informations disponibles, les auteurs utilisent à la fois les informations sociales et les informations textuelles pour construire le profil du document. Ensuite, pour le documentd_n, son profil de document est défini par son ensemble de tags$tags_{dn}$et son ensemble de termes$terms_{dn}$, de la manière suivante :

$$profile(d_n) = tags_{dn} \cup terms_{dn}$$

Où :

$tags_{dn}$: représente l'ensemble des tags utilisées pour annoter le documentdn. Qui est défini comme suit :

$$tags_{dn} = \{tag_i, tf_{dn}(tag_i) | \ i \in [1..P]\}$$

$terms_{dn}$: représente l'ensemble des termes qui appartiennent au contenu du document. Qui est défini comme suit :

$terms_{dn}$=$\{term_j, freq_{dn}(term_j) | \ j \in [1..R]\}$.

Où :

- P : le nombre de tags utilisés pour annoter le documentdn.
- R : le nombre de termes qui appartiennent au documentdn.
- $tf_{dn}(tag_i)$: est la fréquence des tags basées sur le document, ce qui signifie combien de fois le document dn est tagué par le tagtag_i.
- $tf_{um}(term_j)$: est la fréquence du terme $term_j$ dans le documentdn.

Selon(Smyth, 2011), 66 % de nos recherches Web sont similaires à celles effectuées par nos collègues. Sur la base de cette idée, pour adaptée le profil ils préfèrent trouvent les utilisateurs ayant des intérêts similaires (amis proches) pour l'utilisateur actuel basé sur la requête initiale, pour trouver les listes de tags similaires pour chaque terme de requête. Pour trouver ces listes de tags similaires, ils utilisent la mesure d'amitié et la mesure d'expansion de tags inspirés du travail de(Bender, et al., 2008). Pour obtenir le profil utilisateur adapté, les auteurs faisant une intersection entre le profil utilisateur original et l'union des listes de balises similaires comme suit :

$$profile_{adapted}(u_m, q) = profil(u_m) \cap \bigcup_{t_{i \in q}} sim(t_i)$$

La requête de l'utilisateur est reformulée en ajoutant les tags similaires de chaque terme de requête, en fonction des amis proches de l'initiateur de la requête. La nouvelle requête est pondérée par la formule suivante inspirée de(Rocchio, 1971) 1971] :

$$q^{new}(u_m) = a.q^{old} + \frac{b}{|\bar{\bar{T}}|}\sum_{i \in T} t'$$

Où :

- q^{new} : Est la requête développée.
- q^{old} : Est l'ancienne requête donnée par l'utilisateur.
- A et b sont des constants, a, b ∈[0,1].
- $\bar{\bar{T}}$: Représente la liste des tags similaires des anciens termes de requête. Cette liste est établie selon la formule proposée suivante :

$$\bar{\bar{T}} = \bigcup_{t_i \in q^{old}} t', t' = \{t' \in sim(t_i) | TagSim_{Total}(t') > avg_{simT}\}$$

Où :

$$T = \bigcup_{t_i \in q^{old}} sim(t_i)$$

- $sim(t_i)$: représente la liste des tags qui sont similaires au terme de requêtet$_i$.
- $TagSim_{Total}(t')$ *:* représente la somme des similitudes de tag t' dans toutes les listes de tags similaires

Pour calculer le score d'amitié moyen, nous utilisons la formule proposée suivante :

$$TagSim_{Total}(t') = \sum_{\substack{t_i \in q^{old} \\ t' \in sim(t_i)}} TagSim(t', t_i)$$

Où :

- $TagSim(t', t_i)$: représentela similarité entre deux tags t,t_i.
- avg_{simT} : représente la similarité moyenne dans la liste T, qui se calcule comme suit :

$$avg_{sim_T} = \frac{\sum_{i \in T} TagSim_{Total}(t')}{|T|}$$

3 Profil d'utilisateur

Selon (Kechid, 2009),un profil peut être défini comme un modèle personnalisé d'accès à l'information, alors qu'une requête est l'expression d'un besoin circonstancié que l'utilisateur souhaite voir en tenant compte de son profil.

Le profil d'utilisateur que nous proposons dans cette contribution est un profil qui combine l'aspect thématique et l'aspect social des intérêts de l'utilisateur : ***l'aspect thématique*** représente l'ensemble des documents jugés pertinents par l'utilisateur, ainsi que les termes pertinents extraits à partir de ses documents, et ***l'aspect social*** représente l'ensemble des tags les plus utilisés par l'utilisateur pour annoter les documents pertinents.

Le profil proposé se compose de :

3.1 Données personnelles

Cette composante représente la partie statique de notre profil. Ces données sont utilisées pour établir l'identité de l'utilisateur. Ces informations peuvent inclure : nom,

prénom, mot de passe, E-mail, âge, sexe, langue, adresse et domaine d'intérêts initial (Figure 3.1).

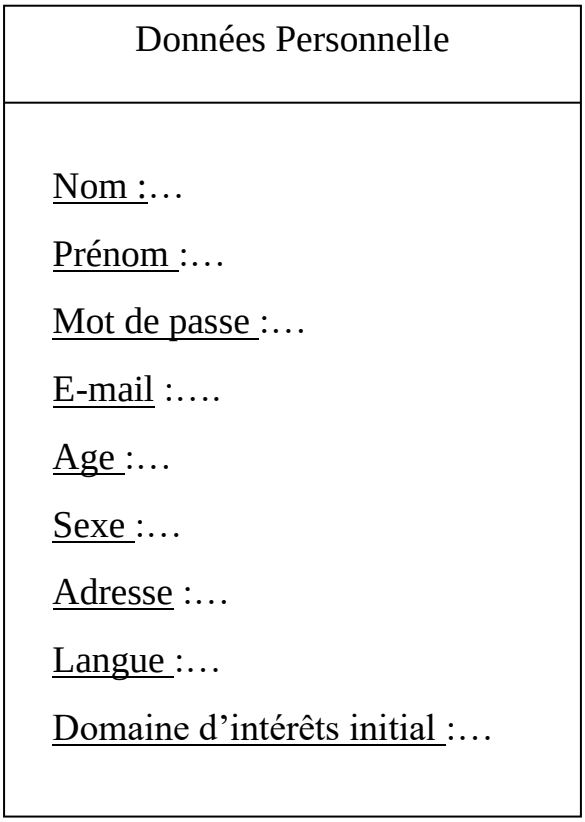


Figure 3.1:*Modèle de données personnelles.*

3.2 Historique de recherche pertinent

Cette composante sert à sauvegarder l'historique pertinent de l'utilisateur et elle comporte comme indique-la (Figure 3.2), les deux sous-ensembles suivants :

3.2.1 Les termes pertinents

Les termes pertinents forment un ensemble de termes les plus importants extraits des documents jugés pertinents par l'utilisateur. Ces termes représentent l'orientation

thématique de l'utilisateur et ils sont pondérés en fonction de leur importance dans l'historique de recherche. Nous calculons la pondération de ces termes selon la formule tf*idf du modèle vectoriel (Salton & Buckley, 1988) suivante :

$$W(t) = \left(0.5 + 0.5 * \frac{tf}{MAXtf}\right) * log(\frac{N}{n}) \qquad (1)$$

Où :

- *tf* : fréquence totale de terme *t* dans les documents de l'historique.
- *Max tf :* la plus grande fréquence des termes dans les documents de l'historique.
- *N :* nombre de documents de l'historique.
- *n :* nombre de documents contenant le terme *t* de l'historique.

3.2.2 Les tags pertinents

L'ensemble de tags les plus utilisés par l'utilisateur pour annoter les documents pertinents de son historique de recherche. Ces tags servent à détecter une pertinence supplémentaire fournis par les utilisateurs pour exprimer leurs opinions, leurs sentiments, leurs impressions, etc.

Les tags sont également pris en compte dans l'historique d'annotation sociale de la même manière que les termes. Ainsi, le poids d'un tag g est calculé selon la méthode suivante :

$$W(g) = \left(0.5 + 0.5 * \frac{gf}{MAXgf}\right) * log(\frac{N}{n}) \qquad (2)$$

Où :

- *gf* : fréquence totale de tag *g* dans les documents de l'historique.
- *Max* gf : la plus grande fréquence totale des tags dans les documents de l'historique.
- *n* : nombre de documents annotés par le tag *g* dans l'historique.
- *N* : nombre de documents dans l'historique.

Historique de Recherche Pertinent
Termes Pertinents : V_t= *{(t1, wt1), (twt2),…}*
Tags Pertinents: V_g= *{(g1,wg1),(g2,wg2),…}*

Figure 3.2:*Modèle de l'historique de recherche pertinent.*

3.3 Base de liens de cooccurrence terme-terme

La collection des termes liés à un domaine d'intérêt reflète un premier niveau pour caractériser un utilisateur et son centre d'intérêts. Nous voulons passer à un niveau suivant d'exploitation des termes pertinents qui consiste à identifier et évaluer les relations de cooccurrence entre les termes. Nous percevons les relations de cooccurrence entre les termes comme une pseudo-sémantique, soulignant la fréquence avec laquelle deux termes pertinents apparaissent ensemble dans les documents pertinents de l'historique de recherche.

Pour nous, l'apparition fréquente de deux termes dans plusieurs documents indique qu'ils décrivent ensemble un sujet commun. Par exemple, l'apparition des deux termes « Virus » et « Ordinateur » peut indiquer qu'il s'agit d'un sujet d'informatique, tandis que l'apparition de « Virus »et « Vaccin » indique un sujet de médecine.

L'évaluation de la relation de cooccurrence entre les termes pertinents est représentée par la similarité thématique, notéesim_cooccur _Them(t_i, t_j), qui représente la liaison entre chaque terme t_i et d'autres termest_j, et est calculée selon la formule suivante :

$$\text{Sim_cooccur _Them}(t_i, t_j) = \frac{|d(t_i,t_j)|}{|d(t_i)|+|d(t_j)-d(t_i,t_j)|} \qquad (3)$$

Avec :

- $|d\ (t_i, t_j)\ |$: nombre de documents contenant $t_i$et $t_j$collectivement.
- $|d\ (t_i)|$: nombre de documents contenant ti.
- $d\ (t_j)$- $d\ (t_i, t_j)\ |$: nombre de documents contenant t_j et ne contenant past_i.

La base de relations de cooccurrence des termes pertinents comme l'affiche la (Figure 3.3) est utilisée pour stocker ces similarités entre les termes pertinents de l'historique du profil utilisateur. Cette base est désignée sous le nom de "Base_cooccur_them".
Notre but, à partir de cette composante, est de mettre en lumière les relations de cooccurrence entre les termes pertinents pour améliorer les résultats de recherche.

Base de cooccurrence Terme-Terme (Base_cooccur_them)

Terme / Terme	t_1	t_2	…	t_k
t_1	-	Sim_cooccur _Them(t_1, t_2)	…	Sim_cooccur _Them(t_1, t_k)
t_2	Sim_cooccur _Them(t_2, t_1)	-	…	Sim_cooccur _Them(t_2, t_k)
…	…	…	-	…
t_k	Sim_cooccur _Them(t_k, t_1)	Sim_cooccur _Them(t_k, t_2)	…	-

Figure 3.3:*Modèle de la base des relations de cooccurrence de terme-terme.*

3.4 La base de liens de cooccurrence terme-tag

Étant donné que les tags d'utilisateurs peuvent également servir à enrichir les requêtes, il est essentiel de ne pas dérouter la recherche en sortant du contexte actuel de l'utilisateur. Ainsi, un tag ne doit être intégré dans une requête que s'il entretient une relation de cooccurrence très significative avec les termes de cette requête.

Pour cette raison, nous allons mettre en valeur les liens de cooccurrence entre les termes pertinents de l'historique de recherche et les différents tags utilisés par l'utilisateur pour annoter les documents pertinents. Nous appliquons la même logique précédente pour mettre en évidence les liens de cooccurrence entre les termes et les tags, en utilisant une similarité sociale, notée Sim_cooccur _scl (t_i, g_j). Cette similarité exprime la relation entre chaque terme pertinent ti et tous les tags pertinentsg_j, et elle est calculée selon la formule suivante :

$$\text{Sim}_{\text{cooccur}_{\text{scl}(t_i,g_j)}} = \frac{|d(t_i,g_j)|}{|d(t_i)|+|d(g_j)-d(t_i,g_j)|} \quad (4)$$

Avec :

- $|d(t_i, g_j)|$: Nombre de documents contenant t_i et annotés par le tagg_j.
- $|d(t_i)|$Nombre de documents contenantt_i.
- $|d(g_j) - d(t_{i,}g_j)$: Nombre de documents annotés par g_j et ne contenant past_i.

La base de relations de cooccurrence terme-tag (Base_cooccur_scl) comme l'affiche (figure 3.4), sert à sauvegarder les liens de similarité entre les termes pertinents et les tags fréquents de l'historique de profil utilisateur.

Base de cooccurrence Terme-Tag (Base_cooccur_scl)

Tag / Terme	$\mathbf{g_1}$	$\mathbf{g_2}$	…	$\mathbf{g_k}$
$\boldsymbol{t_1}$	Sim_cooccur _scl(t_1, g_1)	Sim_cooccur _scl(t_1, g_2)	…	Sim_cooccur _scl(t_1, g_k)
$\boldsymbol{t_2}$	Sim_cooccur _scl(t_2, g_1)	-	…	Sim_cooccur _scl(t_2, g_k)
…	…	…	-	…
$\boldsymbol{t_k}$	Sim_cooccur _scl(t_k, g_1)	Sim_cooccur _scl(t_k, g_2)	…	…

Figure3.4: *Modèle de la base des relations de cooccurrence de terme-tag.*

4 Processus de Recherche

Notre processus de recherche (Figure 3.5) comprend les étapes suivantes :

- Le système reçoit la requête utilisateur Q.
- Le système consulte le profil utilisateur pour lui proposer plusieurs variantes de reformulation de la requête initiale.
- Après la réponse de l'utilisateur, le système envoi la variante choisie en recherche.
- Le système met à la disposition de l'utilisateur les résultats de recherche.
- Le système observe les réactions et les comportements de l'utilisateur (choix de variante de reformulation, documents pertinents dans les 5 tops) pour mettre à jour son profil.

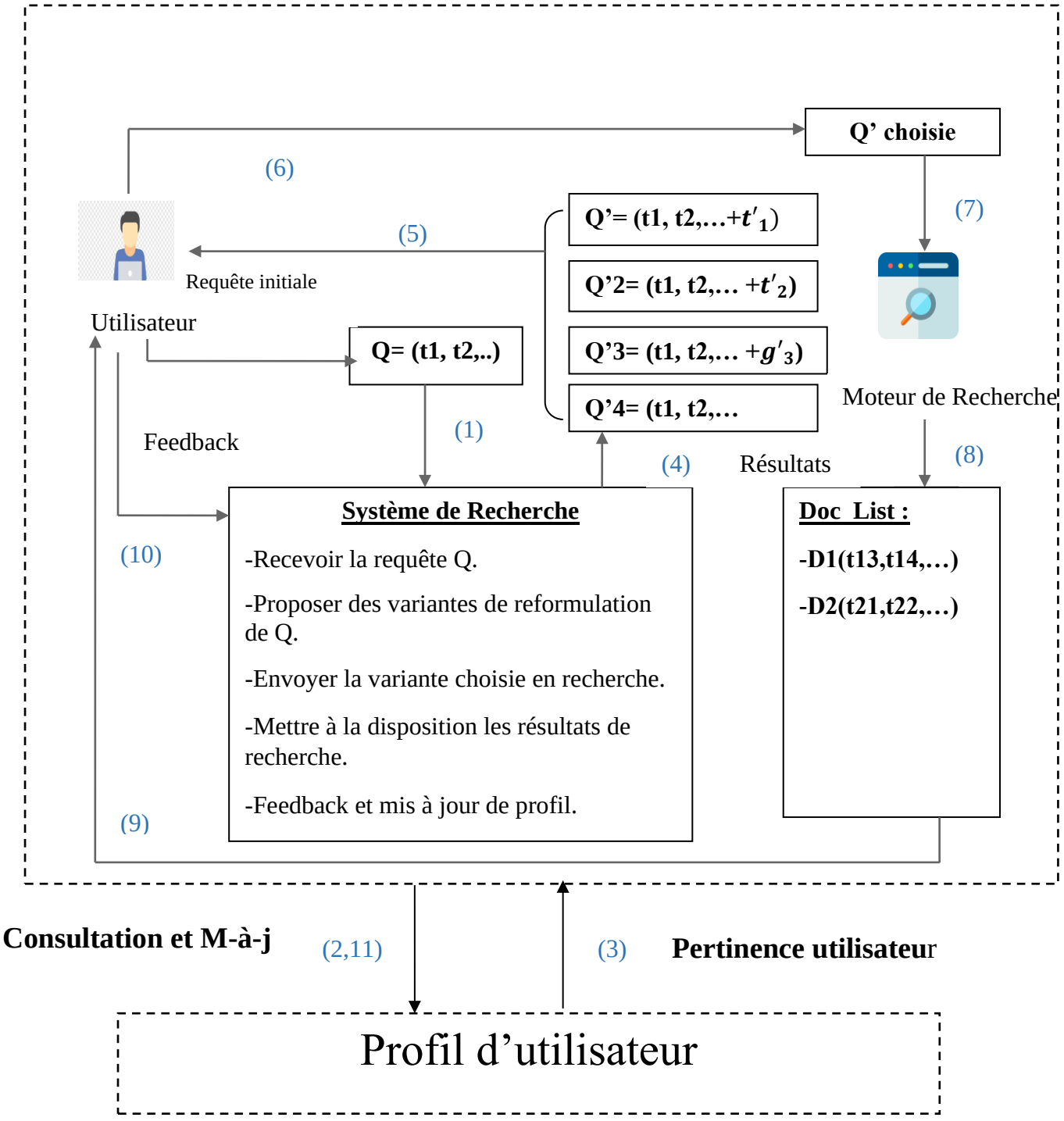


Figure 3.5: *Processus de Recherche.*

Dans ce qui suit, nous présentons les différentes phases de notre processus de recherche.

4.1 La reformulation de requête

L'objectif de reformulation de requête par réinjection de pertinence est de produire une nouvelle requête mieux adaptée aux besoins en information de l'utilisateur. La

reformulation de requête implique l'ajout de nouveaux termes à une requête initiale ou la modification des poids des termes existants afin d'orienter la recherche vers les documents pertinents. La génération de la nouvelle requête peut se faire de manière automatique ou interactive.

Notre objectif dans cette phase de reformulation de requête est de remédier aux problèmes de désorientation de recherche dus à des reformulations abusives des requêtes initiales de l'utilisateur. Nous voulons assurer deux points :

- Les éléments ajoutés à la requête soient pertinents et en relation avec le sujet de la requête.
- La prise en compte de l'avis de l'utilisateur.

L'opération de reformulation de requête se fait en deux étapes, la proposition de plusieurs variantes de reformulation possible de la requête, et le choix final de la reformulation à envoyer à la recherche est faite par l'utilisateur lui-même.

4.1.1 Proposition des variantes de reformulation

Nous proposons quatre variantes de reformulation de la requête originale Q, comme suit :

***Variante de reformulation N 1* :**

$$Q' = Q + t'_1 \qquad (05)$$

Avec :

- Q : la requête initiale.

-Q' : la nouvelle requête reformulée.

-t'_1: Leterme t qui maximise la similarité (3) avec tous les termes de requête Q.

Variante de reformulation N 2 :

$$Q'_2 = Q + t'_2 \qquad (06)$$

Avec :

-Q : la requête initiale.

-Q' : la nouvelle requête reformulée.

$-t'_2$: Le terme t qui maximise la similarité (3) avec le terme le plus important (selon le poids) de requête Q.

Variante de reformulation N 3 :

$$Q'_3 = Q + g'_1 \qquad (07)$$

Avec :

- Q : la requête initiale.

-Q' : la nouvelle requête reformulée.

-g'_1: Letag g qui maximise la similarité (4) avec tous les termes de requête Q.

Variante de reformulation N 4 :

$$Q'_4 = Q + g'_2 \qquad (08)$$

Avec :

-Q : la requête initiale.

-Q' : la nouvelle requête reformulée.

$-g'_2$: Le tag g qui maximise la similarité (4) avec le terme le plus important (selon le poids) de requête Q.

4.1.2 Pondération et classement des variantes

L'utilisateur est confronté à quatre variantes de reformulation de sa requête. Le choix de l'utilisateur parmi ces variantes revêt une grande importance car il influence directement la direction de sa recherche ultérieure. Cependant, il est crucial de déterminer quelles variantes doivent être présentées en priorité à l'utilisateur. Pour cela, nous établissons une pondération pour classer ces variantes. Ce poids est calculé en prenant compte du choix de l'utilisateur des quatre variantes et le nombre de documents pertinents retournés parmi les cinq premiers résultats de la recherche. La formule suivante est utilisée pour déterminer ce poids :

$$Poid_var(V_i) = \frac{Nbr_choix(V_i) + Nb_Doc_Pert(V_i)}{(Nb_Doc_Pert(V_i)+1)*|Q|} \qquad (09)$$

Avec :

-$Nbr_choix(V_i)$:Le nombre total de combien de fois l'utilisateur à choisi la variante (V_i).

-$Nb_Doc_Pert(V_i)$: le nombre total de documents pertinents retrouvés dans les 5 top documents retournés chaque fois l'utilisateur à choisi la variante(V_i).

-$|Q|$: le nombre total de requêtes dans l'historique de recherche.

Nous notons que cette mesure, le Poids_var(V_i), est initialisée à zéro pour les quatre variantes, ce qui implique que les requêtes reformulées sont classées de manière aléatoire au début de la recherche. Cependant, en observant les choix de l'utilisateur, nous constatons que ce critère évolue au fil du temps. Ainsi, les requêtes reformulées sont classées en fonction du score le plus élevé à un moment donné. Par exemple, si l'utilisateur opte pour la varianteQ'_3, le poids de cette variante devient plus élevé que celui des autres variantes. Tandis que les scores des autres variantes sont diminués. Logiquement pour la prochaine recherche la variante Q'_3sera classée en premier.

4.1.3 Intervention de l'utilisateur

Il est essentiel de prendre en compte l'avis de l'utilisateur lors de la reformulation de la requête. Parmi les quatre variantes de reformulation proposées, l'utilisateur sélectionne celle qu'il estime être la plus pertinente pour ses besoins de recherche. Ce choix revêt une importance capitale car il détermine quelle reformulation sera employée pour la recherche suivante.

4.1.4 Mis à jour des poids des variantes

Apres le choix d'une variante de reformulation par l'utilisateur, cette variante est envoyée à la recherche. Ensuite le système observe la réaction (jugement de pertinence) de l'utilisateur avec les résultats de recherche retournés. Le nombre de documents jugés pertinents par l'utilisateur dans les cinq premiers documents retournés est exploité pour mettre à jour les poids des quatre variantes selon la formule (09).

5 Evaluation de l'approche

D'après (Toms, Agosti, Fuhr, & Pe, 2013) et (Khazri, Tmar, Abid, & Boughanem, 2008), la notion de pertinence dans le contexte de la recherche personnalisée d'informations est hautement subjective, car elle est étroitement liée aux préférences de l'utilisateur. Seul cet utilisateur est en mesure d'évaluer cette pertinence selon son point de vu. Cette réalité fait de l'évaluation d'une approche de recherche d'information personnalisée une tache très problématique. Parce qu'il n'y a pas de collection de tests standards permettant à l'utilisateur d'évaluer ses propres performances. Une collection de tests comportant des interrogations et des pré-jugements de pertinence réalisées par un expert du domaine exprime la pertinence des

documents pour cet expert lui-même, mais pas nécessairement pour les autres utilisateurs.

5.1 Stratégie d'évaluation

Dans l'absence d'un cadre d'évaluation dédié à un contexte de personnalisation déjà indiqué et mentionné dans plusieurs approches PIR ((Bouidghaghen, 2011), (Panjarkar, Shete, & Kahate, 2014), …), où chaque approche développe son propre cadre d'évaluation, Nous conduirons notre phase d'évaluation en suivant les étapes ci-après :

- ***Construction des profils des utilisateurs :*** Nous avons créé 10 profils pour 10 utilisateurs provenant de différents domaines professionnels, afin d'avoir une diversité d'intérêts. Chaque profil a été construit en collaboration avec l'utilisateur concerné. Les termes et les tags pertinents sont pondérés selon les formules proposées, et les liens de cooccurrence ont été calculés et sauvegardés dans les bases Base_cooccur_them et Base_cooccur_scl.
- ***Besoins en information et requêtes* :** Chaque utilisateur formule trois requêtes correspondant à trois besoins en information. Quelques requêtes expriment des besoins d'information spécifiques pour l'utilisateur, et quelques autres requêtes sont communes entre les utilisateurs.

  Nous utilisons *Google* comme moteur de recherche.
- ***Jugement de pertinence :*** Les utilisateurs évaluent manuellement la pertinence des documents. Chaque document est jugé comme étant soit non pertinent **(0)**, soit pertinent **(1)**.
- ***Evaluation*** : l'évaluation consiste à calculer le facteur de précision (le nombre de document pertinents) dans les 5, les 10 puis les 15 premiers documents retournés.
- ***Comparaison*** : Afin de démontrer les avantages de notre approche par rapport à d'autres approches de recherche, nous calculons la précision selon les 3 scénarios suivants : (A) requête envoyée sans reformulation, (B) requête avec reformulation automatique personnalisée, (C) requête reformulée avec notre approche de reformulation interactive personnalisée.

5.2 Expérimentation

L'expérimentation implique l'exécution des requêtes des utilisateurs sur notre modèle de recherche selon les trois scénarios susmentionnés. Notre objectif est de démontrer

que nos variantes de reformulation sont plus performantes que les autres méthodes de reformulation.

5.2.1 Construction des profils et requêtes :

Le tableau (Tableau3.1) suivant, présente les utilisateurs, leurs centres d'intérêts et les requêtes à envoyer à la recherche :

Utilisateur	**Centres d'intérêts**	**Requête**
U_1: "***Etudianten informatique***"	**Termes :** (réseau 0.521), (java 0,5), (sécurité 0.108), (informatique 0.751), (application 0.654), (neurone 0.155), (script 0.641), (swing 0.279), (internet 0.256)**…** **Tags :**(cryptographie 0.274), (couche de neurone 0.199), (micro édition 0.113), (cyber sécurité 0.056) **…**	• Réseau • Java • Sécurité
U_2: "***Etudianten informatique***"	**Termes :** (mac 0.412), (application0.649), (développement 0.466), (linux 0.602), (androïde 0.117), (web 0.588), (site 0.317), (Système 0.445), (exploitation 0.488), (mobile 0,165) **…** **Tags : (**interface utilisateur 0.202), (sémaphore 0,089), (machine virtuel 0.028), (mise à jour 0,056), (création site web 0,455)**…**	• Application • Système d'exploitation • Site web
U_3: "**Intéresseau médecine**"	**Termes :** (consultation 0,43)**,** (génique 0,244), (thérapie 0,213), (pdf 0,098), (médicale 0,687), (santé 0,609), (sang 0,789), (fibrinogène0,376)**…** **Tags :** (suivi médicale 0.224), (traitement génique 0.309), (visite médicale 0.256), (maladies géniques 0.277), (production de dégradation de fibrinogène 0.398)**…**	• Thérapie • Consultation • Pdf

U_4: "**Intéresseau médecine**"	**Termes :** (chirurgie 0.231), (urinaire 0,143), (menstruel 0,584), (pédiatrique 0.227), (générale 0.014), (traitement 0.163), (cancer 0,588), (diurne 0.063), (cycle 0,07), (ovarien 0.198), (nocturne 0,145), (analyse 0,243) **…** **Tags : (**femme enceinte 0.523), (menstruation0.623), (abdominale 0.223), (ovulation 0,322), (traitement de cancer 0,211)**…**	• Cycle • chirurgie • Analyse urinaire
U_5: "**Intéresseau cuisine**"	**Termes : (**plat 0.502), (cuisine 0.645), (algérienne 0.082), (repas 0.533), (légumes 0.109), (Pâtisserie 0.132) (méditerranéenne 0,651) **…** **Tags :**(plat tunisienne 0.599), (tartelettes 0.299), (quiche salé 0.249), (plats estivaux 0.199)**…**	• Cuisine • Fond • Cuisine méditerranéenne
U_6: "**Intéresseaucuisine**"	**Termes :**(Recettes 0.251), (cuisine0.196),(décore 0.391), (maison 0.309), (typique 0.460), (gâteau 0.420) , (japonaise0.391), (chocolat 0.410),(anniversaire 0.450) ,(pizza 0.237),(italienne 0.243), (sushi 0.383)**…** **Tags :** (CuisineItalienne0.360), (Desserts 0.085), (Cuisine Végétarienne 0.075)**…**	• Décorer gâteau • Recette pizza • Cuisine japonaise
U_7: " **Intéresseau tourisme**"	**Termes :** (hôtels 0.022),(voyage 0.210), (écotourisme 0.287),(Guide 0.171),(touriste 0.165), (historique 0.056),(Forest 0.255), (Amazon 0.266), (bon plan 0.224), paris (0.155), **…** **Tags :**	• Guide touristique • Bon plan • Amazon

	(france0.159),(vacances 0.025), (Aventure safari 0.014), (Vacances en famille 0.017),**…**	
U_8: "**Intéresseau tourisme**"	**Termes :** (Méditerranée 0.201), (historique 0.315), (voyages 0.193), (destination 0.424), (bateau 0.220), (tourisme 0.12), (monuments 0.555), (croisière 0.27), (ile 0.395), (vacances 0.055), (visite 0.335), (touristique 0.21), (rêve 0.401) **…** **Tags :** (historiques Algérie0.560), (croisières Algérie 0.234), (croisières Méditerranée 0.471)**…**	• Voyages en bateau • Destination de rêve • Visiter les monuments
U_9: "**Psychiatre**"	**Termes :** (compréhension 0,611), (humaines 0.523), (psychologie 0.517), (émotionnelle 0.627), (neuroscience0.419),symptoms(0,378) , (psychiatrie 0.410) **…** **Tags :**(psycholoy humain 0.365), (psychoéducation 0.540), (Relaxation 0.034), (Santé mental 0.157)**…**	• Compréhension émotionnelle • Perception • Neuroscience
U_{10}: "**Psychiatre**"	**Termes :** (conflit 0.420), (sensorielle 0.229), (thérapie 0.387), (super ego0.516), (symptoms 0.324), (freud 0.483),(intérieur 0.398), (traumatique 0.330), (psychanalyse0.519), (ACT 0,410), (stress 0.334)**…** **Tags :(**Sensations0.120),(thérapie techniques 0.412), (Psychologie 0.332) (Sigmund freud's 0.513)	• Conflit • Stress • ACT

Tableau 3.1*: Profils d'utilisateurs.*

5.2.2 Résultats et discussion

Dans une première partie, nous présentons

Dans le tableau suivant (Tableau 3.2), nous présentons les résultats expérimentaux obtenus concernant la précision dans les 5, 10 et 15 documents retournés, pour chaque requête, chaque utilisateur et chaque scénario.

	Sans reformulation de requête			Avec reformulation automatique			Avec reformulation interactive		
User1	**P@5**	**P@10**	**P@15**	**P@5**	**P@10**	**P@15**	**P@5**	**P@10**	**P@15**
Réseau	0	0	0	0	0	0	1	0,9	0,8
Java	0	0	0	0	0	0	1	1	1
Sécurité	0	0,1	0,066	1	1	0,933	1	1	0,933
Moyen_ Précision	**0**	**0,033**	**0,022**	**0,333**	**0,333**	**0,311**	**1**	**0,966**	**0,911**
User2	**P@5**	**P@10**	**P@15**	**P@5**	**P@10**	**P@15**	**P@5**	**P@10**	**P@15**
Application	0	0	0	1	0,9	0,866	1	0,9	0,866
Système d'exploitation	0	0,1	0,066	0	0	0,133	1	0,9	0,866
Site web	0	0	0	0	0	0	0,8	0,7	0,666
Moyen_ Précision	**0**	**0,033**	**0,022**	**0,333**	**0,3**	**0,333**	**0,933**	**0,833**	**0,799**
User3	**P@5**	**P@10**	**P@15**	**P@5**	**P@10**	**P@15**	**P@5**	**P@10**	**P@15**
Thérapie	0	0,3	0,266	0,6	0,7	0,6	1	1	0,866
Consultation	0	0,1	0,066	0,8	0,8	0,866	0,8	0,8	0,866

PDF	0	0	0	0,4	0,2	0,533	0,4	0,5	0,533
Moyen_ Précision	**0**	**0,133**	**0,11**	**0,6**	**0,566**	**0,666**	**0,733**	**0,766**	**0,755**
User4	**P@5**	**P@10**	**P@15**	**P@5**	**P@10**	**P@15**	**P@5**	**P@10**	**P@15**
Cycle	0,2	0,4	0,4	0,6	0,5	0,466	1	1	1
Chirurgie	0,4	0,3	0,2	0,6	0,6	0,533	1	0,9	0,8
Analyse urinaire	0	0	0	0,2	0,1	0,066	0,8	0,6	0,466
Moyen_ Précision	**0,2**	**0,233**	**0,2**	**0,466**	**0,4**	**0,355**	**0,933**	**0,833**	**0,755**
User5	**P@5**	**P@10**	**P@15**	**P@5**	**P@10**	**P@15**	**P@5**	**P@10**	**P@15**
Cuisine	0	0	0	0	0	0	1	1	1
Fond	0	0	0	0	0	0	1	1	0,866
Cuisine méditerranéenne	0,2	0,2	0,333	0,4	0,5	0,533	0,8	0,9	0,866
Moyen_ Précision	**0,066**	**0,066**	**0,111**	**0,133**	**0,166**	**0,177**	**0,933**	**0,966**	**0,91**
User6	**P@5**	**P@10**	**P@15**	**P@5**	**P@10**	**P@15**	**P@5**	**P@10**	**P@15**
Décorer gâteau	0.4	0.4	0.266	0.6	0.5	0.466	0.8	0.9	0.8
Recette pizza	0.4	0.3	0.266	0.6	0.3	0.333	0.8	0.7	0.667
Cuisine japonaise	0.4	0.5	0.466	0.8	0.6	0.533	0.8	0.8	0.8
Moyen_ Précision	**0.4**	**0.4**	**0.332**	**0.733**	**0.466**	**0.444**	**0.8**	**0.8**	**0.755**
User7	**P@5**	**P@10**	**P@15**	**P@5**	**P@10**	**P@15**	**P@5**	**P@10**	**P@15**
Guide touristique	0.2	0.2	0.133	0.6	0.7	0.8	0.8	0.8	0.866
Bon plan	0.4	0.3	0.2	1	1	1	1	1	1

Amazon	0	0	0	0.6	0.6	0.6	0.8	0.7	0.8
Moyen_ Précision	**0.2**	**0.166**	**0.111**	**0.733**	**0.766**	**0.8**	**0.866**	**0.833**	**0.888**
User8	**P@5**	**P@10**	**P@15**	**P@5**	**P@10**	**P@15**	**P@5**	**P@10**	**P@15**
Voyages en bateau	0.6	0.5	0.533	0.8	0.8	0.866	1	0.9	0.866
Destination de rêve	0.6	0.7	0.666	1	0.8	0.733	1	0.8	0.733
Visiter les monuments	0.2	0.2	0.133	0.6	0.7	0.6	1	0.9	0.933
Moyen_ Précision	**0.466**	**0.466**	**0.444**	**0.8**	**0.766**	**0.710**	**1**	**0.866**	**0.866**
User9	**P@5**	**P@10**	**P@15**	**P@5**	**P@10**	**P@15**	**P@5**	**P@10**	**P@15**
Compréhension émotionnelle	0.2	0.3	0.333	0.4	0.6	0.5	0.8	0.8	0.8
Perception	0.2	0.2	0.266	0.8	0.7	0.666	1	1	0.866
Neuroscience	0.4	0.4	0.466	0.8	0.7	0.8	1	1	1
Moyen_ Précision	**0.2**	**0.2**	**0.222**	**0.666**	**0.666**	**0.655**	**0.933**	**0.933**	**0.888**
User10	**P@5**	**P@10**	**P@15**	**P@5**	**P@10**	**P@15**	**P@5**	**P@10**	**P@15**
Conflit	0	0	0	0.4	0.3	0.4	0.8	0.9	0.9
Stress	0.2	0.2	0.133	0.8	0.8	0.666	1	1	1
ACT	0	0	0	1	0.8	0.866	1	1	1
Moyen_ Précision	**0.066**	**0.066**	**0.044**	**0.733**	**0.633**	**0.644**	**0.933**	**0.966**	**0.966**

Tableau 3.2:*Résultats expérimentaux détaillés des précisions.*

D'après ces résultats (Tableau 3.2), nous constatons que notre modèle est performant et affiche des améliorations notables en termes de précision à p@5, p@10 et p@15 par rapport au modèle classique (sans reformulation) et automatique (reformulation automatique).

Nous remarquons que l'intégration de l'utilisateur a conduit à une augmentation de la précision par rapport aux approches classiques et automatique, et surtout dans la sélection des cinq premiers documents susceptibles d'être consultés par l'utilisateur.

Cas d'égalité des performances

Le tableau suivant (Tableau 3.3) représente les cas où le choix de l'utilisateur se confonde par hasard avec la reformulation automatique (mêmes valeurs de précision)

	Reformulation automatique			Reformulation interactive		
User1	**P@5**	**P@10**	**P@15**	**P@5**	**P@10**	**P@15**
Sécurité	1	1	0,933	1	1	0,933
User2	**P@5**	**P@10**	**P@15**	**P@5**	**P@10**	**P@15**
Application	1	0,9	0,866	1	0,9	0,866
User3	**P@5**	**P@10**	**P@15**	**P@5**	**P@10**	**P@15**
Consultation	0,8	0,8	0,866	0,8	0,8	0,866
User7	**P@5**	**P@10**	**P@15**	**P@5**	**P@10**	**P@15**
Bon plan	1	1	1	1	1	1
User8	**P@5**	**P@10**	**P@15**	**P@5**	**P@10**	**P@15**
Destination de rêve	1	0.8	0.733	1	0.8	0.733

Tableau 3.3*:* *Cas d'égalité des performances.*

Exemples de reformulation

Voici quelques exemples (Tableau3.4) illustrant comment la reformulation systématique et interactive se déroule :

User	**Requête initiale (Q)**	**Variantes**	**Reformulation automatique**	**Reformulation interactive**
User1	Q_1 = réseau	$\mathbf{Q'_3}$: Le tag qui maximise la similarité(4) avec les termes de Q.	Q '=réseau+ ***informatique***	Q'=réseau+***couche de neurone***
	Q_2 = java	$\mathbf{Q'_3}$: Le tag qui maximise la similarité(4) avec les termes de Q.	Q '=java+***swing***	Q '=java+***script***
	Q_3 = sécurité	$\mathbf{Q'_3}$: Le tag qui maximise la similarité(4) avec les termes de Q.	Q '=sécurité+ ***informatique***	Q '=réseau+***cryptographie***
User2	Q_1 = application	**Q '** : le terme qui maximise la similarité(3) avec les termes de Q.	Q'_1 = application + ***web***	Q'_1 = application + ***web***
User3	Q_1 = thérapie	$\mathbf{Q'_3}$: Le tag qui maximise la similarité(4) avec les termes de Q.	Q'_1 = thérapi + ***médicale***	Q'_1=thérapie + ***traitement génique***
…	…	…	…	…

Tableau3. 4:*Quelques exemples de déroulement de reformulation automatique et interactive.*

Nous remarquons que cette illustration explique le processus de reformulation systématique (automatique) et la reformulation interactive. Dans la reformulation automatique, le système toujours ajoute un terme qui maximise la similarité, comme dans l'exemple de l'utilisateur 1 où le système ajoute le terme « informatique » qui maximise la similarité (3) avec la requête initiale « réseau ». Dans la reformulation interactive, c'est l'utilisateur qui choisit l'une des variantes proposées par le système pour reformuler sa requête. Par exemple, dans le cas de l'utilisateur 1, il choisit la variante 3, le tag « couche de neurone », qui maximise la similarité (4) avec la requête initiale « réseau ».

Ci-dessous, nous présentons quelques exemples d'améliorations des résultats obtenus après la reformulation de requêtes dans notre approche de personnalisation :

Exemple 01 :

Les premiers résultats de la requête « Thérapie » envoyée sans reformulation par l'utilisateur 3, qui s'intéresse à la médecine, ne satisfont pas ses besoins (Figure3.6).

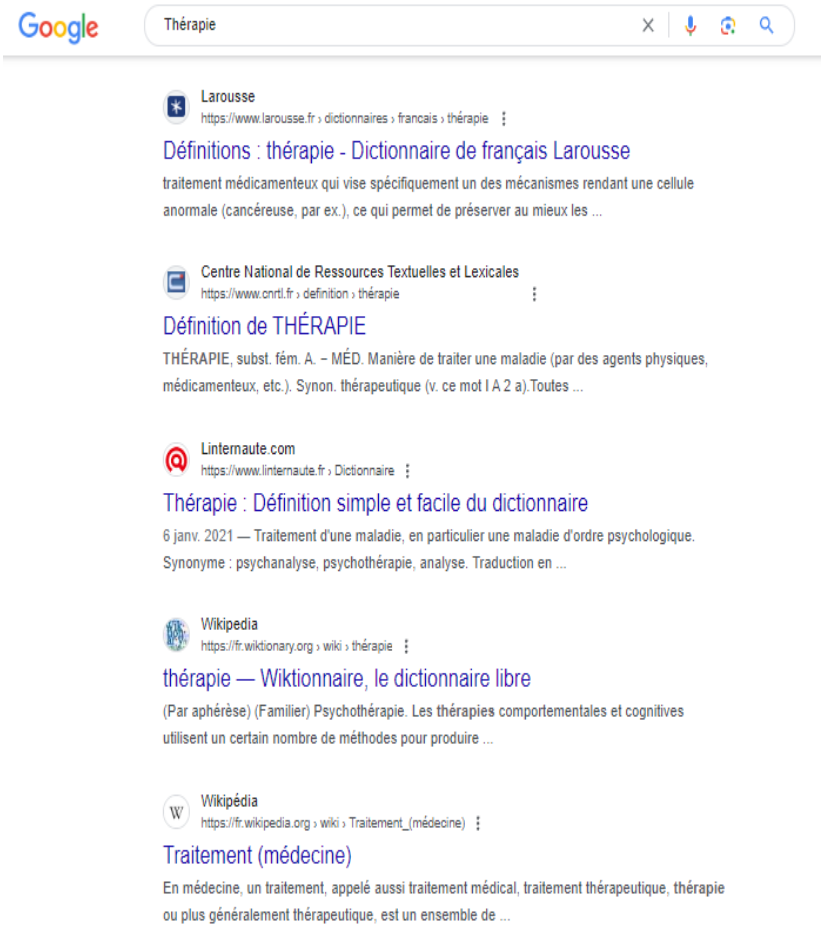


Figure3.6 : *Représentation des résultats de requête "Thérapie" sans reformulation.*

Deuxièmement, on utilise la reformulation automatique de la requête par l'ajout du terme « médicale » extrait à partir de son centre d'intérêt. Nous remarquons que les premiers résultats retournés par le moteur de recherche sont peu satisfaisants aux besoins d'information de l'utilisateur (Figure 3.7).



Figure3.7 : *Représentation des résultats de requête "Thérapie" après la reformulation automatique.*

À la fin, nous utilisons notre proposition pour reformuler la requête « Thérapie », en considérant l'avis de l'utilisateur et en ajoutant à la requête le tag « traitement génique » pour répondre à ses besoins (Figure3.8).

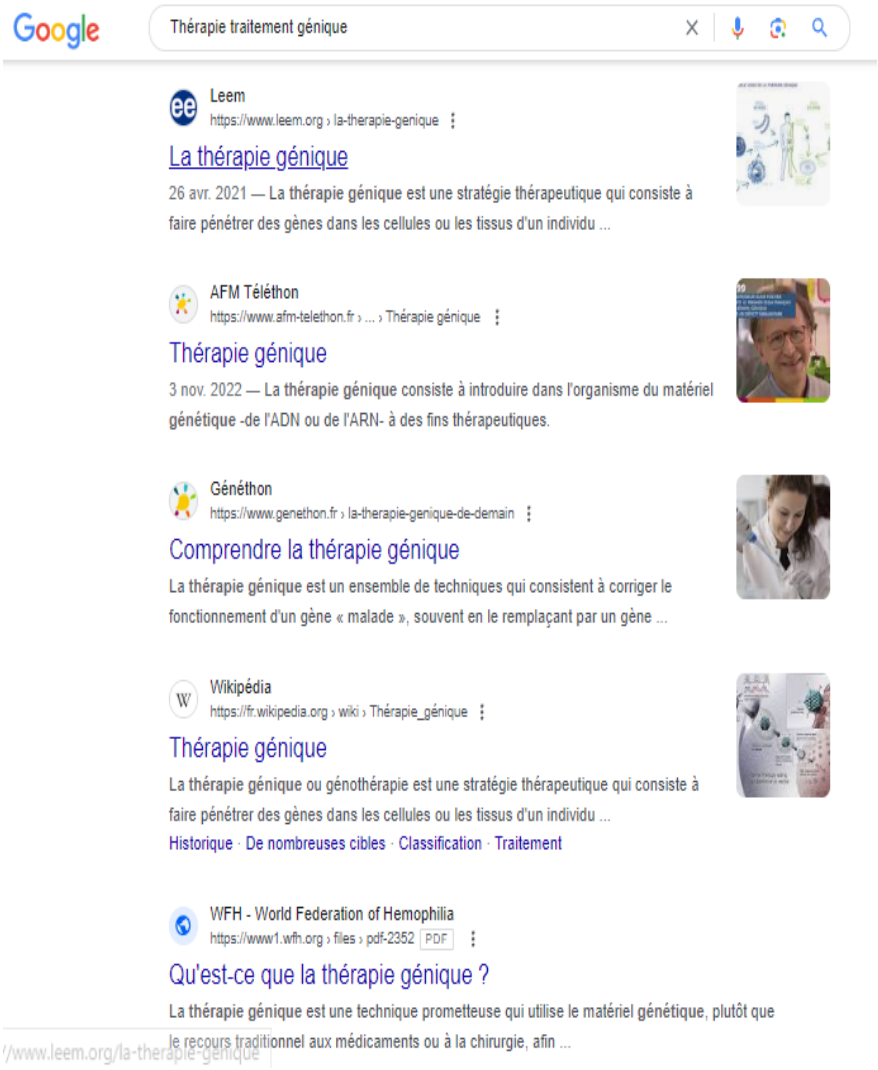


Figure3.8 : *Représentation de résultats de requête "Thérapie" après la reformulation interactive.*

Exemple 02 :

L'utilisateur8 a soumis la requête «Visiter les monuments »au moteur de recherche pour obtenir des résultats répondant à son besoin, et les résultats étaient les suivants (Figure3.19) .

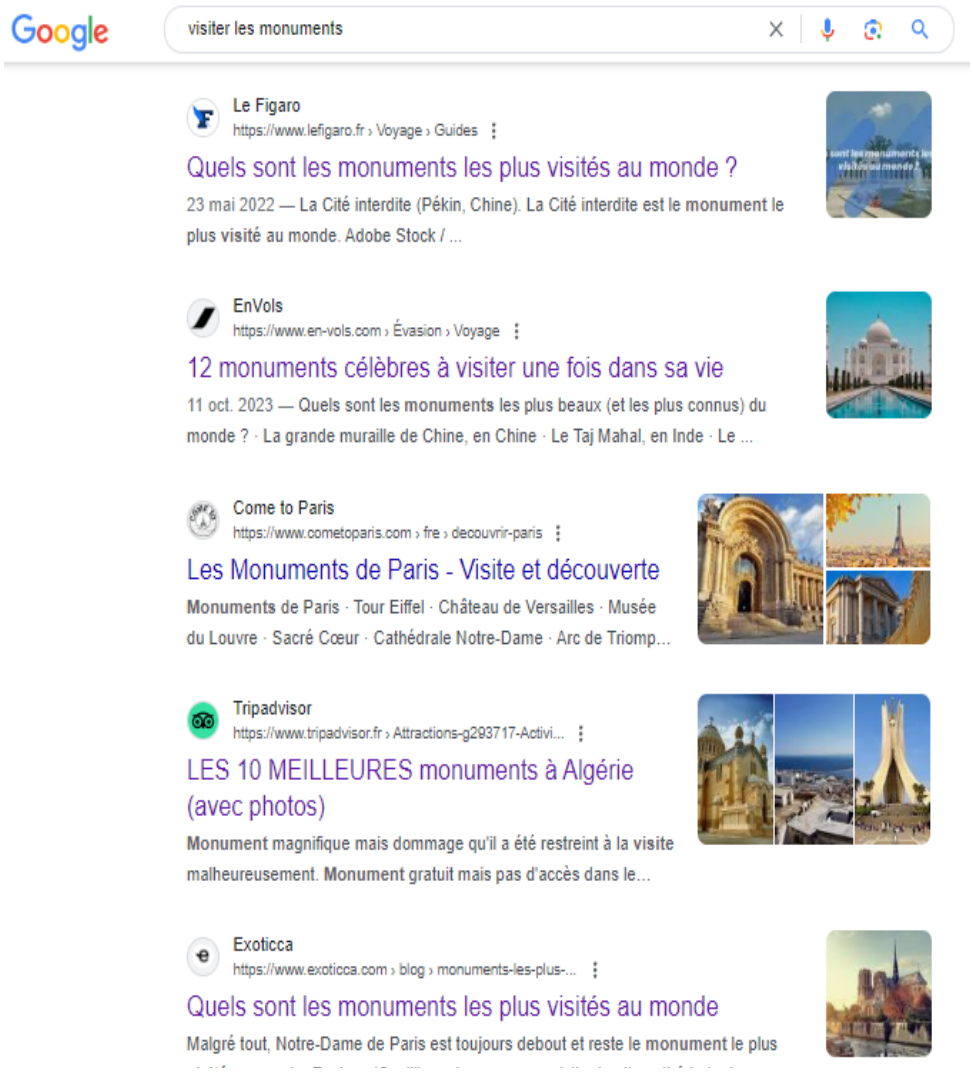


Figure3.9: *Représentation des résultats de requête "visiter les monuments" sans reformulation.*

Ces résultats ne satisfont pas ses besoins. Pour cela, l'utilisateur adopte la reformulation automatique, le système ajoute le terme « Historique », extrait à partir du centre d'intérêt de l'utilisateur 8. Les résultats se sont améliorés, mais l'utilisateur n'est pas entièrement satisfait (Figure 3.10).

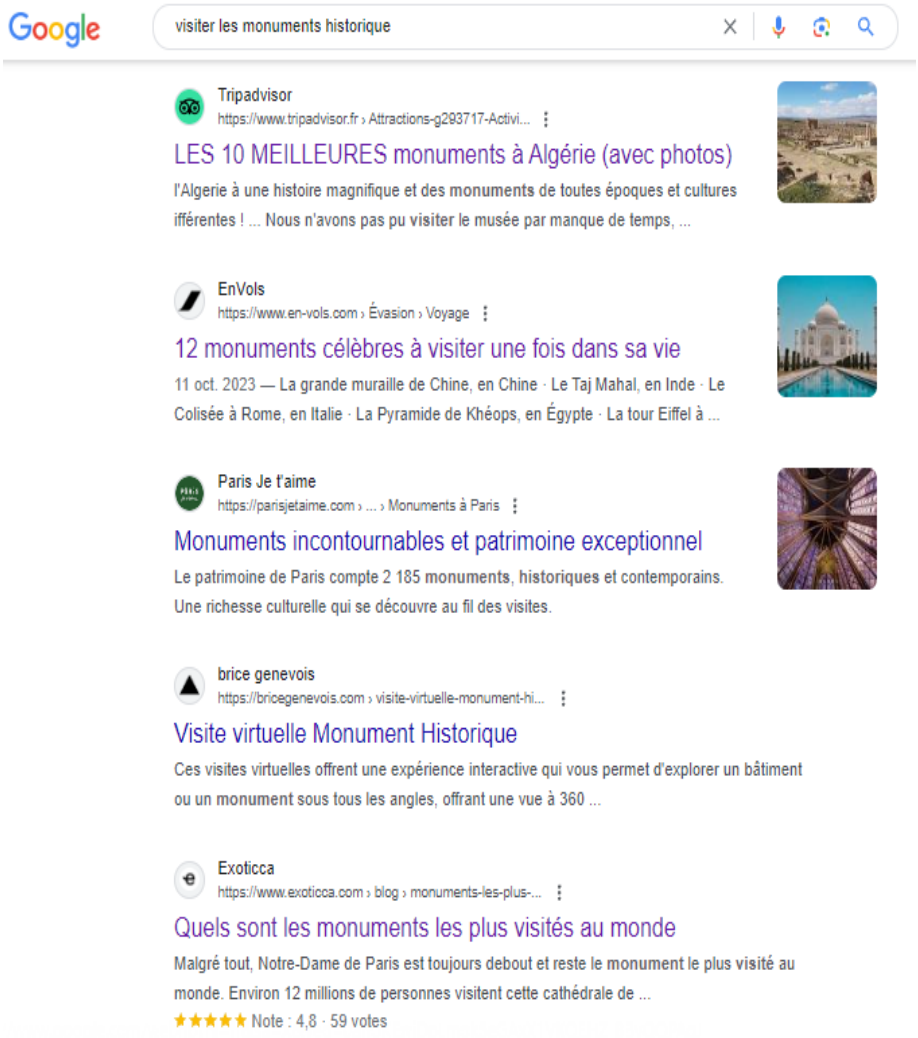


Figure 3.10:*Représentation des résultats de requête "visiter les monuments" après la reformulation automatique.*

Ces résultats se sont améliorés, mais l'utilisateur n'est pas entièrement satisfait. Pour cette raison, il utilise la reformulation interactive en considérant son avis et en ajoutant le terme « Algérie ». Les résultats obtenus sont les suivants (Figure 3.11) :

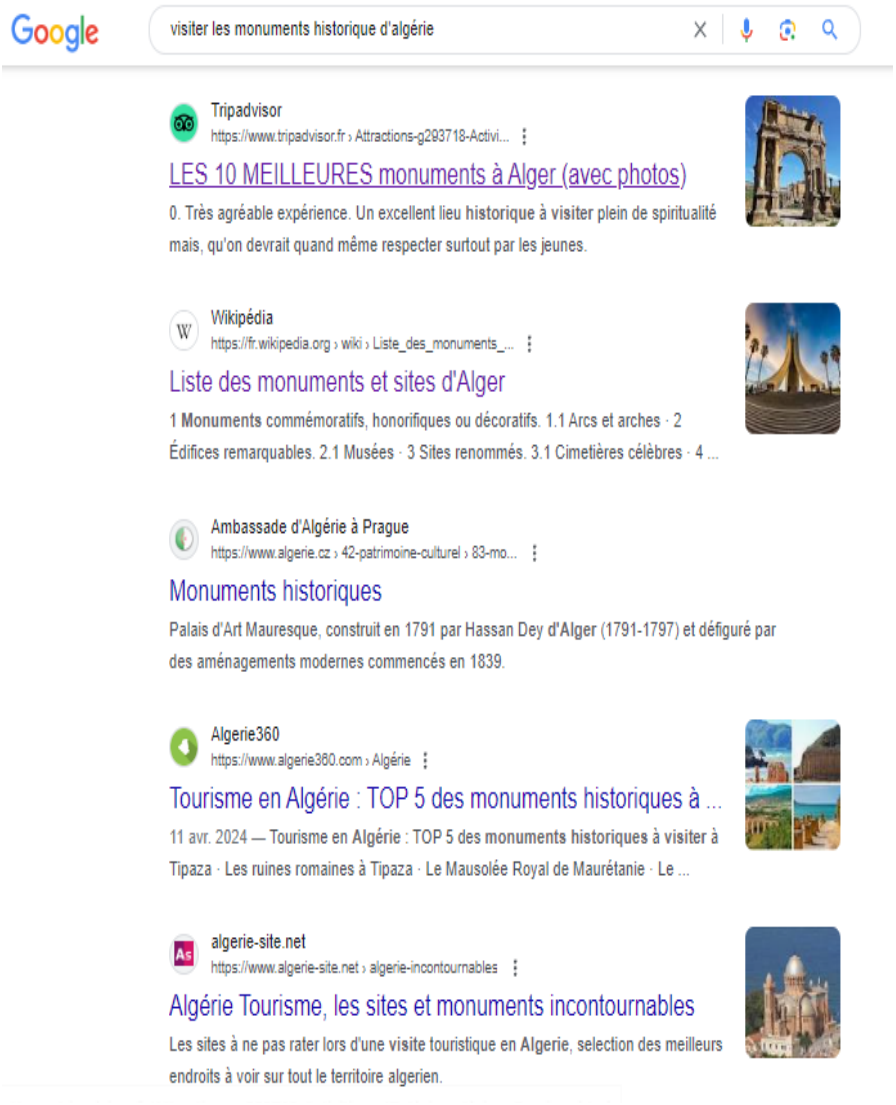


Figure 3.11: Représentation des résultats de requête "visiter les monuments" après la reformulation interactive.

Nous constatons que, après cette reformulation de la requête, les résultats ont beaucoup mieux amélioré et l'utilisateur été satisfait.

Résultats globaux de précision

Pour conclure, nous exposons le tableau (Tableau 3.4) qui résume les résultats de notre expérience et qui affiche la moyenne des précisions des résultats obtenus pour toutes les requêtes envoyées par les utilisateurs, selon les 3 scenarios déjà expliqués.

	Scenario (A) *Sans reformulation*	**Scenario (B)** *Reformulation automatique*	**Scenario (C)** *Reformulation interactive*
P@5	0.173	0.553	0.906
P@10	0.190	0.506	0.866
P@15	0.175	0.509	0.825
P_moy	**0.179**	**0523**	**0.865**

Tableau 3.5*:La moyenne des précisions des résultats pour toutes les requêtes soumises par les utilisateurs.*

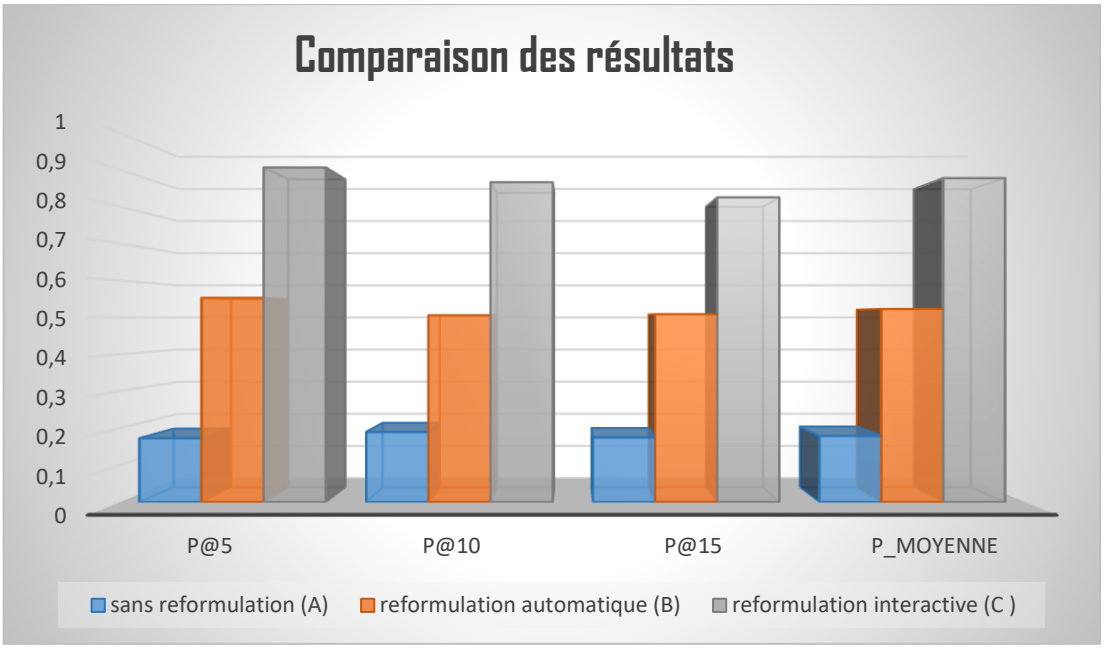


Figure 3.12: *Diagramme à barres de comparaison des résultats des 3 scénarios.*

	Taux d'amélioration (B-A)	**Taux d'amélioration (C-B)**	**Taux d'amélioration(C-A)**
P@5	0.38	0.353	0.733
P@10	0.316	0.36	0.676
P@15	0.334	0.316	0.65
P_moy	**0.343**	**0.354**	**0.698**

Tableau 3.6:*La moyenne des précisions des résultats pour toutes les requêtes soumises par les utilisateurs.*

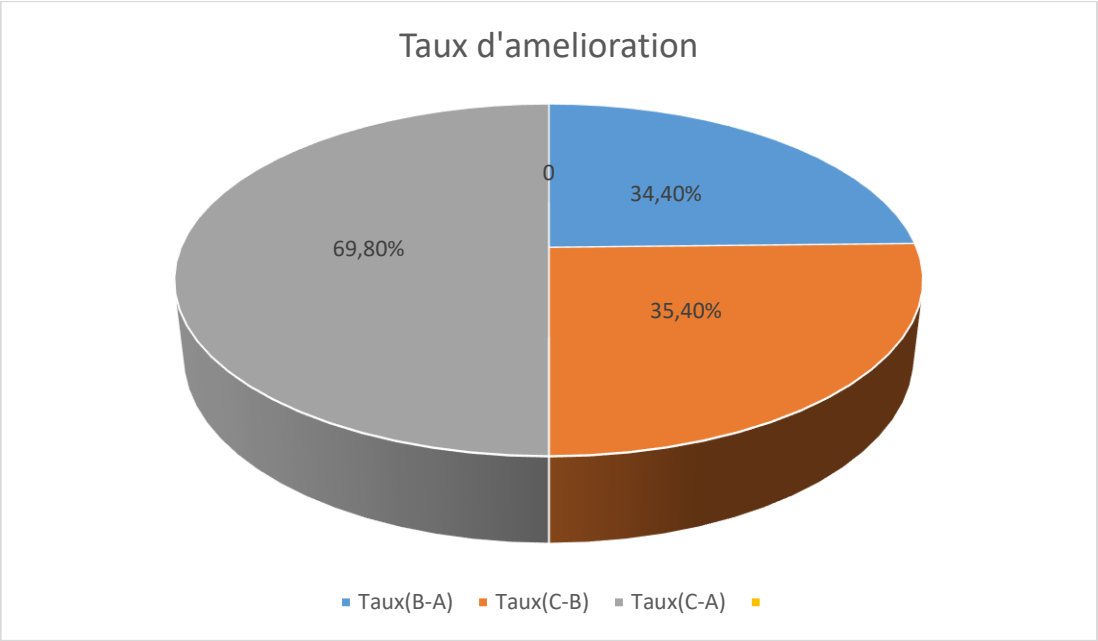


Figure 3.13: *Cercle relatif de comparaison de taux d'amélioration des 3 scénarios.*

Les résultats de la simulation montrent une amélioration significative des précisions avec la reformulation automatique (B) par rapport au scénario sans reformulation (A). En outre, notre proposition de reformulation interactive (C) offre une amélioration encore plus significative par rapport à la reformulation automatique (B). À l'issue, l'intégration de l'utilisateur dans la phase de la reformulation des requêtes améliore considérablement la précision des résultats de recherche.

5.3 Présentation d'application

Puisque l'expérimentation n'est pas encore implémentée intégralement, l'objectif de cette implémentation est de démontrer que notre solution proposée est applicable.

5.3.1 Modélisation de la base de données

Les différentes composantes du profil utilisateur sont représentées à travers un Modèle Conceptuel des Données (MCD) de la manière suivante :

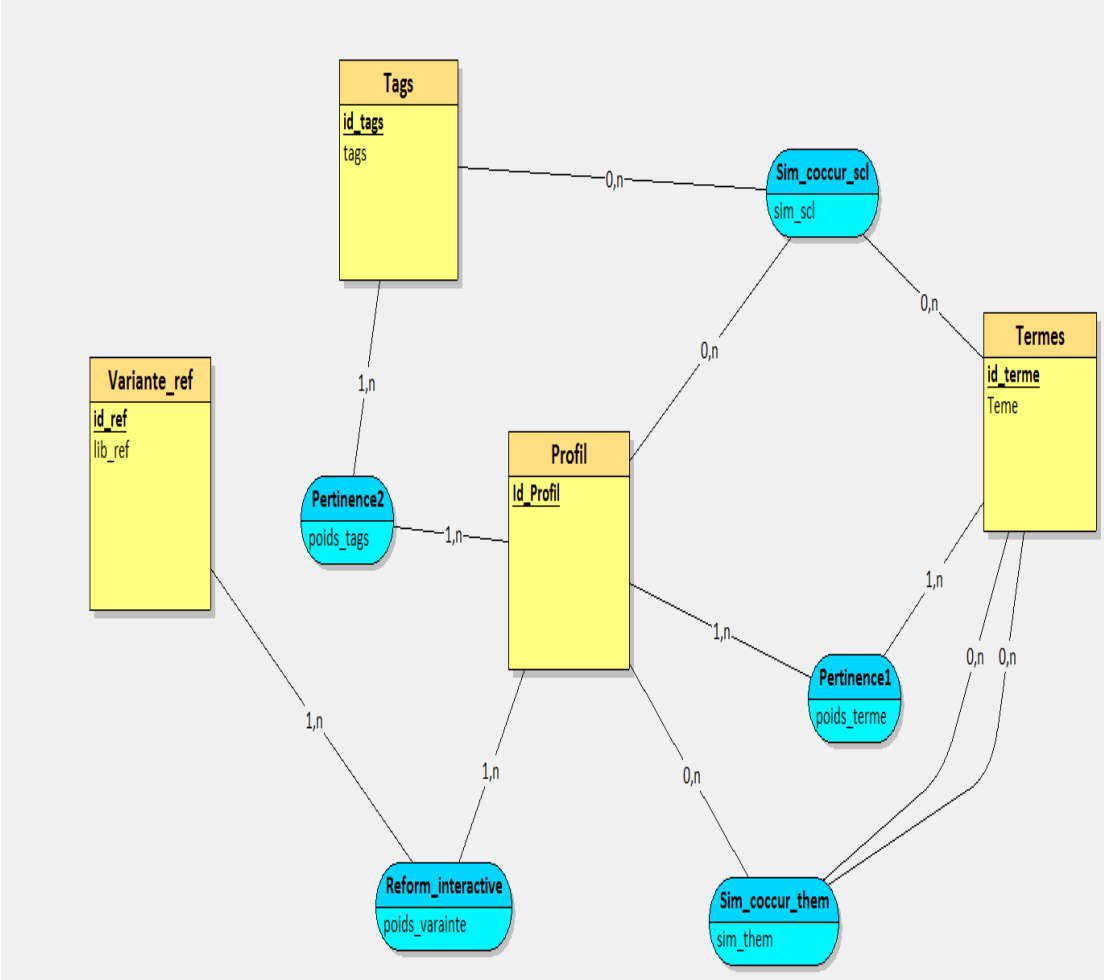


Figure3.14 : *Modèle de conception du profil utilisateur.*

a) Les entités :

N°	Entité	Identifiant	Propriétés et types
1	Profil	Id_profil	- Nom(VARCHAR) - Prénom(VARCHAR) - Sexe(CHAR) - Adresse(INT) - Mot de passe(VARCHAR) - E-mail(VARCHAR) - Age(INT)

			- Langue(CHAR) - Domaine(VARCHAR)
2	Terme	Id_terme	- Terme(VARCHAR)
3	Tags	Id_tags	- Tags(VARCHAR)
4	Variante_ref	Id_ref	- Libelle_ref(VARCHAR)

Tableau 3.7:*Entités du Modèle Conceptuel de Données (MCD).*

b) Les relations :

N°	Relation	Entités participants	Identifiant	Propriétés et types
1	Pertinence_terme	Profil Terme	Id_profil, Id_terme	- Poid_terme(FLOAT) - Freq_max_t(INT) - Freq_t(FLOAT - Nbr_doc(INT) - Nbr_con_t(INT)
2	Pertinence_tags	Profil Tags	Id_profil, Id_tags	- Poid_tags(FLOAT) - Freq_max_tg(INT) - Freq_tag(FLOAT) - Nbr_doc(INT) - Nbr_con_tg(INT)
3	Sim_coccur_them	Profil Terme1 Terme2	Id_profil, id_terme1, Id_terme2	- Sim_cooccur_them (FLOAT)

				- Nbr_doc_com(INT) - Nbr_doct1(INT) - Nbr_doct2(INT)
4	Sim_coccur_scl	Profil Terme Tags	Id_profil, id_terme, Id_tags	- Sim_cooccur_scl (FLOAT) - Nbr_doc_com(INT) - Nbr_doct(INT) - Nbr_doctg(INT)
5	Ref_interactive	Profil Variante_ref	Id_profil, Id_reform	- Nbr_choix_variante (INT) - Nbr_doc_ $pert_{5top}$ (INT) - Poid_var(FLOAT)

Tableau 3.8*:Relations d'association dans le Modèle Conceptuel de Données (MCD).*

5.3.2 Présentation de la base de données

Pour la création des tables de la base de données nous avons utilisé le *MySQL Workbench*. Dans ce qui suit nous allons présenter :

5.3.2.1 Présentation de MySQL

MySQL Workbench est un environnement intégré de développement et de gestion de bases de données MySQL. Il permet la modélisation visuelle des bases de données, la génération de scripts SQL, l'édition de données, le développement SQL avancé, ainsi que l'administration et la migration de bases de données. Son interface conviviale en fait un outil essentiel pour les développeurs et les administrateurs de bases de données MySQL.

5.3.2.2 Création des tables de base de données

Notre base de données se compose des tables suivantes :

Table	Action	Lignes	Type	Interclassement	Taille	Perte
pertinence_tags	Parcourir Structure Rechercher Insérer Vider Supprimer	43	InnoDB	utf8_general_ci	48,0 kio	-
pertinence_term	Parcourir Structure Rechercher Insérer Vider Supprimer	96	InnoDB	utf8_general_ci	48,0 kio	-
profile	Parcourir Structure Rechercher Insérer Vider Supprimer	10	InnoDB	utf8_general_ci	16,0 kio	-
reform_interactive	Parcourir Structure Rechercher Insérer Vider Supprimer	143	InnoDB	utf8_general_ci	48,0 kio	-
sim_coccur_scl	Parcourir Structure Rechercher Insérer Vider Supprimer	122	InnoDB	utf8_general_ci	80,0 kio	-
sim_coccur_them	Parcourir Structure Rechercher Insérer Vider Supprimer	254	InnoDB	utf8_general_ci	64,0 kio	-
tags	Parcourir Structure Rechercher Insérer Vider Supprimer	43	InnoDB	utf8_general_ci	16,0 kio	-
termes	Parcourir Structure Rechercher Insérer Vider Supprimer	99	InnoDB	utf8_general_ci	16,0 kio	-
variante_ref	Parcourir Structure Rechercher Insérer Vider Supprimer	4	InnoDB	utf8_general_ci	16,0 kio	-
9 tables	Somme	814	InnoDB	utf8_general_ci	352,0 kio	0 o

Figure3.15:*Les tables de la base de données.*

5.3.2.3 Liste des tables de la base de données

Notre base de données comprend les tableaux ci-dessous :

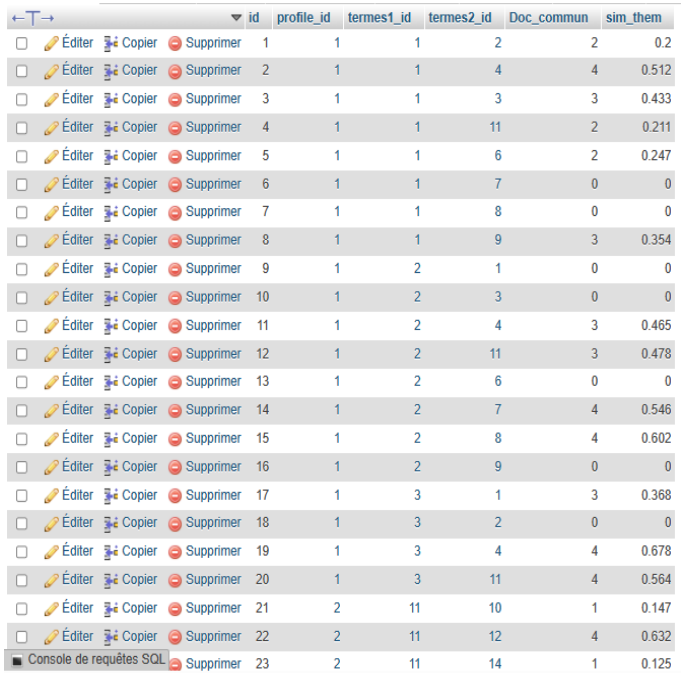
- *Table « terme »* : représente les termes pertinents de chaque utilisateur.
- *Table « tags »* : représente les tags pertinents pour chaque utilisateur, représenter comme suit(Figure3.16).

←T→	id_tags	tags
Éditer Copier Supprimer	1	Cryptographie
Éditer Copier Supprimer	2	Couche de neurone
Éditer Copier Supprimer	3	Micro édition
Éditer Copier Supprimer	4	Cyber sécurité
Éditer Copier Supprimer	5	Interface utilisateur
Éditer Copier Supprimer	6	Sémaphore
Éditer Copier Supprimer	7	Machine virtuel
Éditer Copier Supprimer	8	Mise à jour
Éditer Copier Supprimer	9	Création site web
Éditer Copier Supprimer	10	Suivi médicale
Éditer Copier Supprimer	11	Traitement génique
Éditer Copier Supprimer	12	Visite médicale
Éditer Copier Supprimer	13	Maladies géniques
Éditer Copier Supprimer	14	Production de dégradation de fibrinogène
Éditer Copier Supprimer	15	femme enceinte

Figure 3.16:*Exemple de la table "tags".*

- *Table «sim_coccur_scl »* : contient les relations de cooccurrence entre les tags et les termes pertinents de chaque utilisateur.

- *Table « sim_coccur_them »* : contient les relations de cooccurrence entre les termes pertinents de chaque utilisateur comme indiqué dans (Figure 3.17) :



	id	profile_id	termes1_id	termes2_id	Doc_commun	sim_them
Éditer Copier Supprimer	1	1	1	2	2	0.2
Éditer Copier Supprimer	2	1	1	4	4	0.512
Éditer Copier Supprimer	3	1	1	3	3	0.433
Éditer Copier Supprimer	4	1	1	11	2	0.211
Éditer Copier Supprimer	5	1	1	6	2	0.247
Éditer Copier Supprimer	6	1	1	7	0	0
Éditer Copier Supprimer	7	1	1	8	0	0
Éditer Copier Supprimer	8	1	1	9	3	0.354
Éditer Copier Supprimer	9	1	2	1	0	0
Éditer Copier Supprimer	10	1	2	3	0	0
Éditer Copier Supprimer	11	1	2	4	3	0.465
Éditer Copier Supprimer	12	1	2	11	3	0.478
Éditer Copier Supprimer	13	1	2	6	0	0
Éditer Copier Supprimer	14	1	2	7	4	0.546
Éditer Copier Supprimer	15	1	2	8	4	0.602
Éditer Copier Supprimer	16	1	2	9	0	0
Éditer Copier Supprimer	17	1	3	1	3	0.368
Éditer Copier Supprimer	18	1	3	2	0	0
Éditer Copier Supprimer	19	1	3	4	4	0.678
Éditer Copier Supprimer	20	1	3	11	4	0.564
Éditer Copier Supprimer	21	2	11	10	1	0.147
Éditer Copier Supprimer	22	2	11	12	4	0.632
Console de requêtes SQL Supprimer	23	2	11	14	1	0.125

Figure3.17:*Exemple de la table "sim_coccur_them".*

- *Table « pertinence_terme»* : représente le degré de pertinence des termes de chaque utilisateur, elle contient le poids de chaque terme, tel que présenté dans (Figure 3.18) :

	id	termes_id_termes	profile_id_profile	poids_terme
Éditer Copier Supprimer	1	1	1	0.521
Éditer Copier Supprimer	2	2	1	0.5
Éditer Copier Supprimer	3	3	1	0.108
Éditer Copier Supprimer	4	4	1	0.751
Éditer Copier Supprimer	5	11	1	0.654
Éditer Copier Supprimer	6	6	1	0.155
Éditer Copier Supprimer	7	7	1	0.641
Éditer Copier Supprimer	8	8	1	0.279
Éditer Copier Supprimer	9	9	1	0.256
Éditer Copier Supprimer	10	10	2	0.412
Éditer Copier Supprimer	11	11	2	0.649
Éditer Copier Supprimer	12	12	2	0.466
Éditer Copier Supprimer	13	14	2	0.602
Éditer Copier Supprimer	14	13	2	0.117
Éditer Copier Supprimer	15	15	2	0.588
Éditer Copier Supprimer	16	16	2	0.317
Éditer Copier Supprimer	17	18	2	0.445
Éditer Copier Supprimer	18	17	2	0.488
Éditer Copier Supprimer	19	19	2	0.165
Éditer Copier Supprimer	20	20	3	0.43
Éditer Copier Supprimer	21	21	3	0.244
Éditer Copier Supprimer	22	22	3	0.213

Figure 3.18:*Exemple de la table "pertinence_terme".*

- *Table « pertinence_tags » :* représente le degré de pertinence des tags de chaque utilisateur, elle contient le poids de chaque tag.
- *Table « profil » :* cette table contient les informations personnelles qui identifier chaque utilisateur et représente comme suit (Figure 3.19) :

	id_profile	nom	prenom	email	password	age	sexe	adresse	langue	domaine
Éditer Copier Supprimer	1	Douma	Fatma	doumafatma@email.com	fatma22	22	femme	Médéa	Français	Informatique
Éditer Copier Supprimer	2	Ramdani	Fathia	fathia@email.com	fathia123	22	femme	ksar el boukhari	Anglais	Informatique
Éditer Copier Supprimer	3	Benaissa	Rihab	benaissa@email.com	rihab	24	femme	Batna	Français	médcine
Éditer Copier Supprimer	4	Douma	Ikram	doumaikram@email.com	ikram	26	femme	médéa	Français	médecine
Éditer Copier Supprimer	5	Kenouche	Asma	asma@email.com	asmakenouche	33	femme	Alger	Arabe	cuisine
Éditer Copier Supprimer	6	Sahraoui	Fatiha	fatiha@email.com	fatihasahraoui	45	femme	Blida	Arabe	cuisine
Éditer Copier Supprimer	7	Gamaz	Amir	gamaz@email.com	amiramir	32	Homme	béjaia	Anglais	Tourisme
Éditer Copier Supprimer	8	Benkhlfa	Abdellah	abdellah@email.com	abdellah	38	Homme	Sétif	Arabe	Tourisme

Figure 3.19:*Exemple de la table "profil".*

- *Table « variante_ref »* : représente les quatre variantes de reformulation.
- *Table « reformulation_interactive »* : représente le poids des variantes de reformulation pour chaque utilisateur, elle contient nombre de choix de variante, nombre de documents dans les cinq tops, tel que démontré dans (Figure 3.20) :

	id	profile_id	variante_ref_id	Nbr_choix	Nbr_doc_pertinent	poids_variante
Éditer Copier Supprimer	1	1	1	0	0	0
Éditer Copier Supprimer	2	1	2	0	0	0
Éditer Copier Supprimer	3	1	3	0	0	0
Éditer Copier Supprimer	4	1	4	0	0	0
Éditer Copier Supprimer	24	2	1	0	0	0
Éditer Copier Supprimer	25	2	2	0	0	0
Éditer Copier Supprimer	26	2	3	0	0	0
Éditer Copier Supprimer	27	2	4	0	0	0
Éditer Copier Supprimer	41	3	1	0	0	0
Éditer Copier Supprimer	42	3	2	0	0	0
Éditer Copier Supprimer	43	3	3	0	0	0
Éditer Copier Supprimer	44	3	4	0	0	0
Éditer Copier Supprimer	54	4	1	0	0	0
Éditer Copier Supprimer	55	4	2	0	0	0
Éditer Copier Supprimer	56	4	3	0	0	0
Éditer Copier Supprimer	57	4	4	0	0	0

Figure 3.20:*Exemple de la table "reformulation_interactive".*

5.3.3 Environnement de développement de l'application

Pour réaliser cette application, nous devons utiliser certains langages de programmation et logiciels :

PHP :

PHP est un langage de script open-source principalement utilisé pour le développement web côté serveur. Il permet de créer des pages web dynamiques et interactives en s'intégrant facilement avec les bases de données[1].

[1] https://www.w3schools.com/php/



XAMPP:

XAMPP est une distribution de serveur web open-source qui combine Apache, MySQL, PHP et Perl, facilitant ainsi le développement et les tests d'applications web sur un environnement local[2].



HTML :

HTML (HyperText Markup Language) est le langage standard utilisé pour créer et structurer le contenu des pages web. Il permet d'organiser le texte, les images, les liens et autres éléments multimédias sur un site internet[3].



CSS :

CSS (Cascading Style Sheets) est un langage de feuille de style utilisé pour décrire l'apparence et la mise en page des documents HTML. Il permet de contrôler les couleurs, les polices, les espacements et la disposition des éléments sur une page web[4]



JavaScript :

JavaScript est un langage de programmation utilisé principalement pour rendre les pages web interactives et dynamiques. Il permet de manipuler le contenu, le style et le comportement des éléments HTML en temps réel[5].

[2] https://www.w3schools.com/xampp/
[3]https://www.w3schools.com/html/
[4] https://www.w3schools.com/css/

[5] https://www.w3schools.com/javascript/

VUE.JS :

Vue.js est un frame Work JavaScript progressif et open-source utilisé pour construire des interfaces utilisateur interactives et réactives dans les applications web. Il se concentre sur la simplicité et la flexibilité, facilitant ainsi le développement d'applications complexes[6].

5.3.4 Présentation des interfaces

Cette page (Figure 3.21) permet à l'utilisateur de se connecter à l'application en introduisant son e-mail et son mot de passe.

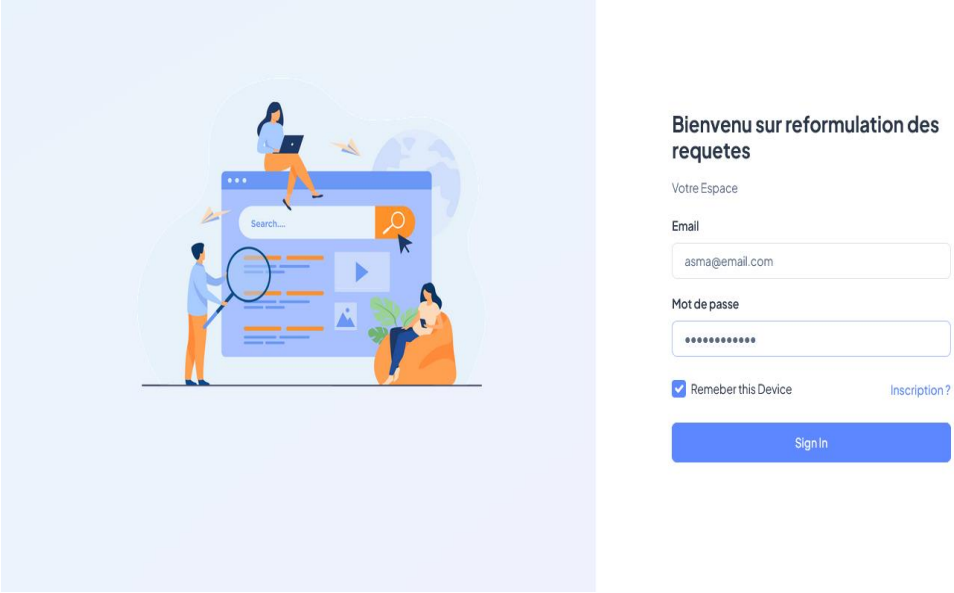


Figure3.21:*Page d'identification.*

[6] https://www.w3schools.com/vue.js/

Une fois l'utilisateur authentifié, la page de recherche apparaîtra (Figure 3.22). Sur cette page, l'utilisateur peut effectuer une recherche en mode classique, systématique ou interactif.

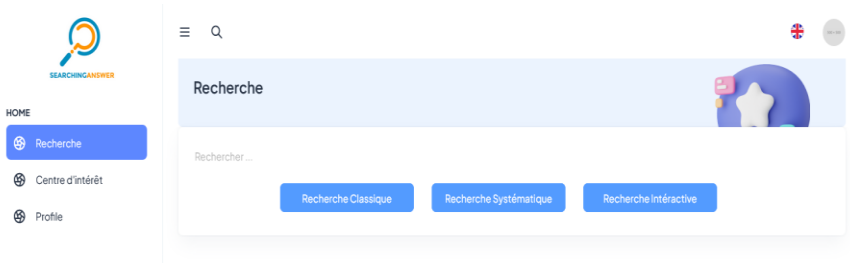


Figure3.22: *Page de recherche utilisateur.*

Lorsque l'utilisateur saisit sa requête et lance une recherche en mode classique, il est redirigé vers la page des résultats du moteur de recherche Google, comme illustré par la figure suivante (Figure 3.23) :

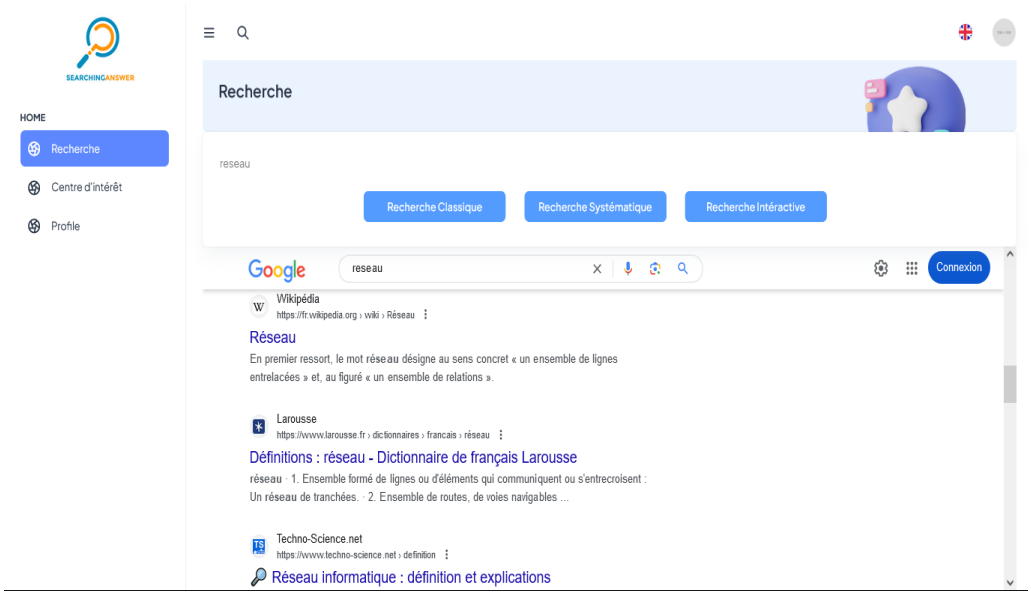


Figure 3. 23 : *Résultats de recherche en mode classique.*

En revanche, s'il décide de lancer la recherche en mode systématique (Figure 3.24), le système sélectionne le terme « sans fil » qui maximise la similarité (3) avec la requête initiale « réseau ».

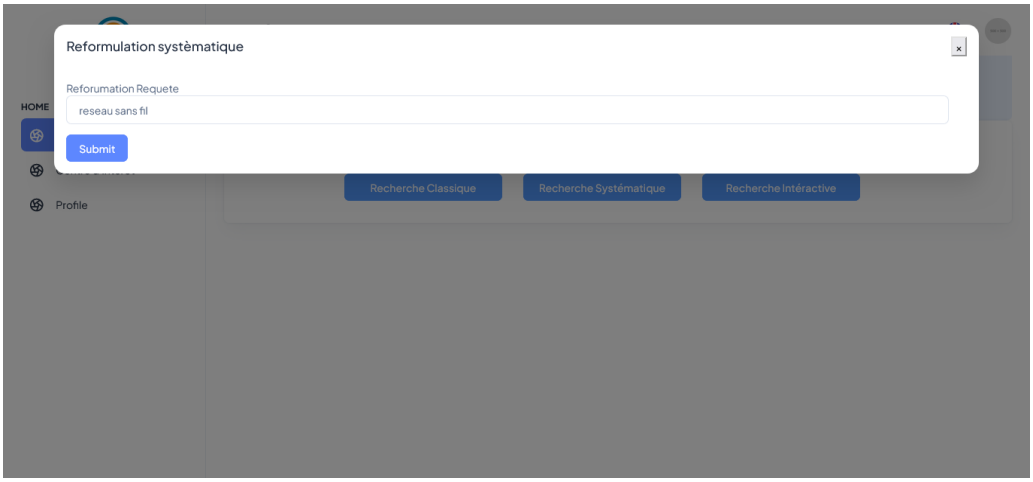


Figure 3. 24 : *Reformulation automatique de la requête.*

Lorsque l'utilisateur clique sur le bouton « Sumbit », la requête est modifiée et envoyée au moteur de recherche Google, comme le montre la (Figure 3.25) :

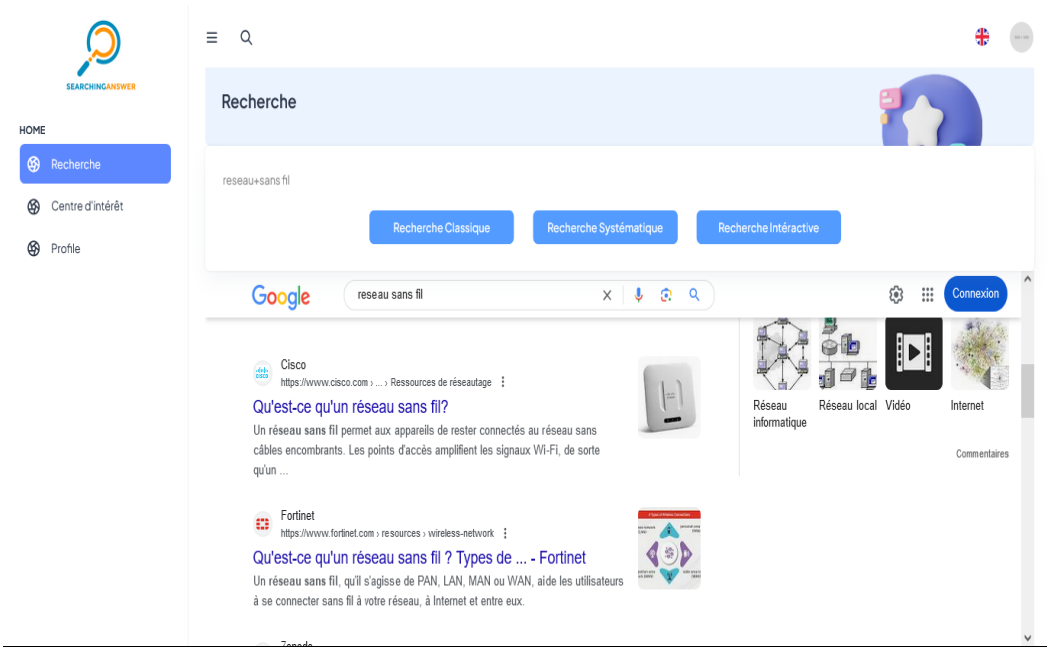


Figure 3. 25: *Résultats de recherche après la reformulation automatique.*

Cependant, s'il choisit de lancer la recherche en mode interactif, un processus complet est suivi pour reformuler la requête avant de l'envoyer au moteur de recherche, afin d'obtenir des résultats adaptés à ses besoins. Les figures (Figure 3.26, Figure 3.27) illustrent les étapes de ce processus de recherche :

Classement des variantes :

Lorsque l'utilisateur envoie sa requête au moteur de recherche, le système lui propose quatre variantes : deux pour les termes (informatique, sans fil) et les autres pour les tags (wifi, couche de neurone). Ces variantes sont construites à partir de l'extraction des termes ou des tags qui ont la plus grande similarité de pondération (Figure 3.26).

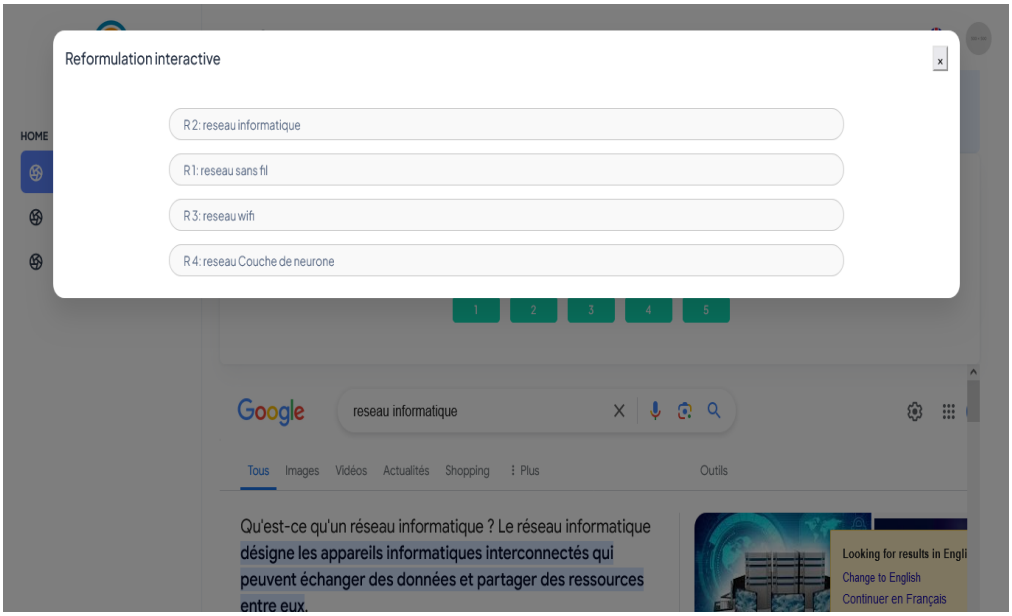


Figure 3.26 : *Classement des variantes.*

Choix d'une variante :

Une fois que le système renvoie à l'utilisateur les quatre variantes, celui-ci choisit la variante N3(le tag qui maximise la similarité 4 avec la requête « réseau ») la plus adéquate à son besoin d'information (Figure 3.27).

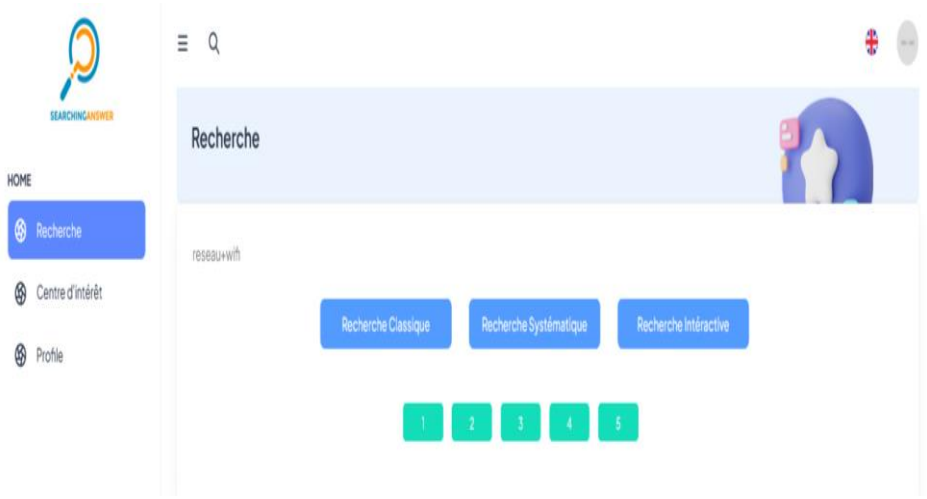


Figure 3.27 : *Choix d'une variante.*

Résultats de recherche :

Lorsque l'utilisateur choisit la variante N3, la requête initiale "réseau" est modifiée en ajoutant la variante N3 (réseau+wifi) et envoyée au moteur de recherche Google, comme le montre la (Figure 3.28) :

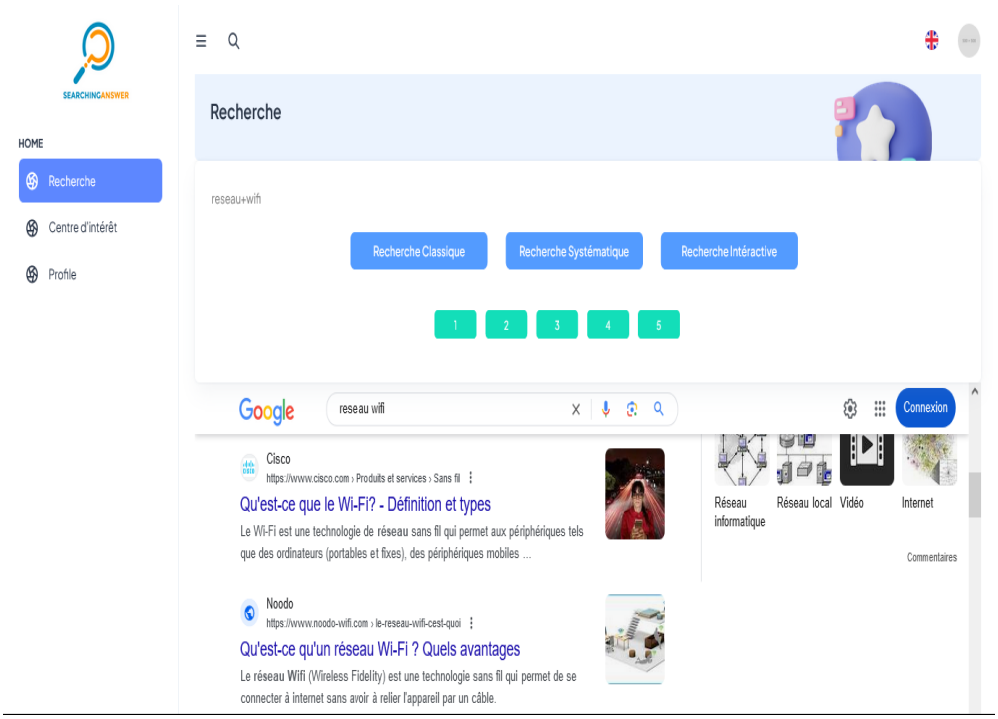


Figure 3.28 : *Résultats de recherche après la reformulation interactive.*

Mis à jour des poids :

Après que l'utilisateur a sélectionné la variante N3 de reformulation, le nombre de choix, le nombre de documents pertinents dans les cinq tops et les poids des variantes étaient nuls (Figure 3.29).

	id	profile_id	variante_ref_id	Nbr_choix	Nbr_doc_pertinent	poids_variante
Éditer Copier Supprimer	1	1	1	0	0	0
Éditer Copier Supprimer	2	1	2	0	0	0
Éditer Copier Supprimer	3	1	3	0	0	0
Éditer Copier Supprimer	4	1	4	0	0	0

Figure 3.29: *Poids avant mis à jour.*

Mais après la sélection, il devient 1. De même, le nombre de documents jugés pertinents par l'utilisateur parmi les cinq premiers résultats devient 3, pour mettre à jour les poids des quatre variantes (Figure 3.30).

	id	profile_id	variante_ref_id	Nbr_choix	Nbr_doc_pertinent	poids_variante
Éditer Copier Supprimer	1	1	1	0	0	0
Éditer Copier Supprimer	2	1	2	0	0	0
Éditer Copier Supprimer	3	1	3	1	3	0.666667
Éditer Copier Supprimer	4	1	4	0	0	0

Figure 3.30 : *Mis à jour des poids.*

6 Conclusion

Dans ce chapitre, nous avons présenté le cadre conceptuel de notre approche pour reformuler la requête utilisateur, basée sur un profil d'utilisateur social. Ce profil combine deux aspects : thématique, représenté par les termes pertinents, et social, représenté par les annotations sociaux (tags). Nous avons également présenté les relations de cooccurrence entre les termes de l'historique pertinent de recherche et les tags, dans le but de valoriser ces relations pour sélectionner les termes les plus adéquats pour reformuler la requête de l'utilisateur ambiguë. Nous avons proposé quelques variantes pour les termes et les tags et fait intervenir l'utilisateur pour choisir la variante la plus convenable à son besoin d'information.

Cette approche vise à améliorer la pertinence des résultats de recherche en reformulant la requête initiale de l'utilisateur.

Conclusion général

Conclusion générale

Conclusion et perspectives

Dans ce travail, nous avons présenté une nouvelle approche de reformulation interactive et personnalisée de la requête utilisateur pour améliorer la pertinence des résultats de recherche d'information. Cette approche repose, en premier lieu, sur la définition d'un profil d'utilisateur qui regroupe ses indicateurs d'intérêts dérivés des deux contextes thématique et social de ses recherches antérieures. L'aspect thématique est représenté par les termes pertinents, et l'aspect social est représenté par les annotations sociales (tags) effectuées sur les documents pertinents. Et en deuxième lieu, sur la proposition d'un ensemble de variantes de reformulations possibles de la requête d'utilisateur, et de donner la décision finale à l'utilisateur en choisissant la reformulation qu'il trouve plus pertinente pour lui. Ce qui donne un aspect interactif à notre approche.

Comme perspectives de ce travail, on peut citer :

- Utiliser un corpus, générale ou spécialisé, pour tirer les liens de cooccurrence entre les termes et les tags pour avoir des relations issues d'une utilisation humaine plus large, et d'exploiter le profil d'utilisateur pour le classement des variantes pour assurer la pertinence utilisateur.
- Introduire les techniques de deep-learning pour la description des intérêts de l'utilisateur et pour l'extraction des liens sémantiques entre les composantes pertinentes de profil d'utilisateur.
- Elargir le champ d'évaluation pour avoir des résultats mieux reflétant la réalité.

Bibliographie

Bibliographie

Bibliographie

Akermi, I., Boughanem, M., & Faiz, R. (2015). Une approche de recommandation proactive dans un environnement mobile. *INFORSID 2015.* France.

Allan, J., & Salton, G. (1994). Automatic Text Decomposition and Structuring. *Actes du Congrès RIAO'94, Intelligent Multimedia Information Retrieval Systems and Management.* New York, USA.

Amato, G., & Straccia, U. (1999). User Profile Modeling and Applications to Digital Libraries. *Proceedings of the 3rd European Conference on Research and Advanced Technology for Digital Libraries (ECDL).*

Andreasen, T. (2003). An approach to knowledge-based query evaluation: Fuzzy Databases. *Fuzzy Sets and Systems* .

Badache, I. (2016). *Recherche d'information sociale: exploitation des signaux sociaux pour améliorer la recherche d'information.* Thèse de doctorat, Université Paul Sabatier - Toulouse III, France.

Badache, I., & al. (2019). Recherche d'information sociale en langue arabe: Cas de facebook. *16th CORIA'19.* Lyon, France.

Baeza-Yates, R., & Ribeiro-Neto, R. A. (1999). *Modern Information Retrieval.* ACM Press (New York) ; Addison-Wesley (Harlow, England).

Basile, P., Caputo, A., & Semera, G. (2014). Extending an Information Retrieval System through Time Event Extraction. *8th International Workshop on Information Filtering and Retrieval, DART 2014, Co-located with XIII AIIA Symposium on Artificial Intelligence, AIIA 2014.*

Baziz, M. (2005). *Indexation conceptuelle guidée par ontologie pour la recherche d'information.* Thèse de doctorat.

Bendella, M., & Quafafou, M. (2019). Expansion de requêtes à base de motifs et de plongements de mots pour améliorer la recherche de microblogs. *16th CORIA'19.* Lyon, France.

Bender, M., Crecelius, T., Kacimi, M., Michel, S., Neumann, T., Parreira, J. X., et al. (2008). Exploiting social relations for query expansion and result ranking. Dans IEEE (Éd.), *Proceedings of IEEE 24th International Conference on Data Engineering Workshop (ICDEW 2008).*

Biancalana, C., & al, e. (2013). Social Semantic Query Expansion. *ACM Transactions on Intelligent Systems and Technology , 4* (4).

Bila, N., Cao, J., Dino, R., Ho, T., & Hull, R. K. (2008). Mobile user profile acquisition through network observables and explicit user queries. *9th International Conference on Mobile Data Management.* Beijing, China.

Bibliographie

Bischoff, k., & al. (2008). Can All Tags Be Used for Search? *Proceedings of the 17th ACM Conference on Information and Knowledge Management (CIKM '08).* Californie, USA.

Borlund, P. (2003). The concept of relevance in IR. *Journal of the American Society for Information Science and Technology , 54* (10).

Borlund, P. (2003). The Concept of Relevance in IR. *Journal of the American Society for Information Science and Technology , 54* (10).

Borlund, P., & Ingwersen, P. (1998). Measures of relative relevance and ranked half-life: performance indicators in interactive IR. Dans C. W. Eds (Éd.), *Proceedings of the 21st ACM SIGIR International Conference on Research and Development.*

Boubekeur, A. F. (2008). *Contribution à la définition de modèles de recherche d'information flexibles basés sur les CP-Nets.* Thèse de doctorat, Université Toulouse III - Paul Sabatier, France.

Boughanem, C. S.-D., & Chrisment, C. (1999). Query modification based on relevance back-propagation in an ad hoc environment. *Information Processing & Management (Inf. Process. Manage.)* .

Bouhini, C. (2014). Impact des réseaux sociaux sur le processus de recherche d'information. Université Jean Monnet de Saint Étienne et Universités de Lyon, France.

Bouhini, C., Géry, M., & Large, C. (2014). Integrating user's profile in the query model for Social Information Retrieval. *Information Science (RCIS), 2014.*

Bouhini, C., Gery, M., & Largeron, C. (2016). Personalized information retrieval models integrating the user's profile. *IEEE 10th International Conference on Research Challenges in Information Science (RCIS 2016).* Grenoble, France.

Bouhini, C., Gery, M., & Largeron, C. (2016). Personalized information retrieval models integrating the user's profile. *IEEE 10th International Conference on Research Challenges in Information Science (RCIS 2016).* Grenoble, France.

Bouidghaghen, O. (2011). Accès contextuel à l'information dans un environnement mobile : approche basée sur l'utilisation d'un profil situationnel de l'utilisateur et d'un profil de localisation des requêtes. Université Toulouse 3, France.

Bouidghaghen, O. (2009). Recherche contextuelle d'information dans un environnement mobile. *CORIA'09.* Presqu'île de Giens, France.

Bouramoul, A., Kholladi, M.-K., & Doan, B.-L. (2011). Evaluation of Information Retrieval Systems Towards a New Context-Based Approach. *4th International Conference on Information Systems and Economic Intelligence.* Marrakech, Morocco.

Challam, V., Gauch, S., & Chandramouli, A. (2007). Contextual Search Using Ontology-Based User Profiles. *Proceedings of RIAO 2007.* Pittsburgh, USA.

Bibliographie

Chaudiron, S., & Ihadjadenem. (2002). *Recherches récentes en sciences de l'information : Convergences et dynamiques.* (G. R. V. Couzinet, Éd.)

Daoud, M. (2009). *Accès personnalisé à l'information : approche basée sur l'utilisation d'un profil utilisateur sémantique dérivé d'une ontologie de domaines à travers l'historique des sessions de recherche.* Thèse de doctorat.

Daoud, M., Tamine, L., & Boughanem, M. (2011). A personalized search using a semantic distance measure in a graph-based ranking model. *Journal of Information Science , 37* (6).

Dounas, T., & Ould Fella, M. (2015). *Implémentation d'un modèle d'appariement pour un Système de Recherche d'Information Personnalisé.* Tiziouzou.

Duong, T., Uddin, M., & Nguyen, C. (2013). Personalized semantic search using ODP: A study case in academic domain. *ICCSA 2013.*

El ghali, B. (2016). *Modèles Contextuels pour la Recommandation et l'Expansion de Requêtes en Recherche d'Information.* Thèse de doctorat.

Fox, C. (1992). *Lexical analysis and stoplists. In Information retrieval.* Pertinence,Hall.

Gauch, s., Chaffe, J., & Pretschner, A. (2003). Ontology-based personalized search and browsing. *Web Intelligence and Agent System* .

Gowan, J. (2003). *A multiple model approach to personalised information access.* Mémoire de Master, Université de College Dublin.

Haddad, L. (2015). *Un modèle de langue pour une recherche d'information personnalisée.* Mémoire de master, Université Mouloud Mammeri, Tizi-Ouzou, Algérie.

Hattori, S., Tezuka, T., & Tanaka, K. (2007). Context-aware query refinement for mobile web search. *2007 International Symposium on Applications and the Internet Workshops.*

Heymann, P., Ramage, D., & Garcia-Moli, H. (2008). Social Tag Prediction. *Proceedings of the 31st ACM SIGIR'08.* Singapour.

Jarvelin, K., & Kekalainen, J. (2002). Cumulative Gain-Based Evaluation of IR Techniques. *ACM Transactions on Information Systems , 20* (4).

Jarvelin, K., & Kekalainen, J. (2000). IR Evaluation Methods for Highly Relevant Documents. Dans B. e. al (Éd.), *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval.*

Kanhabua, N., & Nørvag, K. (2012). Learning to Rank Search Results for Time-sensitive Queries. *Proceedings of the 21st ACM International Conference on Information and Knowledge Management (CIKM '12).*

Kechid, S. (2009). *Integration du modèle utilisateur dans un système de recherche d'information distribuée.* Thèse de doctorat, Université USTHB, Algérie.

Bibliographie

Khan, L. R. (2000, Août). Ontology-based Information Selection. Faculty of the Graduate School, University of Southern California.

Khazri, M., Tmar, M., Abid, M., & Boughanem, M. (2008). Proposition pour l'intégration des réseaux petits mondes en recherche d'information. (M. K. Assogba, Éd.) *11*.

Kostadinov, D. (2003). *La personnalisation de l'information, définition de modèle de profil.*

Krir, H. M. (2012). *Un système de recherche d'information personnalisée basé sur la modélisation multidimensionnelle de l'utilisateur.* Thèse de doctorat, Université Paris Sud – Paris XI, France.

Kumar, H., Lee, S., & Kim, G. (2014). Exploiting Social Bookmarking Services to Build Clustered User Interest Profile for Personalized Search. *Information Sciences , 281*.

Lieberman, H. (1997). Autonomous interface agents. *Proceedings of the SIGCHI conference on Human Factors in Computing Systems.* New York, USA.

Liu, F., Yu, C., & Meng, W. (2004). Personalized Web Search For Improving Retrieval Effectiveness. *IEEE Transactions on Knowledge and Data Engineering , 16* (1).

Ma, Z., Pant, G., & Sheng. (2007). Interest-based personalized search. *ACM Transactions on Information Systems* .

Maron, M. E., & Kuhns, J. L. (1960). On relevance, probabilistic indexing and information retrieval. *Journal of the ACM (J. ACM) , 7* (3).

McGill, M., & Salton, G. (1983). *Introduction to Modern Information Retrieval.* New York: McGraw-Hill.

Menegon, D., Mizzaro, S., Nazzi, E., & Vassena, L. (2009, Avril). Benchmark evaluation of context-aware Web search. (Springer, Éd.) Toulouse, France.

Mobasher, B., Sieg, A., & Burke, R. (2007). Web search personalization with ontological user profiles. *CIKM'07 : Proceedings of the sixteenth ACM conference on information and knowledge management.* New York, NY, USA.

Mohamed, S., & Abdelmoty, A. I. (2017). Spatio-semantic user profiles in location-based social networks. *International Journal of Data Science and Analytics , 4*.

Moulahi, B. (2015). *Définition et évaluation de modèles d'agrégation pour l'estimation de la pertinence multidimensionnelle en recherche d'information.* These de dectorat, Université Toulouse III Paul Sabatier.

Noll, M. G., & Meinel, C. (2007). Web Search Personalization via Social Bookmarking and Tagging. *Proceedings of the 6th International Conference of the Semantic Web and 2nd Asian Conference on Asian Semantic Web Conference.*

Ogden, C. K., & Richards, I. A. (1923). The Meaning of Meaning.

Bibliographie

Ould Amer, N. (2020). *Recherche Sociale et Personnalisée d'Information.* Thèse de doctorat, Université de Grenoble Alpes, France.

Ould Amer, N., Mulhem, P., & Géry, M. (2016). Axiomatic Term-Based Personalized Query Expansion Using Bookmarking System. *Database and Expert Systems Applications - 27th International Conference, DEXA 2016.* Porto, Portugal.

Panjarkar, A. G., Shete, D. S., & Kahate, S. A. (2014). Personalized Web Search with User's Profile in Re-ranking. *International Journal of Science, Engineering, and Technology (IJSET) , 2* (8).

Rafa, T. (2023). *Combinaison d'un Profil Social et un Profil Situationnel pour la Recherche d'Information.* Thèse de doctorat.

Rafa, T., & Kechid, S. (2018). A Semantic-Based Personalized Information. *DEXA 2018 Workshops, CCIS 903.* Regensburg, Allemagne.

Raghavan, V., & Sever, H. (1995). On the reuse of past optimal queries. *Proceedings of the 18th annual international ACM SIGIR conference on Research and development in information retrieval.* NY, USA.

Raza, M. A., Mokhtar, R., & Ahmad, N. (2018). A survey of statistical approaches for query expansion. (S. Nature, Éd.) *Journal of Knowledge and Information Systems* .

Razmerita, L. (2009). *User modeling and personalization of advanced information systems.*

Robertson, s., & Sparck Jones, K. (1976). Relevance weighting for search terms. *Journal of The American Society for Information Science , 27* (3).

Rocchio, J. J. (1971). *The SMART retrieval system: Experiments in automated document processing.* (G. Salton, Éd.) Englewood Cliffs, NJ: Prentice-Hall.

Saidl'haj, L. (2007). Recherche Conceptuelle d'Information Modèle d'Indexation Mixte.

Saint Requier, A., Dupont, G., & Sébastien, A. (2010). *Évaluation d'outils de reformulation interactive de requêtes.* Université de Rouen, LITIS, France.

Salton, G. (1968). *A comparison between manual and automatic indexing methods.* Ithaca, NY, USA.

Salton, G. (1969). A comparison between manual and automatic indexing methods. *American Documentation* .

Salton, G. (1971). *The SMART Retrieval System: Experiments in Automatic Document Processing.* (P.-H. Inc, Éd.) NJ (New Jersey).

Salton, G., & Allan, J. (1994). Automatic Text Decomposition and Structuring.

Salton, G., & Buckley, C. (1988). Term Weighting Approaches in Automatic Text Retrieval. *Information Processing and Management* .

Bibliographie

Saoud, Z. (2017). *Exploitation des réseaux sociaux pour la recherche d'information.* These de doctorat, USTHB (Université des Sciences et de la Technologie Houari-Boumediene), Alger, Algérie.

Saoud, Z., & Kechid, S. (2016). Integrating social profile to improve the source selection and the result merging process in distributed information retrieval. *Information Sciences , 336*.

Saracevic, T. (2007). Relevance: A Review of the Literature and a Framework for Thinking on the Notion in Information Science. Part III: Behavior and Effects of Relevance. *Journal of the American Society for Information Science , 58* (13).

Shafiq, O., & al. (2015). On personalizing web search using social network analysis. *Information Sciences , 314*.

Sieg, A., Mobasher, B., Burke, R., Prabu, G., & Ly, S. (2004). Using concepts hierarchies to enhance user queries in web-based information retrieval.

Smyth, B. (2011). Web search: Social & collaborative. *roceedings of 8th French Information Retrieval Conférenceen Recherche d'Infomations et Applica-tions.* Avignon, France.

Smyth, B. (2011). Web search: Social & collaborative. *Proceedings of the 8th French Information Retrieval Conference on Recherche d'Informations et Applications (CORIA).* Avignon, France.

Tamine, L., & Daoud, M. (2018). Evaluation in contextual information retrieval: Foundations and recent advances within the challenges of context dynamicity and data privacy. *ACM Computing Surveys , 51* (4).

Tamine, L., Boughanem, M., & Daoud, M. (2010). Evaluation of contextual information retrieval effectiveness: overview of issues and research. (Springer, Éd.) *Journal of Knowledge and Information Systems , 24* (1).

Tamine, L., Boughanem, M., & Zemirli, W. (2008). Personalized document ranking: Exploiting evidence from multiple user interests for profiling and retrieval. *ournal of Digital Information Management , 6* (5).

Tan, B., Shen, X., & Zhai, C. (2006). Mining long-term search history to improve search accuracy. *Proceedings of the 12th international conference on Knowledge Discovery and Data Mining.* University of Illinois at Urbana-Champaign.

Taylor, A. R., Cool, C., Belkin, V. J., & Amad, W. J. (2007). Relationships between Categories of Relevance Criteria and Stage in Task Completion. *Information Processing and Management , 43* (4).

Toms, E., Agosti, M., Fuhr, N., & Pe. (2013). *Evaluation Methodologies in Information Retrieval.* article.

Uddin, M. N., Duong, T., Sean, V., & Jo, G.-S. (2012). Construction of semantic user profile for personalized web search. Dans Springer (Éd.), *ICCCI 2012.*

Bibliographie

Vallet, D., Cantador, I., & Jose, J. M. (2010). Personalizing Web Search with Folksonomy-Based User and Document Profiles. *Advances in Information Retrieval, 32nd ECIR 2010.* Royaume-Uni.

Vandenbussche, P. Y., & Teissèdre, C. (2011). Events Retrieval Using Enhanced Semantic Web Knowledge. *Workshop DeRIVE 2011, in conjunction with the 10th International Semantic Web Conference (ISWC 2011).*

Wahlster, W., & Kobsa, A. (1986). Dialogue-based user models. *Proceedings of IEEE , 74* (7).

Wen, J. R., Lao, N., & Ma, W. Y. (2004). Probabilistic model for contextual retrieval. Dans A. Press (Éd.), *Proceedings of the 27th Annual International ACM SIGIR 2004 Conference.*

Wong, S., Ziarko, W., & Wong, P. (1985). Generalized vector spaces model in information retrieval. *Proc. of the 8th ACM-SIGIR conference.* Montreal, Quebec.

Xu, S., & al. (2008). Exploring Folksonomy for Personalized Search. *Proceedings of the 31st ACM SIGIR'08.* Singapour.

Yang, C., & Salton, G. (1973). On the specification of term values in automatic indexing. *Journal of Documentation , 29.*

Yates, R. B., & Neto, R. (1999). *Modern Information Retrieval.* (A. W. ACM Press, Éd.)

Zargayouna, H. (2005). *Indexation sémantique de documents XML.* Thèse de doctorat, Univ Paris XI, France.

Zargayouna, H., Roussey, C., & Chevallet, J. P. (2015). Recherche d'information sémantique: état des lieux. *Traitement Automatique des Langues (TAL) , 56* (3).

Zemirli, N. W. (2008). *Modèle d'accès personnalisé à l'information basé sur les Diagrammes d'Influence intégrant un profil utilisateur évolutif.* Thèse de doctorat, Université Paul Sabatier de Toulouse III.

Zemirli, W. (2004). le développement d'un système de recherche d'information personnalisé intégrant le profil utilisateur.

Zhou, D., Lawless, S., & Wade, V. (2012). Improving Search via Personalized Query Expansion Using Social Media. *Information Retrieval , 15.*