REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université DR.YAHIA FARES de MEDEA

Faculté des sciences

Département de Mathématique et d'Informatique



Mémoire de fin d'étude

Pour l'obtention du diplôme de master

En : Informatique

Option : Ingénierie des Systèmes et Technologies du Web

Thème :

Conception d'un Profil Utilisateur pour la Personnalisation de la Recherche d'Information

Réalisé par :

- Melle. BACHENE Leila
- Melle. TOUBAL SEGHIR Nassima

Dirigé par :

Mr. RAFA Tahar

Année Universitaire : 2017/2018

Résumé

L'objectif de la recherche d'information personnalisée est d'identifier le réel besoin en information de l'utilisateur caché derrière sa requête et d'adapter les résultats de recherche à ce besoin.

Les systèmes de recherche d'information personnalisée (SRIP) utilisent ce qu'on appelle « *Profil d'Utilisateur* », qui est un ensemble de données explicites et implicites sur l'utilisateur et ses intérêts.

Dans ce travail, nous proposons une approche de personnalisation de recherche d'information basée sur un profil d'utilisateur qui combine le contexte thématique de l'utilisateur représenté par les termes pertinents de son historique de recherche et son contexte social représenté par les liens d'amitié sur le site social.

Le profil d'utilisateur est intégré dans le processus de recherche d'information pour améliorer la pertinence des résultats de recherche.

Mots clés :

Recherche d'information, recherche d'information personnalisée, profil utilisateur, profil utilisateur social, réseau social.

Abstract

The purpose of the personalized information retrieval is to identify the real user information need hidden behind the user query and to adapt the search results to this need.

Personalized Information Retrieval systems use what is known as "User Profile", which is a set of explicit and implicit data about the user and his interests.

In this work, we propose a personalized information retrieval approach based on a user profile that combines the user's thematic context represented by the relevant terms of his search history and his social context represented by the friendship links on the social site.

The user profile is integrated into the information retrieval process to improve the relevance of search results.

Key words:
Information search, personalized information retrieval, user profile, social user profile, social network.

ملخص

الغرض من عملية البحث عن المعلومات الشخصية هو تحديد معلومات المستخدم الحقيقية المخبأة خلف طلبه وتكييف نتائج البحث مع هذه الحاجة.

تستخدم أنظمة استرداد المعلومات الشخصية (أنظمة الاسترجاع) ما يُعرف باسم "ملف تعريف المستخدم"، وهو عبارة عن مجموعة من البيانات الواضحة والضمنية حول المستخدم واهتماماته.

في هذا العمل، نقترح نهجًا شخصيًا لاسترجاع المعلومات استنادًا إلى ملف تعريف المستخدم الذي يجمع بين السياق الموضوعي للمستخدم والمتمثل في المصطلحات ذات الصلة من سجل البحث الخاص به وسياقه الاجتماعي الممثل بروابط الصداقة على الموقع الاجتماعي.

لقد تم دمج ملف تعريف المستخدم في عملية استرجاع المعلومات لتحسين ملائمة نتائج البحث.

كلمات البحث:

البحث عن المعلومات، البحث الشخصي عن المعلومات، ملف تعريف المستخدم، ملف تعريف المستخدم الاجتماعي، الشبكة الاجتماعية.

Remerciement

*Nous tenons à remercier en premier lieu le bon **Dieu, le tout Puissant** de nous avoir donné la force et le courage pour accomplir ce modeste travail.*

*Notre profonde gratitude et sincères remerciements vont à notre promoteur Mr **RAFA Tahar,** pour nous avoir confié ce travail, pour son suivi, sa disponibilité, ses orientations et ses conseils.*

Nos remerciements vont aussi aux membres de jury qui ont accepté de juger notre travail.

*A tous nos enseignants et les membres du département informatique de l'université **YAHIA FARES**.*

Nous n'oublions pas de remercier, vivement, tous ceux qui ont contribué, de près ou de loin à la réalisation de ce mémoire de fin d'étude.

Merci à tous

Dédicaces

Je dédie ce mémoire à :

Mes chers parents, que nulle dédicace ne puisse exprimer mes sincères sentiments, pour leurs patiences illimitées, leur encouragement contenu, leur aide, leur soutien et leurs prières tout au long de mes études, en témoignage de mon profond amour et respect pour leurs grands sacrifices.

Cher mari et sa famille : Mustapha, j'aimerai bien que tu trouve dans ce travail l'expression de mes sentiments de reconnaissance les plus sincères car grâce à ton aide et à ta patience avec moi que ce travail a pu voir le jour.

Mes chères sœurs : Nesrine , Amel, Fatima zahraa pour leurs encouragements permanents, et leur soutien moral.

Mes chers frères : Mohamed, Youcef, Roudwan, pour leur appui et leur encouragement

A toute ma famille pour leur soutien tout au long de mon parcours universitaire et à tous ceux que j'aime.

Mes meilleurs amis : Wafaa, Ilham, Nassima, Meriem, Imane.

A chaque cousins et cousines.

Aux jeunes enfants qui ont apporté joie et plaisir dans ma vie : Riyad, Ibtihel, Chaymaa , Achraf, Hadil.

Merci d'être toujours là pour moi.

Leila

Dédicaces

Je dédie ce travail qui n'aura jamais pu voir le jour sans les soutiens indéfectibles et sans limite de mes chers parents qui ne cessent de me donner avec amour le nécessaire pour que je puisse arriver à ce que je suis aujourd'hui. Leurs conseils ont toujours guidé mes pas vers la réussite. Leur patience sans fin, leur compréhension et leur encouragement sont pour moi le soutien indispensable qu'ils ont toujours su m'apporter.

Que dieux vous protège et que la réussite soit toujours à ma portée pour que je puisse vous combler de bonheur.

Je dédie aussi ce travail à mes chères sœurs : Dounia et Yasmine, ma chère famille sans exception spécialement à mes chers cousins et cousines.

A mon binôme et chère amie : Leila.

A tous mes amies, particulièrement : Manel et Nesrine.

Tous mes professeur qui m'ont enseigné et tous mes collègues de promotion : 2017/2018.

Nassima



Table des matières

Introduction générale 12
Chapitre 1 : Recherche d'Information Classique 14
1. Introduction 15
2. La recherche d'information 15
2.1. Notions de base 15
2.2. Les principaux modèles de la recherche d'information 16
2.2.1. Le modèle booléen 16
2.2.2. Le modèle vectoriel 16
2.2.3. Le modèle probabiliste 17
3. Système de recherche d'information 17
3.1. Définition : 17
3.2. Processus de RI : 17
3.2.1. L'indexation : 18
3.2.1.1. Modes d'indexation 18
3.2.2. L'appariement requête/document : 20
3.3. Reformulation des requêtes 20
3.3.1. Reformulation manuelle 20
3.3.2. Reformulation semi-automatique 20
3.3.3. Reformulation automatique 20
4. Evaluation de SRI 20
4.1. Notion de pertinence 21
4.1.1. Pertinence système 21
4.1.2. Pertinence utilisateur 21
4.2. Processus d'évaluation des systèmes de recherche d'information 21
4.2.1. Collection de test 21
4.2.2. Critères et mesures d'évaluation d'un SRI 22
4.2.3. Campagnes d'évaluation 25
5. Limites de la recherche classique 26
6. Conclusion 26
Chapitre2 : La recherche d'information personnalisée 27

Table des matières

1. Introduction 28
2. De la RI classique vers la RI adaptative 28
   2.1. Nouveaux besoins et limites de la RI classique 28
   2.2. Reformulation de requête par réinjection de pertinence 28
3. L'émergence de la personnalisation 30
   3.1. C'est quoi un contexte ? 30
   3.2. Les différents contextes de recherche 30
      3.2.1. Contexte personnel 30
      3.2.2. Contexte social 31
      3.2.3. Contexte spatio-temporel 31
   3.3. La recherche d'information contextuelle 31
      3.3.1. Modélisation du contexte 31
      3.3.2. Lacune de RI contextuelle et le besoin à la personnalisation 33
4. Accès personnalisé à l'information 33
   4.1. Définition de système d'accès personnalisé à l'information 33
   4.2. Profil utilisateur 33
      4.2.1. Définition de profil utilisateur 33
      4.2.2. Modélisation de profil 34
         4.2.2.1. Modèles de représentation de profil 34
         4.2.2.2. Construction du profil 35
         4.2.2.3. Exploitation du profil 36
         4.2.2.4. Evolution de profil 36
5. Evaluation des systèmes de recherche personnalisés 36
   5.1. Problèmes de l'évaluation d'une tache d'accès personnalisé 37
   5.2. Processus d'évaluation pour l'accès personnalisé 37
      5.2.1. Mesures d'évaluation 37
   5.3. Collection de test 39
6. Conclusion 39

Chapitre 3 : Contribution 40

1. Introduction 41
2. Profil d'Utilisateur 41
   2.1. Les données personnelles 42
   2.2. Historique de recherche 43
   2.3. Base de relations de cooccurrence des termes pertinents 44
   2.4. Liste des amis les plus importants 45

Table des matières

3. Processus de recherche 46
3.1. Sélection de la base de cooccurrence la plus similaire à la requête 47
3.2. Reformulation de requête 47
3.3. Reclassement des résultats 48
3.4. Mise à jour de profil 48
4. Evaluation et Implémentation 49
4.1. Stratégie d'évaluation 49
4.2. Expérimentation 50
4.2.1. Construction des profils et requêtes 50
4.2.2. Résultats et discussion 53
4.3. Implémentation 60
4.3.1. Organisation physique des composantes de profil 60
4.3.2. Présentation de la base de données 60
4.3.2.1. Présentation de MySQL 60
4.3.2.2. Création des tables de base de données 60
4.3.2.3. Liste des tables de la base de données 61
4.3.3. Environnement de développement de l'application 63
4.3.4. Présentation des interfaces 64
5. Conclusion 68
Conclusion et perspectives 70
Références 71

Table des figures

Figure 1. 1: processus en U de la recherche d'information. 17
Figure 1. 2: partition de la collection pour une requête. 23
Figure 1. 3: forme générale de la courbe précision / rappel d'un SRI. 23

Figure 2. 1: processus de réinjection de pertinence. 29

Figure 3. 1: Modèle de profil utilisateur. 42
Figure 3. 2: Modèle de données personnelles. 43
Figure 3. 3: Modèle de l'historique de recherche. 44
Figure 3. 4: Modèle de la base des relations de cooccurrence des termes pertinents. 45
Figure 3. 5 : Modèle de la liste des amis. 46
Figure 3. 6: Le processus de recherche d'information. 47
Figure 3. 7 : représentation des résultats de la requête "bon plan" avant la reformulation. 57
Figure 3. 8 : représentation des résultats de la requête " bon plan " après la reformulation. 57
Figure 3. 9 : représentation des résultats de la requête " France vs Allemagne " avant la reformulation. 58
Figure 3. 10 : représentation de résultats de la requête " France vs Allemagne " après reformulation. 58
Figure 3. 11 : comparaison des résultats des 2 scénarios. 59
Figure 3. 12 : MCD du profil utilisateur. 60
Figure 3. 13 : Les tables de la base de données. 61
Figure 3. 14 : exemple de la table «extrait ». 62
Figure 3. 15 : exemple de la table «profil ». 62
Figure 3. 16 : exemple de la table «termes_pertinents ». 63
Figure 3. 17 : Page d'inscription. 64
Figure 3. 18 : Page d'identification. 64
Figure 3. 19 : Page de recherche utilisateur. 65
Figure 3. 20 : Résultats de recherche en mode classique. 65
Figure 3. 21 : L'étape N°1 de la reformulation de la requête. 66
Figure 3. 22 : L'étape N°2 de la reformulation de la requête. 66
Figure 3. 23 : Résultats de recherche après la reformulation. 67
Figure 3. 24 : Historique de recherche d'un utilisateur. 67
Figure 3. 25 : La liste des amis proches. 68
Figure 3. 26 : Profil d'un utilisateur. 68

Liste des tableaux

Tableau 1. 1: variantes de Fréquence de Terme. 19

Tableau 3. 1: Résultats expérimentaux des requêtes qui acceptent la reformulation pour chaque utilisateur. 55
Tableau 3. 2: Résultats expérimentaux des requêtes qui n'acceptent pas la reformulation pour quelques utilisateurs. 56
Tableau 3. 3: la moyenne des précisions des résultats obtenus sur l'ensemble des requêtes envoyées par les utilisateurs. 59

Introduction générale

Le développement des technologies de l'information et des communications est l'une des plus grandes innovations de ce siècle, qui a pour conséquence la diversité, l'expansion et la multitude des ressources d'information. Ces derniers ont rendu assez difficile à l'utilisateur de trouver l'information pertinente dont il a besoin, car l'information est devenu beaucoup plus disponible mais moins pertinente.

Face à cette réalité, les systèmes de recherche d'information s'orientent vers l'intégration de l'utilisateur dans le processus de recherche d'information par la prise en compte de ses intérêts, on parle alors de la recherche d'information personnalisée, dont l'objectif est de sélectionner à travers une collection de document ceux qui correspondent le mieux possible la requête de l'utilisateur.

Pour réaliser cet objectif les systèmes de recherche d'information personnalisée (SRIP) utilisent une structure appelée profil d'utilisateur. Ce profil est ensemble de données explicites et implicites sur l'utilisateur et ses intérêts, généralement déduites à partir de son historique de recherche.

Dans ce travail nous nous intéressons à la recherche d'information personnalisée (RIP). Nous proposons une approche de recherche basée sur la conception et l'exploitation d'un profil social d'utilisateur.

Nous combinons le contexte thématique de l'utilisateur, représenté par l'ensemble de termes pertinents extraits de son historique de recherche, avec son contexte social, représenté par les liens d'amitiés avec les autres utilisateurs de réseau social. Nous introduisons une vue pseudo-sémantique dans le profil d'utilisateur en valorisant les liens de cooccurrence entre les termes pertinents. Ces liens de cooccurrence sont exploités pour la sélection des termes les plus appropriés pour l'enrichissement de la requête utilisateur.

Notre objectif est de guider les nouvelles recherches d'information de l'utilisateur pour améliorer la pertinence des résultats de recherches, et cela à travers l'intégration de profil de l'utilisateur dans le processus de recherche. Nous visons d'améliorer les deux phases de processus de recherche : (*i*) la reformulation de la requête initiale de l'utilisateur, et (*ii*) l'adaptation des résultats des recherche aux intérêts de l'utilisateur.

Ce mémoire est constitué de trois chapitres, dans ce qui suit nous donnons une brève description de leurs contenus respectifs :

Le chapitre 1 présente les notions de base de la recherche d'information ainsi que les principaux modèles de RI. Puis nous passons à la définition des systèmes d'informations ainsi que leur évaluation, et nous terminons avec les limites de la recherche d'information classique.

Le chapitre 2 présente l'émergence de la personnalisation de la RI, où nous présentons les nouveaux besoins et limites de la RI classique qui ont conduit à l'émergence de la RI adaptative. Nous abordons ensuite la notion de contexte en RI ainsi que l'émergence de la RI contextuelle. Par la suite nous passons à l'accès personnalisé à l'information où nous avons fourni un état de l'art sur le profil utilisateur. Les modèles de représentation, la construction, l'exploitation et l'évolution du profil utilisateur sont aussi présentés. Nous passons ensuite à l'évaluation des systèmes de recherche personnalisée et nous finissons par citant quelques approches des SRIP.

Le dernier chapitre présente notre contribution, il sera deviser en deux parties. La première partie est réservée pour la conception de profil utilisateur ainsi que les résultats expérimentaux obtenus pour l'évaluation de notre approche de personnalisation selon une démarche d'évaluations par simulation. Et nous décrivons en deuxième partie l'application que nous avons développée à cet effet.

Chapitre 1 :

Recherche d'Information Classique

1. Introduction

La recherche d'informations (RI) est une branche de l'informatique qui s'intéresse à l'acquisition, l'organisation, le stockage, la recherche et la sélection d'informations [Salton, 1968].

Elle intègre des modèles et des techniques dont le but est de faciliter l'accès à l'information pertinente dans un corpus (composé de documents d'une ou plusieurs bases de données) pour des requêtes écrite dans un langage naturel, afin de satisfaire les besoins de l'utilisateur. L'essor du web a remis la RI à de nouveaux défis d'accès à l'information, à savoir retrouver une information pertinente dans un espace diversifié et de taille considérable et qui répond au besoin en information spécifique de l'utilisateur [Daoud, 2009].

Nous présentons dans ce chapitre dans une première partie les notions et les modèles de base de la recherche d'information, puis nous passons à la deuxième partie qui décrit le processus de recherche. Et nous terminons le chapitre par une évaluation des systèmes de recherche d'information.

2. La recherche d'information

2.1. Notions de base

- *Document :* on appelle document toute unité qui peut constituer une réponse à un besoin en information ou requête d'un utilisateur, présentés sous des formes variées (texte, image, vidéo, page WEB, etc.) [Zemirli, 2004].

- *Collection de documents :* la collection de document (corpus) constitue l'ensemble des informations exploitable et accessible.

- *Requête :* une requête constitue l'expression du besoin en information de l'utilisateur. Elle contient en général un ensemble de mots clés, qui peut être vue comme une description sommaire des documents ciblés par la recherche [Daoud, 2009].

- *Besoin en informations :* Le besoin d'information est une expression mentale d'un utilisateur, on distingue trois types de besoins utilisateurs :

  - *Besoin vérificatif* : l'utilisateur cherche à vérifier le texte avec les données connues qu'il possède déjà.
  - *Besoin thématique connu* : l'utilisateur cherche clarifier, au revoir ou à trouver de nouvelles informations dans un sujet et domaine connu.
  - *Besoin thématique inconnu* : pour ce type de besoin l'utilisateur cherche de nouveau concept ou de nouvelles relations hors des sujets ou domaine qui lui sont familiers [Ingwersen, 1992].

- *Pertinence :* une pertinence est la mesure de correspondance entre un document et une requête, on distingue plusieurs pertinences possibles entre un document et un besoin en

information, nous citons les quatre les plus importantes : *Pertinence algorithmique, Pertinence thématique, Pertinence cognitive et Pertinence situationnelle* [Buidguaguen, 2011].

- *Modèle de recherche :* il représente le modèle du noyau d'un système de recherche d'information SRI. Il permet de trouver l'ensemble des documents pertinents associés à une requête.

2.2. Les principaux modèles de la recherche d'information

Un modèle de RI a pour rôle de fournir une formalisation du processus de RI et un cadre théorique pour la modélisation de la mesure de pertinence. Un modèle de RI est formellement décrit par un quadruplet [**D**, **Q**, **F**, **R (q_i, d_j)**] où :

D : ensemble des représentants des documents de la collection,
Q : ensemble des représentants des besoins en informations,
F : schéma du support de représentation des documents, requêtes et relations associées.
R (q_i, d_j) : fonction d'ordre associée à la pertinence [Yates, 1999].

Il existe un grand nombre de modèle de RI, parmi les nous présentons les plus couramment utilisés : le modèle booléen, le modèle vectoriel et le modèle probabiliste.

2.2.1. Le modèle booléen

C'est le premier modèle de RI [Salton, 1971], il est basé sur la théorie des ensembles où un document est représenté par un ensemble de termes et une requête est un ensemble de mots avec des opérateurs booléens : AND (∧), OR (∨), NOT (¬).

L'appariement Exact est basé sur la présence ou l'absence des termes de la requête dans les documents.

$$\text{Appariement } (q, d) = \text{RSV } (q, d) = 1 \text{ ou } 0$$

2.2.2. Le modèle vectoriel

Proposé par Salton dans le système SMART [Salton, 1971], basés sur l'algébrique, plus précisément le calcul vectoriel. Ils englobent le modèle vectoriel, le modèle vectoriel généralisé, LSI (Latent Semantic Indexing) et le modèle connexionniste.
La fonction de calcul du coefficient de similarité entre chaque document 'D_j' et la requête 'Q' est calculé par l'une des fonctions suivantes :

-*Produit scalaire* : $\text{RSV } (q, d_j) = \sum_{i=1}^{T} q_i * d_{ij}$

-*Mesure de Jaccard* : $\text{RSV } (q, d_j) = \frac{\sum_{i=1}^{T} q_i * d_{ij}}{\sum_{i=1}^{T} q_i^2 + \sum_{i=1}^{T} d_{ij}^2 - \sum_{i=1}^{T} q_i * d_{ij}}$

-*Mesure de cosinus* : $RSV(q, d_j) = \frac{\sum_{i=1}^{T} q_i * d_{ij}}{\sqrt{\left(\sum_{i=1}^{T} q_i^2\right) * \left(\sum_{i=1}^{T} d_{ij}^2\right)}}$

Où :

d_{ij} : poids du terme t_i dans le document D_j,

q_i : poids du terme t_i dans la requête Q [Zemirli, 2004].

2.2.3. Le modèle probabiliste

Proposé par Robertson et Sparck Jones [Robertson, 1976], il tente d'estimer la probabilité et d'observer des évènements liés au document et à la requête. Le score d'appariement entre le document D et la requête Q, noté RSV (Q, D) est donné par : $RSV(Q,D) = \frac{P(R|D)}{P(\bar{R}|D)}$

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

3. Système de recherche d'information

3.1. Définition :

Un Système de Recherche d'Informations (SRI) est un système informatique qui permet de retourner à partir d'un ensemble de documents, ceux dont le contenu correspond le mieux à un besoin en information d'un utilisateur (c'est-à-dire les documents pertinents), exprimé à l'aide d'une requête [Zemirli, 2004].

3.2. Processus de RI :

Les différentes étapes de processus de Ri sont présentés par Le processus global de la RI (processus en U) (Voir ***Figure 1.1***).

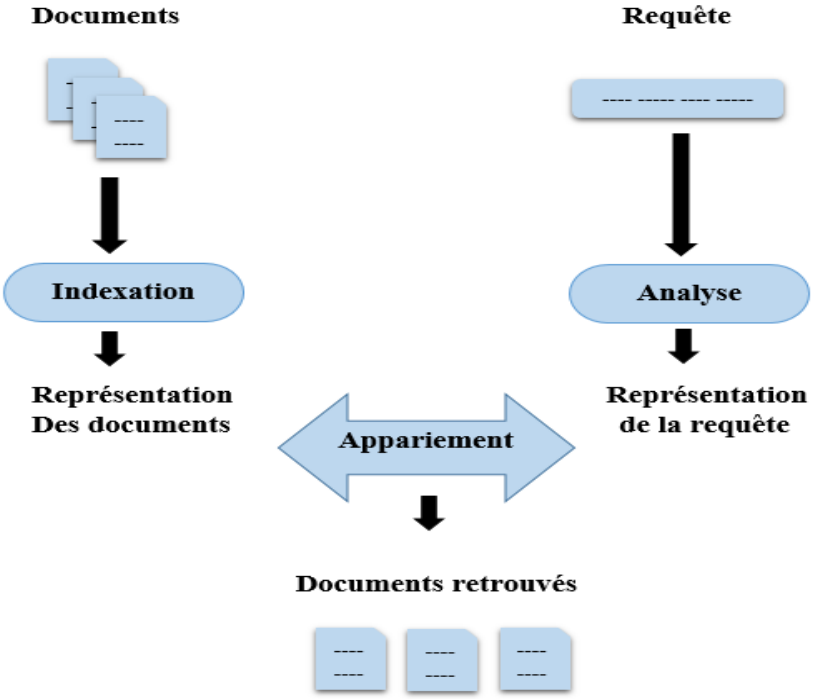


Figure 1. 1*: processus en U de la recherche d'information.*

L'objectif fondamental d'un processus de RI est de sélectionner les documents ''les plus proches'' du besoin en information de l'utilisateur décrit par une requête. Ceci induit deux principales phases dans le déroulement du processus : *indexation* et *appariement requête/documents* [Zemirli, 2004].

3.2.1. L'indexation :

L'indexation est un processus permettant de construire un ensemble d'éléments «clés» permettant de caractériser ou représenter le contenu d'un document pour faciliter de le retrouver en réponse à une requête. Ces éléments clés sont présentés selon une structure basée sur un ensemble de mots simples ou groupe de mots.
On distingue deux types d'indexation :

- ✓ *L'indexation libre* **:** les termes sont extraits du texte à indexer.
- ✓ *L'indexation contrôlée* **:** qui s'appuie sur un ensemble prédéfini de termes (*lexique, thésaurus, ontologie, réseau sémantique).* Elle consiste à sélectionner les termes de cet ensemble qui indexent ce texte.

3.2.1.1. Modes d'indexation

Il existe des différents modes d'indexation : *Manuelle*, *Automatique* ou *Semi-Automatique*.

- ✓ *Indexation manuelle :* chaque document est analysé par un spécialiste du domaine correspondant ou par un documentaliste, pour déterminer les mots clés qui lui semble les plus significatifs pour représenter le document. Elle est basée sur un vocabulaire contrôlé.

- ✓ *Indexation automatique :* c'est le SRI qui génère les indexes des documents. Chaque document est analysé à l'aide d'un processus entièrement automatisé. Elle détecte automatiquement les termes les plus représentatifs du contenu du document. Ce type d'indexation est actuellement la méthode la plus répandue.

- ✓ *Indexation semi-automatique :* le choix final reste au spécialiste du domaine correspondant ou documentaliste, qui intervient souvent pour établir des relations sémantiques entre mots-clés et choisir les termes significatifs.

Les techniques basiques d'indexation automatique se font selon les étapes suivantes :

- ***Extraction des termes (tocknenization)*** : Cette étape représente la segmentation des documents (convertir le texte de document en un ensemble de terme), est basée en générale sur la ponctuation et sur une liste de séparateur.

- ***Normalisation des termes extraits*** : elle contient plusieurs étapes, nous allons expliquer dans ce qui suit les étapes les plus importantes :
  - ✓ *Elimination des mots vides*
  - ✓ *La Racinisation*
  - ✓ *La lemmatisation*
  - ✓ *Extraction des mots composés*
  - ✓ *Extraction des entités nommées*

- ***Pondération des termes normalisés*** : cette étape est généralement basée sur des formules de pondération qui affecte à chaque terme un degré d'importance dans le document où il apparaît. Il existe un grand nombre de formules de pondération dont la plus connue est basée sur deux facteurs : fréquence de terme (TF) et fréquence inverse de document (IDF).

  - ✓ *Fréquence de terme (TF)* : La fréquence du terme (term frequency) est simplement le nombre d'occurrences de ce terme dans le document considéré. L'idée sous-jacente est plus un terme est fréquent dans un document plus il est important dans la description de ce document.
    La fréquence TF du terme dans un document peut être exprimée par l'une des variantes suivantes (voir ***Tableau 1.1***) :

Schéma de pondération	Formule TF
Binaire	0, 1
Fréquence brute	$f_{t,d}$
Normalisation logarithmique	$1 + \log(f_{t,d})$
Normalisation « 0.5 » par le max	$0.5 + 0.5 * \frac{f_{t,d}}{max_{\{t' \in d\}} f_{t',d}}$
Normalisation par le max	$k + (1-k) * \frac{f_{t,d}}{max_{\{t' \in d\}} f_{t',d}}$

Tableau 1. 1*: variantes de Fréquence de Terme.*

  - ✓ *Fréquence inverse de document (IDF) :* La fréquence inverse de document (*inverse document frequency*) est une mesure de l'importance du terme dans l'ensemble du corpus. Elle consiste à calculer le logarithme de l'inverse de la proportion de documents du corpus qui contiennent le terme.
    Cette mesure est exprimée selon l'une des déclinaisons suivantes :

$$IDF_i = log\frac{|N|}{n} \ , \ IDF_i = log\frac{|N-n|}{n}$$

Où n est la proportion des documents contenant le terme et N le nombre total de documents dans collection.

La fonction de pondération de la forme TF-IDF consiste à multiplier les deux mesures TF et IDF comme suit :

$$TF * IDF = \log(1 + TF) * log\frac{|N|}{n}$$

3.2.2. L'appariement requête/document :

L'appariement consiste à détecter les documents pertinents pour la requête posée, et calculer le degré de cette pertinence, appelé aussi score ou RSV (*Retrieval Statuts Value*). Ce score sert à trier la liste des documents retournés.

Il existe deux types d'appariement :

- ✓ *Appariement exact :* L'ensemble des documents respectant exactement la requête sont sélectionnés, mais pas ordonné.

- ✓ *Appariement approché :* Les documents sont sélectionnés selon un degré de pertinence vis-à-vis de la requête et sont ordonnés.

3.3. Reformulation des requêtes

La reformulation de requêtes est l'une des méthodes élaborées pour l'adaptation des SRI aux besoins des utilisateurs. C'est un processus ayant pour objectif de générer une nouvelle requête par ajout de termes significatifs et/ou ré-estimation de leur poids, suppression et/ou repondération des termes pour mieux représenter les besoins de l'utilisateur plus adéquate que celle initialement formulée par l'utilisateur de but d'avoir un SRI plus performant.

3.3.1. Reformulation manuelle

Elle consiste à présenter à l'utilisateur une liste de documents jugés pertinents en réponse à la requête initiale. C'est à l'utilisateur de sélectionner à partir des documents pertinents ceux dont lesquels le système va extraire les termes à rajouter à la requête initiale dans le but d'effectuer une nouvelle recherche.

3.3.2. Reformulation semi-automatique

Cette technique nécessite l'intervention de t'utilisateur qui doit identifier et sélectionner les documents pertinents et les documents non pertinents.

3.3.3. Reformulation automatique

L'extension de la requête est faite sans intervention de l'utilisateur grâce à l'utilisation d'un thésaurus contenant des informations de type linguistique (équivalence, association, hiérarchie) et statistique (pondération des termes).

4. Evaluation de SRI

L'évaluation est une étape importante dans la conception des SRI. Les premiers protocoles d'évaluation adoptés en RI se basent sur une approche de type laboratoire où (laboratory-based model) initié par Cleverdon [Cleverdon, 1967] dans le cadre du projet Cranfield Project II, et

qui constitue le cadre de référence dans lequel s'inscrivent les expérimentations et la validation des systèmes.
Cette approche fournit des ressources de base pour l'évaluation des SRI : une collection de requêtes, une collection de documents, des jugements de pertinence associés à chaque requête et des mesures d'évaluation basées essentiellement sur le rappel et la précision. elle est souvent adoptée dans les compagnes d'évaluation des SRI tells que TREC (Text REtrieval Conference), CLEF (Cross-Language Evaluation Forum), INEX (Initiative for the Evaluation of XML Retrieval–oriente), etc [Buidguaguen, 2011].
Pour atteindre un système de recherche idéal, nous devons atteindre les deux objectifs suivante : retrouver tous les documents pertinents et rejeter tous les documents non pertinents. Ces deux objectifs sont évalués par des mesures de précision et de rappel, qui ont pour but de permettre la comparaison des modèles entre eux ou la mise au point de leurs paramètres et ont été étendus pour permettre une analyse fine des performances de SRI.

4.1. Notion de pertinence

La pertinence représente la question fondamentale posée lors de l'accès à l'information. Elle dépend notamment du centre d'intérêt ou du domaine d'application, mais également du moment, du lieu et du support que l'utilisateur a choisi pour accéder à l'information, et enfin du système qui délivre cette information [Zemirli, 2008].

Dans [Boughanem, 2008], ils ont défini la notion de pertinence comme " La correspondance entre un document et une requête, ou encore une mesure d'informativité du document à la requête". Elle est liée au jugement individuel d'un utilisateur (pertinence utilisateur) et elle est estimée par les SRI (pertinence système).

4.1.1. Pertinence système

L'évaluation par un système de l'adéquation entre des documents et une requête. Elle mesure le degré auquel le système est capable de retrouver et classer des documents appariés à une requête.

4.1.2. Pertinence utilisateur

Se traduit par des jugements de pertinence sur les documents fournis en réponse à une requête pour deux utilisateurs différents ayant soumis la même requête au SRI ne jugent pas de la même manière les réponses du système.

4.2. Processus d'évaluation des systèmes de recherche d'information

4.2.1. Collection de test

La collection ou corpus de test constitue le contexte d'évaluation, pour évaluer le degré de pertinence des réponses d'un système de recherche d'information à une requête, il est nécessaire de connaître l'ensemble des documents pertinents pour une requête donnée. Chaque collection de test est composée d'une collection de documents ou corpus de documents, d'une collection de requêtes ou corpus de requêtes et des jugements de pertinence des documents par rapport à ces requêtes.

- ✓ *Collection de requêtes* [Moulahi, 2015] **:** Aussi appelées "Topics", simule l'activité de recherche de l'utilisateur. Pour exploiter au mieux les caractéristiques de la collection de documents et avoir une évaluation assez objective, il est important de créer un ensemble de quelques dizaines de requêtes (TREC utilise de 25 à 50 Topics comme une norme du nombre de requêtes de test), et qui soient adéquates par leur longueur, les thèmes abordés, leur forme, etc. Les requêtes sont généralement artificielles formulées par des assesseurs qui participent à la compagne d'évaluation, mais elles peuvent aussi être de vraies requêtes extraites à partir de log de recherche Web comme c'est le cas pour la tâche Web de TREC.

- ✓ *Corpus de documents* **:** c'est un ensemble de documents sur lesquels les systèmes de RI posent des requêtes et récupèrent les documents pertinents. Il existe de très nombreux ensembles de documents en accès libre, notamment sur le Web. Le choix d'une collection ou autre dépend de la tâche de recherche que l'on veut évaluer, pour garantir une représentativité par rapport à la tâche. De même que la spécification du volume des collections de documents utilisées dans l'évaluation est relativement dépendante de la tâche de recherche impliquée dans le système de RI à évaluer, pour garantir une diversité des sujets et du vocabulaire [Moulahi, 2015].

- ✓ *Jugements de pertinence* [Moulahi, 2015] : Les jugements de pertinence indiquent pour chaque document du corpus s'il est pertinent, et parfois même à quel degré il l'est, pour chaque requête. Pour établir ces listes de documents pour toutes les requêtes, les utilisateurs (ou des testeurs simulant des utilisateurs) doivent examiner chaque document de la base de documents, et juger s'il est pertinent par rapport à une requête donnée. Dans les programmes d'évaluation tels que TREC, les collections de documents contiennent plus d'un million de documents, ce qui rend impossible le jugement exhaustif de pertinence. Cette technique est souvent utilisée dans les compagnes d'évaluation telles que TREC ou CLEF.

4.2.2. Critères et mesures d'évaluation d'un SRI

- **Rappel et précision**

Les mesures les plus utilisés pour l'évaluation d'un modèle de RI sont le rappel et la précision. Le principe de ces deux mesures est basé sur la connaissance apriori des documents pertinents de la collection d'une part, et d'autre part la partition de l'ensemble des documents restitués par le SRI en deux catégories : documents pertinents et documents non pertinents, deux facteurs permettent d'évaluer ce critère. Le premier est le taux de rappel, il mesure la capacité du système à sélectionner tous les documents pertinents de la collection. Le second est le taux de précision, il mesure la capacité du système à rejeter tous les documents non pertinents.
La ***figure 1.2*** présente la partition de la collection pour une requête :

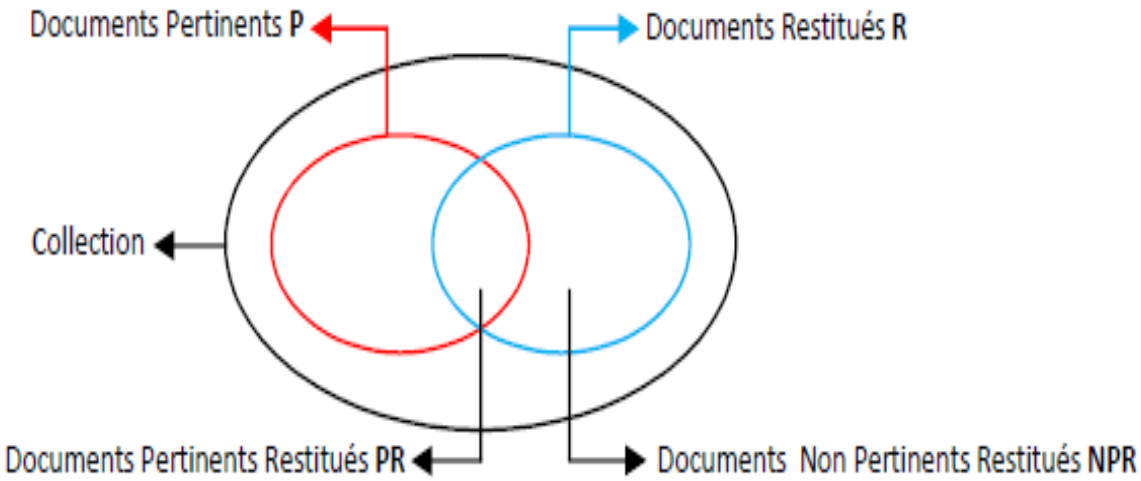


***Figure 1. 2**: partition de la collection pour une requête [Bouramoul, 2011].*

Les taux de précision et de rappel sont formulés comme suit :

$$Taux_rappel = R_d = \frac{Nombre\ de\ documents\ pertinent\ restitués}{Nombre\ de\ documents\ pertinents}$$

$$Taux_précision = P_d = \frac{Nombre\ de\ documents\ pertinents\ restitués}{Nombre\ de\ documents\ restitués}$$

Un SRI idéal est un système qui restitue tous les documents pertinents (rappel = 1), et tous les documents qu'il retrouve sont pertinents (précision =1) pour la requête de l'utilisateur. En pratique, cet idéal n'est jamais atteint puisque ces deux quantités évoluent en sens inverse. Intuitivement, si on augmente le rappel en retrouvant plus de documents pertinents, et si on diminue la précision en retrouvant aussi plus de documents non pertinents. Inversement, une plus grande précision risque de rejeter des documents pertinents diminuant ainsi le rappel [Boubekeur, 2008].

On représente la mesure rappel/ précision par une courbe qui a en générale l'allure suivante :

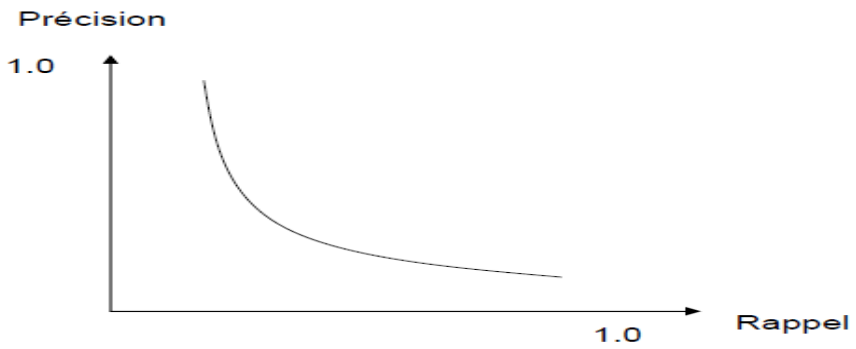


***Figure 1. 3**: forme générale de la courbe précision / rappel d'un SRI.*

- **Précision moyenne (Mean Average Precision-MAP)**

L'évaluation d'un système de recherche d'information est effectuée sur la base d'une collection de requêtes test. Dans ce cas, une valeur moyenne des valeurs de précision sur l'ensemble de requêtes |Q| peut être calculée comme suit :

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

Avec :
AP_q : est la précision moyenne d'une requête q,
Q : est l'ensemble des requêtes
|Q| : est le nombre de requêtes.
Cette mesure peut être qualifiée de globale puisqu'elle combine différents points de mesure. Elle est moins sensible au nombre de documents pertinents que les Pn. Cette précision peut être aussi calculée à différents niveaux de rappel (0%, 10%, 20%, … ,100%), elle est alors appelée : précision moyenne interpolée (MAiP). Pour chaque niveau de rappel, les valeurs calculées sont moyennées sur tout l'ensemble des requêtes.

- **Mesure de haute précision P@X**

C'est la précision à différents niveaux de coupe. Cette précision mesure la proportion des documents pertinents retrouvés parmi les X premiers documents retournés par le système. Les valeurs de X peuvent être fixées à 5, 10, 15, 30, ou 100 documents par exemple. Si la valeur de X est plus grande que le nombre total de documents retrouvés, les documents manquants sont considérés non pertinents. La valeur de haute précision correspondante est alors diminuée.

- **R-Précision**

Cette précision mesure la proportion des documents pertinents retrouvés après que R documents ont été retrouvé. R représente le nombre total de documents pertinents pour la requête. Cette mesure compense les limites des mesures de haute précision quand la précision est calculée pour X documents et que le nombre total de documents |P| est inférieur à X. Si la valeur de R est plus grande que le nombre total de documents retrouvés, tous les documents non retrouvés sont alors considérés non pertinents.

- **Mesure combiné**

C'est une mesure qui combine la précision et le rappel, est principalement motivée par les points suivants : [Korfhage, 1997].
- La difficulté de calcul du rappel maximal dans les collections volumineuses (dénombrement des documents pertinents à une requête),
- L'inadéquation de ces mesures dans le cas où la fonction d'appariement n'est pas une fonction d'ordre faible.
- La nécessité de combiner les deux aspects rappel/précision.
La principale mesure combinée est :

- ✓ *Mesure harmonique (F-mesure)*

Nommée F-mesure ou F-score. Van Rijsbergen [Rijsbergen, 1979] a introduit la F-mesure comme combinaison du rappel et de la précision et définie par :

$$F_\beta = \frac{(1+\beta^2).(précision * rapple)}{(\beta^2.précision + rappel)}$$

Où β traduit l'importance relative du rappel et de la précision. Dans le cas particulier où β=1, la F-mesure définit la moyenne harmonique du rappel et de la précision.

$$F = \frac{2}{\frac{1}{Rappel} + \frac{1}{Précision}}$$

Maximiser la F-mesure revient à trouver le meilleur compromis entre le rappel et la précision. La valeur maximale de la F-mesure est 1. Elle est obtenue quand tous les documents classés sont pertinents, et quand tous les documents pertinents ont été classés.

- **Autres mesures d'évaluation**

D'autres mesures d'évaluation d'un SRI existent. Ainsi, des mesures complémentaires au rappel et à la précision, ont été définies comme suit :

- ✓ La notion de silence et bruit présentent respectivement le taux de documents pertinents non sélectionnés par le système de recherche et le taux de documents non pertinents sélectionnés par le système de recherche, sont formulés comme suit :
  - $Bruit = 1 - précision$
  - $Silence = 1 - Rappel$
- ✓ L'indice de Fallout (ou Hallucination) peut être utilisé à la place du rappel. Il définit le pourcentage de documents non pertinents qui ont été retrouvés [Ishioka, 2003].
- ✓ L'élimination est la mesure complémentaire de Fallout. Elle définit le pourcentage de documents non pertinents non retrouvés [Boubekeur, 2008].

4.2.3. Campagnes d'évaluation

Les campagnes d'évaluations en RI permettent d'évaluer sur des collections différentes plusieurs SRI, afin de valider les différents modèles mis en œuvre, et comparer les systèmes. Les objectifs essentiels des campagnes sont les suivants :

- Encourager la RI sur de grandes collections fermées,
- Développer la communication entre l'industrie, l'académie et le gouvernement en mettant en place un forum ouvert pour faciliter les échanges d'idées sur la recherche,
- Augmenter la vitesse de transfert de la technologie du laboratoire de recherche aux enseignes commerciales,
- Rendre disponible et accessible des techniques d'évaluations appropriées pour les industriels et les académiciens [Hedin, 2004].

Les campagnes d'évaluations présentent l'avantage de cibler une tâche particulière et d'évaluer l'efficacité des systèmes répondant à cette tâche .Les campagnes d'évaluation les plus réputées sont :

- ✓ *TREC :* la conférence qui construit à ce jour l'initiative la plus importante dans le domaine de l'évaluation de SRI au regard de la dimension thématique et qui est l'une des campagnes les plus utilisées en RI, fournissant plusieurs collections de test.
- ✓ *CLEF :* pour l'évaluation de SRI multilingue basée sur les langues européennes.
- ✓ *INEX :* pour l'évaluation de SRI sur des documents XML par l'orientions des tâches de recherche vers des collections de documents structurés.

5. Limites de la recherche classique

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 critères spécifiques de l'utilisateur qui a émis la requête ou du contexte de recherche [Daoud, 2009]. Afin d'améliorer l'activité de RI, ces limites ont fait l'objet d'un nouveau paradigme en RI, les systèmes doivent donner des résultats répondant au mieux aux besoins de l'usager.

6. Conclusion

Au cours de ce chapitre nous avons présenté les principaux concepts et notions de la RI classique. Nous avons particulièrement introduit les notions de base, telles que le document, la collection de documents, la requête, le besoin en informations, la pertinence et le modèle de recherche. Nous avons aussi décrit les processus de base de la RI, à savoir l'indexation et l'appariement requête- document. Ensuite, nous avons étudié les différents modèles de la RI. Enfin, nous avons traité l'évaluation des systèmes de recherche informations.

À travers les différentes sections que nous avons présenté, nous concluons que la limite majeur de la plupart des modèles de la recherche classique est qu'ils retournent la même liste des résultats pour une même requête soumise par des utilisateurs étant dans des contextes ou des situations de recherche différentes.

Pour cela les techniques de la RI classique ont été développées vers la RI adaptative dans le but de remédier les problématiques rencontrés. Il s'agit du développement de techniques de reformulation de la requête et de désambiguïsation du besoin derrière les requêtes [Buidguaguen, 2011].

Plus récemment, les travaux en RI se sont orientés vers une nouvelle génération basée sur l'accès contextuel à l'information. Et pour mieux améliorer l'accès à l'information, les SRI ont orientés vers la recherche personnalisée.

Dans le chapitre suivant nous abordons la recherche d'information personnalisée.

Chapitre2 :

La recherche d’information personnalisée

1. Introduction

Au cours de ce chapitre, nous allons présenter l'émergence de la RI personnalisée, où nous décrivons l'évolution de la RI classique vers la RI adaptative. Plus particulièrement, nous abordons les problématiques majeures de la RI classique en présence du contexte et les techniques de la reformulation des requêtes.

Ensuit nous essayons de définir la notion de contexte, puis nous allons également donner un aperçu sur la recherche d'information contextuelle ainsi que les principaux techniques de modélisation du contexte en RI.

Nous allons aussi présenter en particulier les systèmes de recherche d'information personnalisés, et comme l'objectif de la RI personnalisée est d'intégrer le profil utilisateur dans le processus de recherche, nous allons présenter les différents aspects d'un profil utilisateur : sa définition, sa modélisation (notamment les modèles de représentations), les techniques de construction ainsi que l'évolution et l'exploitation de ce profil. Et nous terminons avec les cadres d'évaluation adaptés à un système de RI Personnalisé.

2. De la RI classique vers la RI adaptative

2.1. Nouveaux besoins et limites de la RI classique

La recherche d'information classique se base principalement sur l'appariement des descripteurs des documents et des requêtes pour la sélection des informations pertinentes.

La limite majeur de la RI classique est qu'elle suppose que l'utilisateur est complètement représenté par sa requête, et que les résultats retournés pour la même requête sont identiques même si elle est exprimée par des utilisateurs différents, car la pertinence entre un document et une requête est estimé par la fréquence d'apparition des mots de la requête dans ces mêmes documents.

Les premiers travaux en RI ayant essayé de trouver des solutions à ce type de problème ont donné naissance à la recherche d'information adaptative.

La RI adaptative est basée sur l'amélioration de l'efficacité du processus de recherche notamment lors de la phase d'exécution de la requête. Les techniques développées ont eu pour but de reformuler et désambiguïser le sens des mots de la requête soumise par l'utilisateur, afin de mieux répondre à son besoin en information.

Nous détaillons dans ce qui suit l'une des approches les plus naturelles et ayant du succès, est la reformulation par réinjection de la pertinence (relevance feedback).

2.2. Reformulation de requête par réinjection de pertinence

L'une des techniques les plus répandues en RI, communément appelée, *Relevance Feedback* (RF). Elle consiste à extraire à partir des documents jugés pertinents par l'utilisateur, les mots clés les plus expressifs et les ajouter à la requête initiale afin d'affiner la recherche. La réinjection de la pertinence pourrait intervenir à deux niveaux : reformuler la requête ou bien réordonner les résultats renvoyés.

Le processus de réinjection de pertinence comporte principalement trois étapes, (voir ***Figure 2.1***).

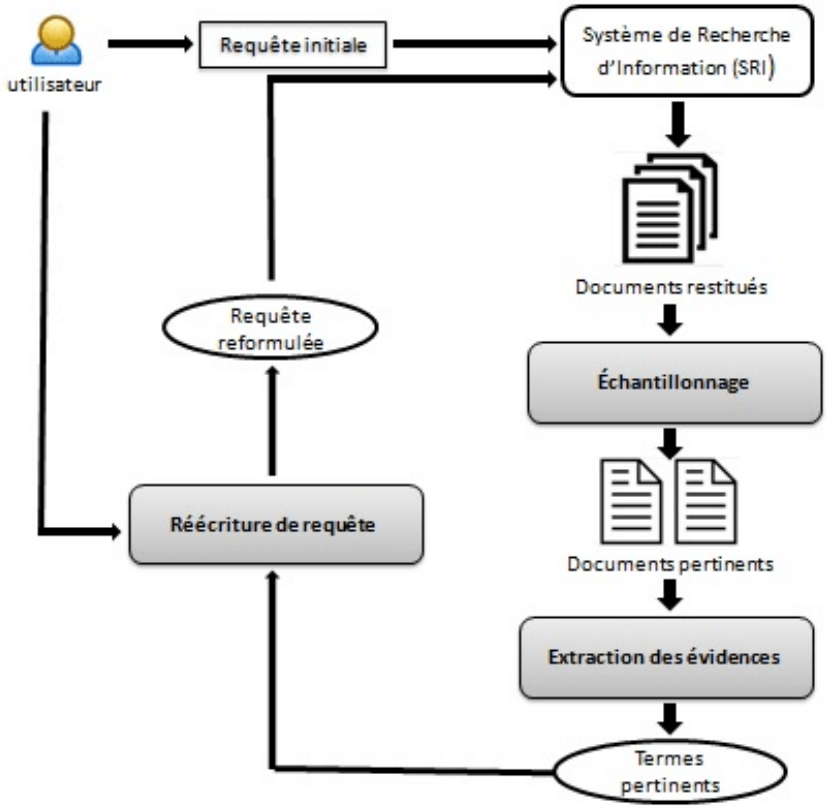


Figure 2. 1*: processus de réinjection de pertinence [Naouar, 2017].*

- ✓ *L'échantillonnage* : cette étape permet de construire un échantillon de documents à partir des éléments jugés par l'utilisateur. Cet échantillon est caractérisé par le nombre d'éléments jugé et le nombre d'éléments jugés pertinents [Bouramoul, 2011].

- ✓ *L'extraction des évidences* est l'étape la plus importante, elle consiste en général à extraire les termes pertinents qui serviront à l'enrichissement de la requête initiale. Plusieurs approches ont été développées. L'approche la plus reconnue est celle de Rocchio [Rocchio, 1971] adaptée au modèle vectoriel, où la nouvelle requête est générée en tenant en compte la distribution des termes dans les documents pertinents et non pertinents retournés, ou générée selon la comparaison entre la distribution de termes dans les documents retournés et l'ensemble de documents de la collection.

- ✓ *La réécriture de la requête* consiste à construire une nouvelle requête en combinant la requête initiale avec les informations extraites dans l'étape précédente [Lobna, 2007].

3. L'émergence de la personnalisation

Malgré le gain de performance apporté par les techniques développées en RI adaptative, elles présentent toute fois des limites. Ces dernières sont principalement liées à la représentation limitée du contexte de l'utilisateur et ne permettent pas d'adapter le processus de recherche à des besoins en information évolutifs tels que les centres d'intérêts.

Afin de combler ces lacunes et de proposer des systèmes répondant plus précisément au besoin utilisateur, le domaine de la RI contextuelle est apparu récemment comme une priorité. L'objectif de la RI contextuelle est de placer l'utilisateur au cœur des modèles en rendant explicites certains éléments du contexte qui peuvent influencer sur les performances des systèmes.

3.1. C'est quoi un contexte ?

Les premiers travaux d'Ingerwersen [Ingerwersen, 1996] et Saracevic [Saracevic, 1997], définissent la notion de contexte comme un ensemble des facteurs cognitifs et sociaux ainsi que les buts et les intentions de l'utilisateur au cours de sa recherche.

Par la suite, plusieurs définitions du contexte sont proposées par la littérature, elles diffèrent essentiellement par ses éléments constitutifs et sa portée à court ou à long terme.

La prise en compte du contexte dans les SRI implique à la fois d'identifier puis de modéliser les différents aspects du contexte utiles : *L'utilisateur*, *L'environnement de recherche* et *Le système*, pour la spécification du besoin de l'utilisateur et de les intégrer dans les méthodes et processus de RI.

- ✓ *L'utilisateur* : il peut être caractérisé par son identité, ses connaissances, ses intentions lors de recherche, et sa perception de la pertinence, etc.
- ✓ *L'environnement de recherche* : lié aux facteurs cognitifs, sociaux ou professionnels qui influencent le comportement de recherche de l'utilisateur et sa perception de la pertinence.
- ✓ *Le système* : parmi ses caractéristiques l'interface de visualisation concerne l'interaction utilisateur-système qui met en évidence l'impact des situations ou de l'environnement sur la rétroaction ou les jugements de pertinence de l'utilisateur.

3.2. Les différents contextes de recherche

D'après notre revue sur la littérature de RI, nous y distinguons trois principales dimensions du contexte abordées en RI contextuelle : *le contexte personnel*, *le contexte social* et *le contexte spatio-temporel*.

3.2.1. Contexte personnel

Est l'une des deux sous-dimensions du contexte de l'utilisateur (c'est une dimension qui permet de modéliser et de stocker les données caractérisant l'utilisateur), elle comprend à son tour les trois sous-dimensions suivantes :

- ✓ *Contexte démographique* **:** inclut des facteurs de préférence personnelles tels que la langue et le sexe, exploitées d'une manière de personnalisée la recherche pour des besoins spécifiques [Buidguaguen, 2011].
- ✓ *Contexte psychologique* **:** comprends des facteurs psychologiques tels que l'anxiété et la frustration, qui ayant un impact sur le comportement de l'utilisateur notamment son jugement de pertinence.
- ✓ *Contexte cognitif* **:** porte particulièrement sur l'expertise et les centres d'intérêt de l'utilisateur à court terme (un domaine d'intérêt spécifique relatif à une session de recherche) ou à long terme (des domaines d'intérêt généraux de l'utilisateur).

3.2.2. Contexte social

Cette dimension accorde une importance particulière sur la communauté à laquelle appartient l'utilisateur telle que les amis, les voisins et les collègues. L'adaptation du processus de recherche d'information consiste essentiellement à considérer les préférences et les profils partagés par la communauté de l'utilisateur plutôt que ses préférences et profil personnels.

3.2.3. Contexte spatio-temporel

Cette dimension comprend deux sous-dimensions : la localisation géographique et le temps. L'adaptation selon cette dimension concerne particulièrement les applications où les informations ont une validité subordonnée au lieu et temps instanciés lors de l'activité de recherche d'information [Tamine, 2008].

3.3. La recherche d'information contextuelle

La RI contextuelle combine entre un ensemble de technologies et des connaissances liées à la requête et le contexte utilisateur, afin de fournir la réponse la plus appropriée (pertinente) à son besoin en information.

La recherche d'information contextuelle est une activité qui fait intervenir en grande partie l'utilisateur. Cette spécificité porte sur l'utilisateur est décrite par son profil et qui couvre ses centres d'intérêts et ses préférences.

Le but fondamental de la plupart des travaux actuels en RI contextuelle, portant particulièrement sur la modélisation du profil de l'utilisateur, est de représenter, construire puis faire évoluer ses besoins en information à court et moyen terme.

3.3.1. Modélisation du contexte

Un modèle de RI classique comprend fondamentalement, selon la vision système, un modèle de représentation des documents et requêtes ainsi qu'un modèle d'appariement requête-document. La vision utilisateur introduisant une nouvelle dimension liée à l'utilisateur et son contexte, y intègre alors en plus, un modèle du contexte de recherche, dûment représenté et/ou exploité comme ressource dans la modélisation de l'appariement requête-contexte-document

[Tamine, 2009]. Il n'existe pas de standard spécifiant quel type d'information nécessite d'être considérée pour modéliser le contexte. La modélisation de contexte réfère souvent à la représentation de contexte. Nous citons les différentes techniques de modélisation du contexte qui ont été proposées dans la littérature.

- ✓ *Modèles clé-valeur* **:** La modélisation la plus simple de l'information de contexte est donnée par : La clé (le nom logique) du contexte, la valeur du contexte. Cette approche est simple d'utilisation mais très limitée en termes d'expressivité et d'évolution du modèle, ne permet pas un raisonnement efficace sur le contexte et ne favorise pas la réutilisabilité.
- ✓ *Modèles à balises* **:** le contexte est présenté sous forme d'une structure de donnée hiérarchique composée de balises associées à des valeurs d'attributs. La plupart de ces modèles sont dérivés du modèle CompositeCapabilities/PréférenceProfiles(CC/PP)[1]. Ces modèles restent spécifiques à un domaine, limités à un ensemble d'aspects du contexte (localisation, environnement, etc.) et incapables de représenter des contextes complexes.
- ✓ *Modèle graphique* **:** Cette approche très connue permet la modélisation d'informations contextuelles selon un graphe conceptuel. Selon la littérature cette modélisation est meilleurs que les modèles clé-valeur et les modèles à balises car ils permettent de représenter les relations, et de résoudre l'ambiguïté dans l'information contextuelle.
- ✓ *Modèles orientés objets* **:** Une approche orientée objet favorise la réutilisabilité, l'encapsulation et l'héritage. Cette approche peut être utilisée pour représenter le contexte, en interne, au niveau du code de programmation .Ces modèles souffrent de deux inconvénients majeurs : le manque de capacités de raisonnement ainsi que la difficulté de validation des conceptions orientées objet en raison de l'absence de normes et de standards dédiés.
- ✓ *Modèles logiques* **:** ces modèles utilisent les faits, les expressions et les règles pour modéliser le contexte. En utilisant un processus de raisonnement ou un moteur d'inférence. Cette approche permet de générer des contextes de haut niveau à partir de contextes de bas niveau. Elle peut être utilisée, comme un supplément, pour améliorer d'autres approches.
- ✓ *Modèles ontologiques* **:** ces modèles utilisent des ontologies pour modéliser le contexte. Cette approche présente beaucoup d'avantages. En effet, il existe de nombreux standards (RDF, RDFS, OWL) et une panoplie d'outils et de moteurs de raisonnement qui peuvent être utilisés selon les besoins de chacun. L'inconvénient majeur de cette approche basée sur les ontologies est la complexité de la recherche et l'interrogation du contexte qui peut nécessiter beaucoup de temps lorsque la quantité des données augmente considérablement.

[1](CC/PP) : proposé par le W3C pour supporter la négociation de contenu entre un navigateur web etun serveur par des dispositifs de plus en plus évolutifs.

3.3.2. Lacune de RI contextuelle et le besoin à la personnalisation

La RI contextuelle émerge comme un domaine à part entière, elle pose des problématiques nouvelles allant de la modélisation du contexte jusqu'à la modélisation de la pertinence cognitive en passant par la modélisation de l'interaction entre un utilisateur et un système de recherche d'information (SRI). La personnalisation tente de répondre à ces problématiques, dans le but de mieux répondre aux attentes et besoins des utilisateurs, les travaux en RI s'orientent actuellement vers des approches dites personnalisation de l'accès à l'information en exploitant des caractéristiques informationnelles spécifiques de l'utilisateur dans les processus d'accès à l'information. Afin de mieux cerner cette évolution, nous discutons dans les sections suivante des principaux facteurs et problématiques ayants conduits à l'émergence de la RI personnalisée.

4. Accès personnalisé à l'information

4.1. Définition de système d'accès personnalisé à l'information

Un système d'accès personnalisé à l'information est un système qui intègre l'utilisateur, en tant que structure informationnelle, tout au long de la chaine d'accès à l'information. Il inclut des structure permettant de représenter l'utilisateur, ces structure traduisent essentiellement les centres d'intérêt, les préférences et le contexte de l'utilisateur.

L'objectif d'un tel système est de délivrer une information pertinente en fonction des caractères spécifiques de l'utilisateur. La personnalisation de l'information peut être alors vue comme un processus de définition, de modélisation, d'exploitation et d'évolution des profils d'un utilisateur afin de répondre à un besoin d'information exprimée par une requête [Tamine& al, 2005].

4.2. Profil utilisateur

4.2.1. Définition de profil utilisateur

Un profil utilisateur c'est toute structure qui permet de modéliser et de stocker les données caractérisant l'utilisateur, il doit représenter l'ensemble des centres d'intérêt, des préférences, des connaissances ou des habitudes de l'utilisateur.
Dans la RI la tache de modélisation de profil de l'utilisateur peut être caractérisée par trois éléments clés :

- *Durée de vie des centres d'intérêt :* selon ce critère on distingue les centres d'intérêt à court terme qui décrivent un intérêt de courte portée lié à la session de recherche en cours, et les centres d'intérêt à long terme qui traduisent en revanche des sujets d'intérêt récurrents de l'utilisateur.
- *Sources d'informations :* en cas d'une approche de modélisation explicite, la source peut être simplement l'utilisateur lui-même. Autrement, dans le cas d'une approche implicite, la source peut être l'historique des clicks, les données de navigation, les pages et les sites favoris, etc…

4.2.2. Modélisation de profil

Les systèmes de recherche d'information tendent principalement à modéliser l'utilisateur selon un profil puis à l'intégrer dans la chaîne d'accès à l'information, afin de mieux répondre à son besoin spécifique. Pour notre part, nous considérons que la modélisation de l'utilisateur ; dans le cadre d'un système d'accès personnalisé l'information, est un processus caractérisé par trois phases la modélisation, l'exploitation et dernièrement l'évolution de profile.

4.2.2.1. Modèles de représentation de profil

De très nombreux modelés de représentation du profil utilisateur ont été proposés dans la littérature pour aider les utilisateurs à faire face à la quantité croissante d'information disponible sur Internet, a pour objectif fondamental de représenter puis faire évoluer ses besoins en information à court et moyen terme. Parmi ces modèles on trouve : le modèle vectorielle, le modèle sémantique et le modèle multidimensionnel.

- ✓ *Modèle vectoriel :*

Il s'agit de l'une des représentations de profil la plus utilisée, en particulier pour sa simplicité, s'appuyant généralement sur le modèle vectoriel Salton [Salton 71] .Le profil est généralement formalisé comme des vecteurs ou classes de vecteurs de termes pondérés extraits des documents correspondants aux centres d'intérêt de l'utilisateur, Par exemple, un profil basé sur les mots-clés associés à l'utilisateur s'appuie sur une pondération TF-IDF issue du domaine de l'indexation en recherche d'information.

Cette représentation apporte l'avantage de la simplicité de mise en œuvre. Il reste aussi à résoudre le problème de l'ordonnancement des préférences et des centres d'intérêts de l'utilisateur. En effet, ces derniers sont très variés et n'ont pas le même degré d'importance pour chaque utilisateur. Il faut donc modéliser le profil utilisateur de façon à prendre en considération l'ensemble des paramètres représentant l'utilisateur [Zemirli, 2004]. On peut citer comme exemple des systèmes : Alipes, WebMate et Surfagent.

- ✓ *Modèle sémantique :*

La représentation du profil met en évidence, dans ce cas, les relations sémantiques entre informations le contenant (Les centres d'intérêts). La représentation est essentiellement basée sur l'utilisation d'ontologies ou des réseaux sémantiques probabilistes et apporte des solutions aux problèmes de dissémination et de synonymie.
Dans ce type d'approche, les centres d'intérêts de l'utilisateur sont appariés aux concepts des ontologies de domaines. Un profil est alors représenté en termes de concepts de l'ontologie intéressant l'utilisateur.

Ce type de représentation offre le double avantage de la structuration et de la représentation associative (relations entre les termes) permettant de considérer l' ensemble des aspects représentatifs du profil [Techuente, 2013].On peut citer comme exemple des systèmes : OBIWAN, SmartPush.

- ✓ *Modèle multidimensionnel :*

Le profil est structuré selon un ensemble de dimensions prédéfinies, représentées selon divers formalismes. A pour objectif de capturer toutes ces caractéristiques informationnelles de l'utilisateur. Leur représentation doit de ce fait être de manière départagée en plusieurs dimensions chaque une contient des informations spécifiques. Les propositions dans [Bouzeghoub et Kostadinov 2004] [Kostadinov, 2003], pour la représentation des profils ont défini huit dimensions : données personnelles, domaines d'intérêt, le centre d'intérêts, la qualité des résultats, la customisation, la confidentialité, le feedback, et les informations générales. Les auteurs de [Tchuente et al., 2013],proposent deux dimensions : une dimension utilisateur (dont les centres d'intérêts sont calculés à partir des activités propres de l'utilisateur) et une dimension sociale (dont les centres d'intérêts sont calculés à partir des activités de son réseau social). La dimension utilisateur et la dimension sociale sont indépendantes l'une de l'autre (calculs, dérivation, etc.).

Pour une application donnée, un utilisateur n'a pas besoin de toutes les dimensions ou sous dimensions ni de toutes les informations caractérisant une dimension. Un profil donné est donc une instanciation partielle de ce méta modèle en fonction des besoins de l'utilisateur, du type d'application et de l'environnement d'exécution de cette application [Zemirli, 2008].

4.2.2.2. Construction du profil

Une fois effectuer le choix de la représentation, l'approche de construction dépend fortement de la représentation choisie pour le profil utilisateur. Plusieurs techniques de représentation des centres d'intérêts constitutifs du profil de l'utilisateur dans les SRI existent, consiste en la collecte des informations qui le représentent et s'effectue soit de manière explicite, soit de manière implicite soit de manière hybride par l'extraction des intérêts de l'utilisateur lors de ses interactions avec le système.

- ✓ *L'approche explicite* : largement utilisée également dans les procédés de construction du profil, se basant sur les informations fournies directement par l'utilisateur via l'interface du système, par exemple, lorsque l'utilisateur visualise un document, il indique son opinion sur le degré de pertinence du document par rapport à sa requête.
- ✓ *L'approche implicite* : permet de collecter les données de l'utilisateur, en observant son comportement et son activité lors de l'utilisation d'un système d'accès à l'information, par exemple, la visualisation d'un document par l'utilisateur peut être interprétée comme une adéquation du document par rapport à sa requête. Ces activité sont souvent appelées les Traces d'usage. Ces traces représentent une suite d'actions effectuées par un utilisateur, elles sont souvent déduites à partir des fichiers log qui constituent une source importante des données de l'utilisateur.
- ✓ *Approche hybride* : Quelques systèmes ont choisi de combiner les deux précédentes méthodes pour obtenir une meilleure performance.

La démarche de collecte des données explicites ou implicites (les traces d'usage) dans le cadre d'un système de personnalisation doit veiller à la mise à jour de des informations collectées. Ainsi, la démarche de collecte relève plus d'efficacité si le profil utilisateur est enrichi au fil de ses activités dans le système. Pour remplir cet objectif, il est nécessaire de définir des solutions permettant l'évolution des modèles utilisateurs et leur extensibilité [KHEMIRI, 2015].

4.2.2.3. Exploitation du profil

L'exploitation du profil de l'utilisateur permet de désambiguïser la recherche, de répondre à des besoins en information récurrents de l'utilisateur, et d'éliciter son besoin en information à un niveau spécifique. Les défis majeurs à ce niveau consistent à représenter et faire évoluer le profil utilisateur à court terme pour répondre à des besoins en information spécifiques de l'utilisateur et à long terme pour répondre à des besoins en informations récurrents, puis à l'exploiter dans un modèle d'accès personnalisé à l'information.

- ✓ *Le profil à long terme* : est identifié et exploité dans le processus de RI pour des requêtes récurrentes initiant des sessions de recherche. Le profil à long terme le plus similaire à la requête est identifié selon une mesure de similarité basé-graphe et sert à l'initialisation du profil à court terme au début de la session de recherche.
- ✓ *Le profil à court terme* : est exploité dans le ré-ordonnancement des résultats des requêtes émises et groupées dans la même session que le profil [Daoud, 2009].

4.2.2.4. Evolution de profil

L'évolution des profils désigne leur adaptation à la variation des centres d'intérêt des utilisateurs qu'ils décrivent, et par conséquent, de leurs besoins en information au cours du temps. Ce qui permet de distinguer les besoins à court terme, construits à partir de la session d'interaction courante, et des besoins à long terme qui sont une réelle représentation des centres d'intérêt persistants de l'utilisateur. Cela permet d'effectuer une évaluation préalable des connaissances de l'utilisateur (adaptation statique). Puis, vient l'évaluation dynamique des connaissances pour améliorer le processus de personnalisation.

L'évolution est davantage abordée comme un problème de représentation de la diversité des domaines d'intérêt de l'utilisateur en utilisant des techniques de classification ou heuristiques liées la notion de cycle de vie artificielle d'un centre d'intérêt.

5. Evaluation des systèmes de recherche personnalisés

Depuis le début des travaux sur la RI, l'évaluation des SRI est considérée comme un élément primordial pour l'évolution dans ce domaine. La démarche de validation en RI se base sur l'évaluation expérimentale des performances du modèle ou du système proposé, selon le modèle de Cranfield [Cleverdon, 1967].

Dans la plupart des protocoles d'évaluation des systèmes d'accès à l'information, les mesures sont construites à partir des jugements exprimés par des utilisateurs ou par des experts. Ces

mesures concernent plusieurs critères : le temps de réponse, la pertinence, la qualité et la présentation des résultats.
L'évaluation orientée vers l'utilisateur est une essentielle composante dans le cadre de l'accès personnalisé à l'information, dont le but est de mesurer l'adéquation des profils utilisateur construits par le système avec les centres d'intérêts effectifs de l'utilisateur.

5.1. Problèmes de l'évaluation d'une tache d'accès personnalisé

D'après [Toms et al, 2013 ; Khazri et al, 2008], la pertinence à améliorer dans un cadre de personnalisation de la recherche d'information est trop subjective, de fait qu'elle dépend en particulier de l'utilisateur qui est le seul qui peut juger cette pertinence selon son point de vue.

Cette réalité fait de l'évaluation d'une approche de recherche d'information personnalisée une tache très problématique, 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 la nouvelle approche pour l'utilisateur. Et d'après notre point de vue, l'exploitation des collections de tests avec des requêtes et pré-jugements de pertinence effectués par des experts, comme c'est fait dans la majorité d'approches de RIP, ne résout pas le problème car il ne s'agit pas de l'avis de l'utilisateur en question.

5.2. Processus d'évaluation pour l'accès personnalisé

Divers travaux ont tenté de mettre en place un cadre d'évaluation approprié aux SRIs 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, 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.

5.2.1. Mesures d'évaluation

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 4.2.2 chapitre 1) quelques autres mesures [Zemirli, 2008] :

- ✓ *La mesure RR (Relative Relevance)* : l'objectif de cette mesure est de considérer les différents types de pertinence dans l'évaluation d'un SRIC. Elle quantifie le degré de concordance entre les types de jugement de pertinence émis dans le cas de deux ensembles de jugements (soit R_1 correspond aux scores de pertinence algorithmique retourné par un SRI et R_2 à des scores de pertinence contextuelle

correspondant à un type de pertinence donnée : situationnelle ou thématique). La valeur de concordance entre R_1 et R_2 est généralement calculée par une mesure de cosinus.

A la différence de la mesure classique de précision, la mesure RR permet de considérer les différents types pertinences ; néanmoins, elle pose un problème mors de l'évaluation. En effet les scores de pertinence algorithmique ne sont pas étalonnés a la même échelle entre les différents SRI, ce qui rend la comparaison de mesure RR non significative.

- ✓ *Les mesures CG (Cumulative Gain) et DCG (Discount Cumulative Gain)* : l'objectif des mesures CG et DCG est d'estimer le gain de l'utilisation en terme de pertinence cumulés en observant les documents situés jusqu'à un rang donné. Les mesures sont définies comme suit :

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

Où G[i] est la valeur de pertinence associée au document de rang i.

$$CG[i] = \begin{cases} G[1], si\ i = 1 \\ CG[i-1] + \dfrac{G[i]}{log_q}, sinon \end{cases}$$

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.

- ✓ *La mesure GRP (Generalized Recall and Precision)* : la mesure GRP est une mesure qui généralise les mesures classiques de rappel et précision en considérant 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 SRI.

D est l'ensemble des documents de la collection.

r(*d)* est la valeur de pertinence graduelle associée au doc *d*.

5.3. Collection de test

La littérature (relativement récente) fait état de deux principaux protocoles de construction de collections de tests [Tamine&al., 2008] :

- ✓ *Réutilisation des collections de tests de TREC* (documents, requêtes et jugements de pertinence) puis leur augmentation par des éléments du contexte tels que l'historique des interactions.
  Le référentiel d'évaluation étant disponible, les mesures agrégées de rappel-précision sont alors exploitées pour évaluer les différences de performances entre les scénarios de recherche basique (ne tenant pas compte du contexte) et scénario de recherche contextuelle.

- ✓ *Construction de collections de tests en menant une campagne d'évaluation* c'est le protocole adopté par la plupart des travaux. Un ensemble d'utilisateurs est identifié et un ensemble de requêtes est construit. On collecte un volume de données-tests issu du croisement des jugements de pertinence pour chaque utilisateur spécifique et chaque requête. Les scénarios d'évaluation consistent alors, de manière classique, à comparer les performances de recherche d'un moteur de recherche classique et d'un moteur de recherche augmenté par la technologie basée sur l'exploitation des contextes des utilisateurs assesseurs de jugements. Dans le cas de l'utilisation de mesures agrégées de rappel-précision, un référentiel est généralement construit sur la base de l'ensemble des documents pertinents jugés par l'ensemble des utilisateurs pour chaque requête. Autrement, l'utilisation de mesures orientées rang évite l'utilisation d'un tel référentiel.

6. Conclusion

Ce chapitre introduit la thématique de recherche de la personnalisation d'information à travers les éléments clés de la personnalisation liés à l'entité utilisateur.

Tous les systèmes d'accès personnalisé s'appuient sur l'idée que la modélisation de l'utilisateur et son intégration dans le processus de recherche augmente sa performance, malgré les difficultés recensées lors du processus d'évaluation.

Le chapitre suivant sera consacrée à la représentation de nos contributions permettant d'intégrer l'utilisateur via son profil dans le processus de recherche.

Chapitre 3 :

Contribution

1. Introduction

Les premiers travaux en recherche d'information sont basés essentiellement sur l'appariement entre les documents et la requête utilisateur, indépendamment de l'utilisateur initiateur de la requête. Avec l'émergence des réseaux sociaux, l'utilisateur d'un système de recherche d'information n'est plus isolé. Son besoin en information peut être influencé par son contexte social [Bouhini, 2014]. Plusieurs travaux en RI personnalisée exploitent les informations sociales (annotations sociales [Saoud et Kechid, 2016 ; Bouhini, 2014], partage de documents [Benjabeur, 2010], relations d'amitié [Jeon et al., 2010], ….) pour décrire l'utilisateur et ses intérêts.

Dans notre travail, nous proposons une approche de personnalisation de la recherche d'information basée sur la construction d'un profil utilisateur. Dans ce profil nous combinons le contexte thématique de l'utilisateur, représenté par l'ensemble de termes pertinents extraits de son historique de recherche, avec son contexte social, représenté par les liens d'amitiés avec les autres utilisateurs de réseau social.

En plus des informations statiques de l'utilisateur, le profil se compose de deux principales composantes : (i) l'historique de recherche pertinent et (ii) la liste des amis les plus similaires à l'utilisateur.

Notre objectif est d'exploiter les informations sociales de l'utilisateur pour inférer ses futurs besoins en information et guider, par conséquent, ses nouvelles recherches d'informations. Le profil d'utilisateur sera intégré dans le processus de recherche d'information pour améliorer les performances des deux phases : (i) la reformulation de la requête et (ii) le reclassement des résultats.

Ce chapitre est organisé comme suit : nous présentons tout d'abord dans la section 2 le profil utilisateur proposé ainsi que ses composants, suivi par la présentation de processus de recherche dans la section 3. Ensuite nous présentons l'évaluation de notre approche dans la section 4 et nous terminons par la section 5 qui est à la conclusion et les perspectives de ce travail.

2. Profil d'Utilisateur

Un profil peut être défini [Kechid, 2009] 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 satisfait en tenant compte de son profil.

Le profil utilisateur peut être construit d'une part à partir d'activités propres à l'utilisateur d'une façon explicite (par l'utilisateur), ou implicite (par l'extraction des intérêts de l'utilisateur lors de ses interactions avec le système) ; et d'autre part par l'utilisation des sources externes comme des utilisateurs similaires ou bien à partir des membres dans son réseau social [Canut et al., 2016].

Dans notre cas, qui se situe dans le contexte de recherche d'information sociale (RIS), nous proposons un profil utilisateur social basé sur deux aspects : ***un aspect thématique*** (représente

l'ensemble des documents jugés pertinents par l'utilisateur, ainsi que les termes pertinents extraits à partir de ces documents), et ***un aspect social*** (représente l'ensemble des amis proches, qui partagent des intérêts communs avec l'utilisateur dans un environnement social).

Le profil social que nous proposons comporte quatre composantes : ***les données personnelles*** qui permettent d'identifier l'utilisateur, ***l'historique de recherche*** qui sert pour sauvegarder l'historique pertinent de l'utilisateur, ***la base de relation de cooccurrence des termes pertinents*** qui sert pour classifier les termes pertinents de l'historique de recherche selon les relations de cooccurrence entre eux et ***la liste d'amis les plus importants*** qui sert pour filtrer et classifier les amis les plus importants à l'utilisateur.

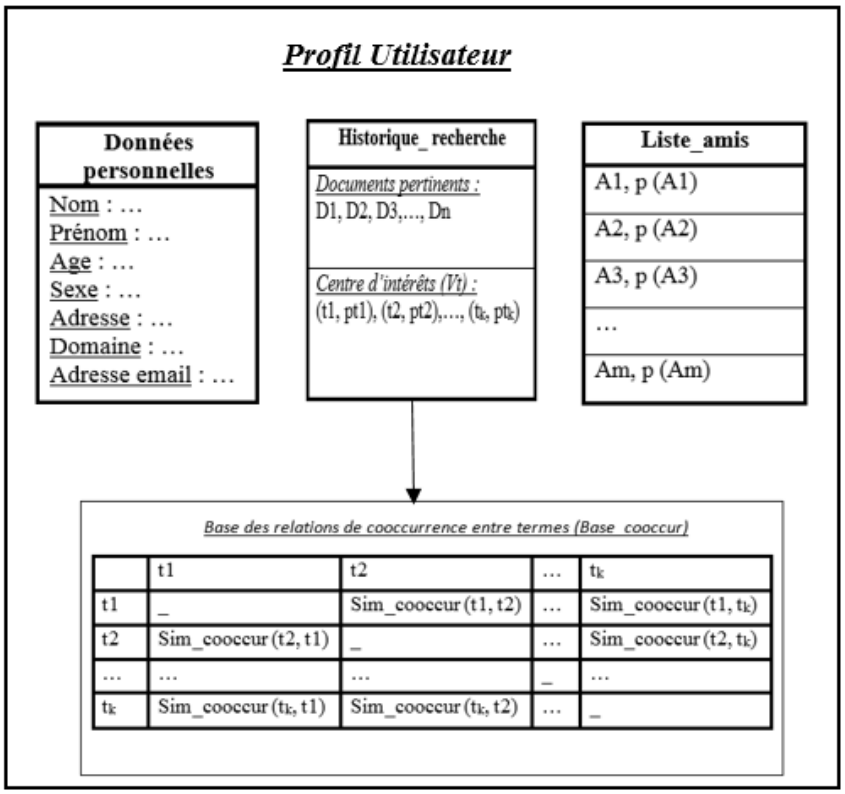


Figure 3. 1*: Modèle de profil utilisateur.*

2.1. Les données personnelles

Les données personnelles sont la partie statique du profil. Elles comprennent l'identité de l'utilisateur (nom, prénom), des données démographiques (âge, sexe, adresse), et d'autres informations comme l'adresse email de l'utilisateur et le domaine d'intérêts initial. Les données personnelles sont relativement stables dans le temps et ne demandent pas de mise à jour automatique par le gestionnaire de profil.

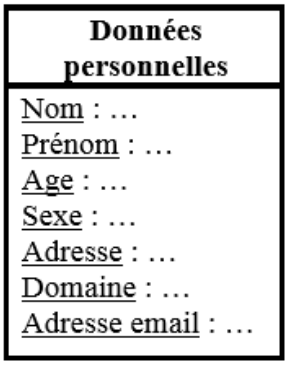


Figure 3. 2*: Modèle de données personnelles.*

2.2. Historique de recherche

Cette composante regroupe l'ensemble des informations collectées sur le comportement de l'utilisateur. Elle sert pour sauvegarder l'historique pertinent de l'utilisateur et elle comporte comme indique la ***Figure 3.3***, les deux sous-ensembles suivants :

(i) *L'ensemble de documents pertinents* : cet ensemble comporte les documents jugés pertinents par l'utilisateur (longuement visités, trop cliqués, sauvegardés, imprimés,…).

(ii) *Le centre d'intérêts de l'utilisateur* : il exprime le domaine d'expertise de cet utilisateur ou son périmètre d'exploration. Il existe plusieurs techniques de représentation des centres d'intérêts constitutifs du profil de l'utilisateur dans les SRI. Une représentation naïve des centres d'intérêts est à base de mots clés. Des techniques plus élaborées sont basées sur la représentation selon des vecteurs de mots clés [Tamine et al., 2008; Gowan, 2003],ou un ensemble de concepts [Sieg et al., 2004a; Liu et al., 2004; Daoud et al., 2008]. Pour nous, un centre d'intérêt est représenté par un vecteur des termes issus des documents pertinents de l'historique de recherche et qui évolue au cours des sessions de recherche, l'importance de chaque terme est définit par une pondération des mots clés du centre d'intérêt. Ces termes sont pondérés selon une formule *tf*idf* du modèle vectoriel [Salton, 1988]. Cette formule est définit comme suit :

$$p_t = (0.5 + 0.5 * tf/Max\ tf) * \log(N/n) \qquad (1)$$

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.

Cet ensemble des termes est sauvegardé dans un vecteur V_t de mots clés avec un poids associé à chaque mot comme suit :

$$V_t = \{(t_1, pt_1), (t_2, pt_2), \ldots, (t_i, pt_i)\} \qquad (2)$$

Où :

- ✓ t_i : représente le i^{eme} terme de centre d'intérêts.
- ✓ pt_i : représente le poids global de t_i par rapport l'historique de recherche.

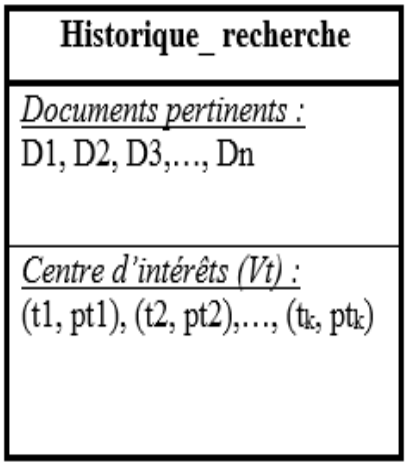


***Figure 3. 3**: Modèle de l'historique de recherche.*

2.3. Base de relations de cooccurrence des termes pertinents

La collection des termes pertinents dans le centre d'intérêts représente le premier niveau d'exploitation de l'historique de recherche dans la description de l'utilisateur et ses intérêts. Nous voulons passer à un deuxième niveau d'exploitation des termes pertinents du centre d'intérêts par la valorisation de leurs relations de cooccurrence. Nous considérons les relations de cooccurrence entre les termes comme une pseudo-sémantique qui permet de mettre en valeur l'apparition fréquente de deux termes pertinents dans des nombreux documents. Pour nous, l'apparition fréquente de deux termes dans nombreux documents indique qu'ils collaborent pour la description du même sujet. Par exemple, l'apparition de deux termes « Apple » et « smartphone » peut indiquer qu'il s'agit d'un sujet de TIC, tandis que l'apparition de « Apple » et « Tree » indique un sujet d'agriculture.

La valorisation des relations de cooccurrence entre les termes pertinents est représentée par une similarité, notée *sim_cooccur* (t_i, t_j), qui exprime la liaison entre chaque terme pertinent t_i avec les autres termes pertinents t_j, et elle est calculée selon la formule suivante :

$$sim_cooccur(t_i, t_j) = \frac{|d(t_i,t_j)|}{|d(t_i)|^2+|d(t_j)-d(t_i,t_j)|} \qquad (3)$$

Avec :

- ✓ $|d\ (t_i,t_j)|$: nombre de documents contenant t_i et t_j conjointement.
- ✓ $|d\ (t_i)|$: nombre de documents contenant t_i.
- ✓ $|d\ (t_j)- d\ (t_i, t_j)|$: nombre de documents contenant t_j et ne contenant pas t_i.

Nous constatons que cette similarité de cooccurrence n'est pas bidirectionnelle, (*sim_ cooccur* $(t_i,t_j) \neq$ *sim_ cooccur* (t_j,t_i)). En effet, il se peut qu'un terme '*x'* apparaisse dans la majorité des documents contenant le terme '*y'*, tandis que '*y'* n'apparait que dans une petite partie des documents contenant '*x'*. ceci traduit le fait qu'un terme peut être plus vaste qu'un autre.

La base de relations de cooccurrence des termes pertinents sert à sauvegarder ces similarités entre les termes pertinents de l'historique de profil utilisateur dans une base appelée (*Base_cooccur*).

Notre objectif d'après cette composante est de mettre en évidence les relations de cooccurrence entre les termes pertinents afin de pouvoir améliorer les résultats de recherche, et cela se fait dans deux phases :

- ✓ la sélection de la *Base_cooccur* la plus adéquate à la requête parmi les bases d'utilisateur et celles de ses amis proches.
- ✓ La reformulation de requête en ajoutant le terme qui maximise la similarité (3) avec les termes de requête.

Base des relations de cooccurrence entre termes (Base_cooccur)

	t1	t2	…	t_k
t1	_	Sim_cooccur (t1, t2)	…	Sim_cooccur (t1, t_k)
t2	Sim_cooccur (t2, t1)	_	…	Sim_cooccur (t2, t_k)
…	…	…	_	…
t_k	Sim_cooccur (t_k, t1)	Sim_cooccur (t_k, t2)	…	_

Figure 3. 4: *Modèle de la base des relations de cooccurrence des termes pertinents.*

2.4. Liste des amis les plus importants

Dans certains cas, le centre d'intérêts de l'utilisateur ne peut pas guider les nouvelles recherches (nouveau ou pauvre profil, changement d'intérêt, nouveau besoins,…). Pour régler ce problème les approches de la RIP demandent de l'aide auprès des amis les plus proches de l'utilisateur qui partagent les mêmes intérêts avec lui.

Dans notre cas, nous dotons notre profil d'utilisateur par une liste dynamique des amis les plus similaires à l'utilisateur. Le but de spécifier une liste des amis importants à l'utilisateur est d'aider à identifier le besoins en information qui peut être dérivé à partir des profils de ses amis.

Cette liste sert à sauvegarder les amis les plus importants (les plus similaires) par rapport à l'utilisateur. Ces amis sont représenté sous forme d'un vecteur d'amis V_A où chaque ami est associé à son poids comme suit :

$$V_A = \{(A_1, pA_1), (A_2, pA_2), \ldots, (A_m, pA_m)\} \quad (4)$$

Où :

- ✓ A_m : représente le i^{eme} ami important.
- ✓ PA_m : représente le poids d'ami Am.

L'importance d'un ami ainsi que son poids sont mesurés à travers une formule de similarité (*sim_ami*) proposée basée sur l'intersection des centres d'intérêts de l'utilisateur avec ceux des autres utilisateurs. La formule se présente comme suit :

$$sim_ami(u_i, u_j) = \frac{|\{V_t(u_i)\} \cap \{V_t(u_j)\}|}{|\{V_t(u_i)\}|} \quad (5)$$

Avec :

- ✓ $|V_t(u_i)|$: vecteur de centre d'intérêt de utilisateur u_i
- ✓ $|V_t(u_j)|$: vecteur de centre d'intérêt de utilisateur u_j
- ✓ $|\{V_t(u_i)\} \cap \{V_t(u_j)\}|$: nombre d'intersection des termes de centre d'intérêt de u_i et de u_j.

Liste_amis
A1, p (A1)
A2, p (A2)
A3, p (A3)
…
Am, p (Am)

***Figure 3. 5** : Modèle de la liste des amis.*

3. Processus de recherche

Notre processus de recherche consiste d'une part à reformuler la requête envoyée par l'utilisateur dans le but d'élargir le champ de recherche, et d'autre part au ré-ordonnancement des résultats pour obtenir en première place les meilleurs résultats correspondants aux intérêts de l'utilisateur selon son profil.

Ce processus est lancé à chaque nouvelle activité de recherche réalisée par l'utilisateur en vue de sélectionner la base de cooccurrence similaire à la requête courante, sur la base d'un calcul de similarité (6). Ensuite nous passons à la phase de reformulation de requête en ajoutant le terme *t* qui maximise la similarité (3) entre les termes de la requête et la base sélectionnée comme la plus adéquate, et pour mettre à la disposition de l'utilisateur les documents les plus pertinents, nous avons pensé à reclasser (ou bien réordonner) les résultats retourné par le moteur de recherche à travers une formule de similarité proposée.

Le scénario suivi par notre processus est représenté de la façon suivante :

1. Le système reçoit la requête utilisateur Q.
2. Le système sélectionne la base de cooccurrence (Base_cooccur) la plus adéquate à la requête parmi la base de l'utilisateur et celles de ses amis proches.
3. Reformulation de la requête Q selon la base sélectionnée.
4. Envoi de la requête reformulée Q' au moteur de recherche classique.
5. Reclassement des résultats retournés par le SRI classique selon leurs scores de similarité vis-à-vis la requête reformulée Q'.
6. Transmission de la nouvelle liste des documents réordonnée à l'utilisateur.
7. Le système observe les réactions et comportements de l'utilisateur pour mettre à jour son profil.

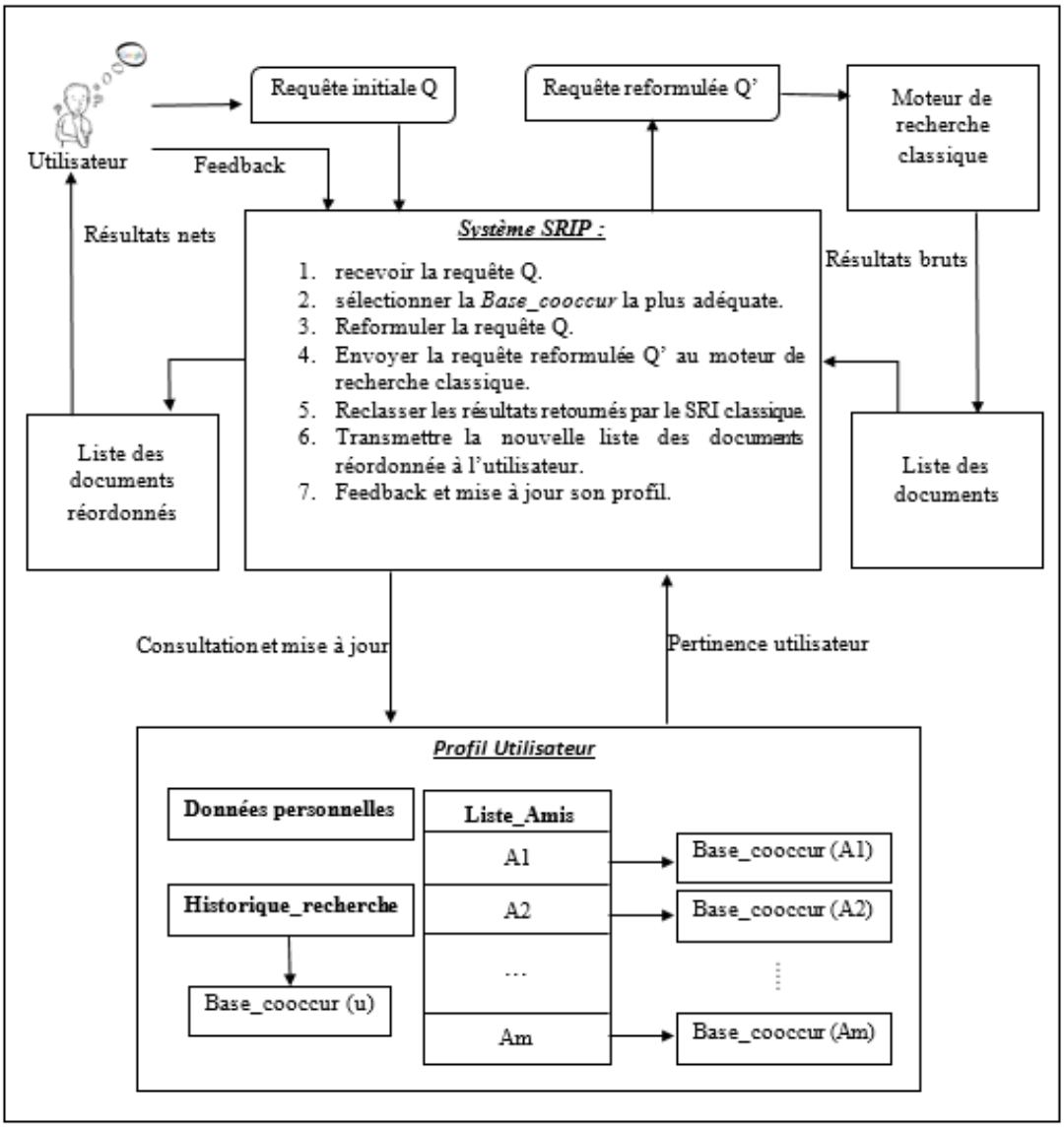


Figure 3. 6*: Le processus de recherche d'information.*

3.1. Sélection de la base de cooccurrence la plus similaire à la requête

Pour pouvoir sélectionner la base de cooccurrence (*Base_cooccur*) la plus adéquate parmi la base de l'utilisateur et celles de ses amis les plus importants, on calcule la similarité de ces différentes bases avec la requête selon la formule suivante :

$$sim\big(Q, Base_cooccur_j\big) = \sum_{t_i \in Q} \sum_{t_j \in Base_cooccur_j} sim_cooccur(t_i, t_j) \qquad (6)$$

La base retenue comme plus adéquate à la requête Q est celle qui maximise la similarité (6).

3.2. Reformulation de requête

La reformulation de la requête permet de ré-exprimer les besoins en information de l'utilisateur par l'injection de certains termes liés aux termes de la requête initiale Q pour la rendre moins ambiguë et élargir son champ de recherche. Nous définissons cette liaison entre les termes par une relation de cooccurrence déduite à partir de la base (*Base_cooccur)* sélectionnée.

Donc la requête original Q est enrichi pour produire une nouvelle requête Q', afin d'obtenir ainsi plus ou moins de documents avec un classement modifié, tel que :

$$Q' = Q+T \qquad (7)$$

Avec :

Q : la requête initiale.

Q' : la nouvelle requête reformulée.

T : le terme *t* qui maximise la relation de similarité avec les termes de la requête Q selon la formule (3).

3.3. Reclassement des résultats

Dans cette phase, nous pensons à reclasser les résultats de moteur de recherche classique, dans le but de personnaliser le classement de ces résultats selon chaque utilisateur pour mieux satisfaire son besoin en information, en lui donnant en première place les documents les plus pertinents.

Cette étape consiste à définir une fonction d'ordonnancement qui permet de quantifier la similarité entre les documents retournés et la requête reformulée Q'. Nous distinguons une fonction d'appariement qui calcule un score de similarité pour chaque document d_i appartient à la liste des *N* premiers documents retournés par le moteur de recherche (*List_N_doc*) par rapport la requête Q'. La formule de cette similarité *sim_re* (d_i, *Q'*) est calculée comme suit :

$$sim(d_i, Q') = log(N/rang(d_i)) * \sum_{t \in Q'} (tf_{di}/|d_i|) \qquad (8)$$

Avec :

- *t* : un élément de la requête reformulée *Q'*.
- tf_{di} : fréquence de l'élément *t* dans le document d_i.
- $|d_i|$: la taille de document d_i.
- *N :* nombre de documents considérés comme *List_N_Doc.*
- *Rang* (d_i) *:* classement initial de document d_i dans la liste *List_N_Doc.*

Et en fonction de ces scores, les documents sont reclassés en ordre décroissant, définissant ainsi la liste finale des documents les plus pertinents à mettre à la disposition de l'utilisateur.

3.4. Mise à jour de profil

Un profil utilisateur ne peut pas être considéré stable dans le temps, une mise à jour de profil est donc requise.

La mise à jour de notre profil (*profil utilisateur social*) doit assurer son enrichissement à partir du comportement social de l'utilisateur et son évolution au fur et à mesure que les données évoluent, comme les intérêts et les préférences. Dans notre contexte la technique d'enrichissement se fait selon l'addition des nouveaux termes pertinents ainsi que leurs poids dans la représentation du profil utilisateur après un traitement prédéfini. De plus l'espace

occupé et le temps de réponse pris par les différents calculs ,sont des critères très important qui permet de suivre l'évolution des intérêts de chaque utilisateur dans le système d'information.

Si le profil de l'utilisateur est enrichi au fil de ses activités dans le système avec les résultats retournés, la démarche de collecte de données explicites ou implicites doit veiller à la mise à jour pour identifier les nouveaux documents jugée pertinent. Nous devons faire une analyse syntaxique (segmentation) sur ces documents pour pouvoir identifier leurs termes pertinents en calculant leur poids selon la formule (1), et les insérés dans la base de cooccurrence en appliquant la formule (3). Ce qui concerne les amis de l'utilisateur, la liste des amis important est filtrée et reclassée selon la formule (5) dans chaque changement des termes de centre d'intérêt de ces amis.

Pour filtrer et alléger l'historique de recherche de l'utilisateur, nous supprimons chaque document dont les termes n'apparaissent pas dans le centre d'intérêts pendant une année.

4. Evaluation et Implémentation

Selon [Toms et al., 2013] et [Khazri et al., 2008], la pertinence à améliorer dans le contexte de la recherche personnalisée d'informations est très subjective, car cette pertinence dépend notamment de l'utilisateur. Cet utilisateur est l'unique capable de juger cette pertinence en fonction de son point de vue. Cette réalité rend l'évaluation d'une approche de recherche d'information personnalisée une tâche très problématique. Car il n'existe pas une collection de test standard qui permet à l'utilisateur d'évaluer son propre travail, en effet une collection de test avec des interrogations et des pré-jugements de pertinence réalisés par un expert de domaine exprime la pertinence des documents pour lui-même et pas forcément pour les autres utilisateurs.

Donc de notre point de vue, l'exploitation des collections de tests, comme la majorité des approches PIR, ne résout pas le problème, car ce n'est pas l'avis des utilisateurs concernés.

Alors, comment pouvons-nous évaluées les améliorations de notre approche ?

Pour nous l'évaluation de notre approche se fait à l'aide d'une simulation. Nous avons donc suivi la stratégie suivante pour un ensemble d'utilisateurs, où chaque utilisateur envoi un ensemble de requêtes à un moteur de recherche afin de satisfaire son besoin en information. Pour cela nous allons appliquer les formules de reformulation proposées sur l'ensemble de requêtes, ensuite les envoyer à un moteur de recherche classique.

Après la réception des résultats, le jugement de pertinence se fait manuellement (satisfait ou non). Et à la fin nous allons faire une comparaison entre les résultats retournés avant et après la reformulation des requêtes.

4.1. 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 ([Buidguaguen, 2011], [Panjarkar et al., 2014], …), où chaque approche développe son propre cadre d'évaluation, nous mènerons notre phase d'évaluation selon les étapes suivantes :

- ✓ *Construction du profil de l'utilisateur* : Nous avons construit 10 profils sociaux pour 10 utilisateurs ayant des intérêts différents dans 5 domaines différents où chaque deux utilisateurs ont des intérêts similaires. Nous avons associé à chaque utilisateur entre 1 et 3 amis proches, l'un de ces amis est de même domaine que l'utilisateur concerné, tandis que d'autres ont des domaines différents.
  Les intérêts de l'utilisateur sont représentés par une liste de termes pertinents qui représente son centre d'intérêt. Pour chaque utilisateur, certains intérêts sont définis d'une façon spécifique selon son domaine, afin de montrer le besoin à la personnalisation de la recherche. La taille des profils varie entre les profils pauvres, moyens et riches.
- ✓ *Requêtes et recherche* : Nous utilisons *Google* comme moteur de recherche où chaque utilisateur envoi 4 requêtes. Pour chaque utilisateur, au moins une requête doit exprimer un besoin informationnel spécifique, et une requête doit exprimer un nouveau besoin informationnel qui peut être aidé à partir des centres d'intérêts de ses amis. Certaines requêtes des utilisateurs sont communes.
- ✓ *Evaluation* : Pour avoir un réel feedback, les jugements de pertinence sur les résultats renvoyés sont faits manuellement par les utilisateurs. L'évaluation consiste à calculer le facteur de précision (taux des documents pertinents retournés) aux 5, 10 et 15 premiers documents retournés.
- ✓ *Comparaison des Performances* : Pour comparer la contribution de notre approche de personnalisation par rapport aux approches classiques, nous calculons la précision selon les 2 scénarios suivants (A) requête envoyée sans reformulation, (B) requête envoyée avec reformulation, en utilisant notre profil social d'utilisateur.

4.2. Expérimentation

L'expérimentation consiste à exécuter les requêtes des utilisateurs sur notre modèles de recherche selon deux scenario : *sans reformulation* et *avec reformulation de requête*. L'objectif est de démontrer que notre modèle de recherche personnalisé basé sur la reformulation de requête à partir des relations sociales, est plus performant que les modèles de recherche classique.

(l'expérimentation n'est pas complétement implémenté, certaines parties sont manuelles).

4.2.1. Construction des profils et requêtes

Nous commençons notre expérimentation par la construction des profils utilisateurs qui définissent les centres d'intérêts de l'utilisateur et ses amis proches ainsi que les requêtes à envoyées, de la façon suivante :

User1 : « étudiant en informatique »

Centres d'intérêts : (informatique, 0.751), (logiciel, 0.412), (langage, 0.63), (programmation, 0.684), (java, 0.511), (eclipse, 0.429), (NetBeans, 0.431), (PHP, 0.567), (SQL, 0.534), (SGBD, 0.366), (application, 0.412), (algorithme, 0.387), (information, 0.628), (système,0.369), (recherche, 0.61), (sécurité, 0.102), (faille, 0.098), (calculs, 0.03), (processus, 0.125), (vulnérabilité, 0.087).

Amis : {user2, 0.4}, {user3, 0.1}.

Requête :

- Java.
- Apple.
- processus de recherche.
- Les outils de création de sites web.

User2 : « étudiant en informatique »

Centres d'intérêts : (réseau, 0.102), (information, 0.023), (informatique, 0.031), (électronique, 0.54), (technologie, 0.27), (matériel, 0.049), (système, 0.256), (smartphone, 0.154), (Apple, 0.06), (Android, 0.117), (protocole, 0.309), (sécurité, 0.023), (recherche, 0.414), (panne, 0.005), (processus, 0.113), (logiciel, 0.028), (application, 0.056).

Amis : {user1, 0.44}.

Requête :

- application Android.
- type de vulnérabilité.
- définition réseau.
- Vérification de niveau de sécurité.

User3 : « étudiant en astronomie »

Centres d'intérêts : (astre, 0.521), (phénomène, 0.578), (eclipse, 0.411), (lune, 0.12), (soleil, 0.196), (terre, 0.135), (planète, 0.108), (étoile-filante, 0.274), (satellite, 0.155), (astronomie, 0.641), (science, 0.101), (étoile, 0.205), (télescope, 0.212), (géographie, 0.086), (NASA, 0.199), (spatiale, 0.079), (calculs 0.051).

Amis : {user4, 0.705}, {user6, 0.411}, {user1, 0.177}.

Requêtes

- c'est quoi eclipse.
- Hubble.
- découverte de marguerite.
- Constellation de la grande ourse.

User4 : « étudiant en astronomie »

Centres d'intérêts : : (univers, 0.413), (céleste, 0.125), (galaxie, 0.001), (astre, 0.14), (loi, 0.228), (Hubble, 0.323), (planète, 0.007), (astronomie, 0.018), (science, 0.236), (étoile, 0.145), (soleil, 0.0123), (terre, 0.109), (télescope0.208), (géographie, 0.144), (NASA, 0.227), (spatiale, 0.023), (satellite, 0.235).

Amis : {user3, 0.705}, {user6, 0.411}.

Requête :

- la situation céleste de rastaban.
- Navette spatiale.
- le point le plus froid de l'univers.
- Phénomène 2018.

User5 : « étudiant en histoire/géographie »

Centres d'intérêts : (guerre, 0.221), (armée, 0.101), (siècle, 0.056), (religion, 0.19), (victoire, 0.377), (mondiale, 0.05), (nucléaire, 0.441), (pays, 0.025), (Europe, 0.113), (peuple, 0.06), (territoire, 0.211), (géographie, 0.175), (histoire, 0.208), (ville, 0.141), (monde, 0.301), (France, 0.196), (Algérie, 0.222).

Amis : {user6, 0.529}.

Requête :

- France vs Allemagne.
- La carte du monde.
- La 2ieme guerre mondiale.
- La bataille de Waterloo.

User6 : « étudiant en histoire/géographie »

Centres d'intérêts : (continent, 0.012), (monde, 0.163), (ville, 0.152), (histoire, 0.301), (géographie, 0.0213), (satellite, 0.001), (spatiale, 0.085), (climat, 0.227), (science, 0.013), (terre, 0.263), (planète, 0.499), (territoire, 0.014), (peuple, 0.158), (nucléaire, 0.012), (pays, 0.13), (Europe, 0.121), (astronomie, 0.231), (carte, 0.114), (Atlas, 0.213).

Amis : {user5, 0.473}, {user3, 0.368}.

Requête :

- Atlas.
- La seconde guerre mondiale.
- La propagande nazie.
- Quelle est la capitale de la Turquie ?

User7 : « intéresse au tourisme »

Centres d'intérêts : (tourisme, 0.144), (touriste, 0.012), (hôtel, 0.523), (destination, 0.401), (loisir, 0.210), (vacance, 0.139), (plage, 0.056), (voyage, 0.171), (restaurant, 0.014), (transport, 0.156), (ville, 0.142), (climat, 0.045), (saison, 0.078), (pays, 0.169), (Europe, 0.266), (agence, 0.652), (bon plan, 0.742).

Amis : {user8, 0.411}, {user6, 0.235}.

Requête :

- voyage pas cher.
- réservation air Algérie.

- Sites touristique en Algérie.
- Monuments célèbres.

User8 : « intéresse au tourisme »

Centres d'intérêts : (tourisme, 0.124), (touriste, 0.309), (voyage, 0.112), (aéroport, 0.456), (pays, 0.222), (avion, 0.233), (hôtel, 0.278), (restaurant, 0.256), (plage, 0.102).

Amis : {user7, 0.777}.

Requête :

- Bon plan.
- Destination de rêve.
- Sites et monuments touristique.
- hôtel Algérie 5 étoiles.

User9 : « étudiant en agriculture »

Centres d'intérêts : (fruit, 0.023),(vegetable, 0.01), (production, 0.0.231), (organic, 0.009), (fertilizer, 0.017), (ground, 0.025), (climate, 0.144), (farm, 0.001), (biological, 0.092), (environment, 0.102), (cultivate, 0.07), (consumer, 0.152), (pesticide, 0.117), (apple, 0.25), (potato, 0.012), (field, 0.014), (farming, 0.023).
Amis : {user10, 0.352}.

Requête :

- Apple.
- Types of plowing.
- production techniques.
- Fertilizer Organic Potato.

User10 : « étudiant en agriculture »

Centres d'intérêts: (production, 0.023), (economy, 0.111), (export, 0.214), (import, 0.132), (organic, 0.012), (consumer, 0.299), (cultivate, 0.142), (ground, 0.213), (trade, 0.125), (farming, 0.001).

Amis : {user9, 0.6}.

Requête:

- Allowed fertilizer in organic.
- Organic food.
- Type of pesticide.
- Hydroponic.

4.2.2. Résultats et discussion

Dans une première partie, nous présentons le tableau (***Tableau 3.1***) des résultats expérimentaux obtenus pour les requêtes de chaque utilisateur (qui acceptent la reformulation seulement).

	Sans reformulation de requête			***Avec reformulation de requête***		
User 1	**p@5**	**p@10**	**p@15**	**p@5**	**p@10**	**p@15**
Java	0.4	0.4	0.333	0.6	0.7	0.53
Apple	0.2	0.3	0.26	0.4	0.4	0.4
Processus de recherche	0	0	0.066	0.4	0.4	0.266
Moyenne taux-précision	**0.2**	**0.233**	**0.219**	**0.466**	**0.5**	**0.398**
User 2	**p@5**	**p@10**	**p@15**	**p@5**	**p@10**	**p@15**
Application Android	0	0	0.133	0.4	0.4	0.466
Type de vulnérabilité	0.6	0.5	0.4	0.8	0.9	0.866
Définition réseau	0.4	0.3	0.266	1	0.9	0.866
Vérification de niveau de sécurité	0.2	0.1	0.066	0.8	0.7	0.6
Moyenne taux-précision	**0.3**	**0.225**	**0.216**	**0.75**	**0.725**	**0.699**
User 3	**p@5**	**p@10**	**p@15**	**p@5**	**p@10**	**p@15**
c'est quoi eclipse ?	0.4	0.3	0.2	0.8	0.5	0.466
Hubble	0	0	0	1	0.8	0.733
Moyenne taux-précision	**0.2**	**0.15**	**0.1**	**0.9**	**0.65**	**0.599**
User 4	**p@5**	**p@10**	**p@15**	**p@5**	**p@10**	**p@15**
la situation céleste de Rastaban	0.6	0.5	0.533	0.8	0.8	0.733
Navette spatiale	0.2	0.3	0.4	0.6	0.5	0.533
le point le plus froid de l'univers	0.4	0.4	0.4	1	0.8	0.666
Phénomène 2018	0.2	0.1	0.006	1	0.9	0.666
Moyenne taux-précision	**0.35**	**0.325**	**0.334**	**0.85**	**0.75**	**0.649**
User 5	**p@5**	**p@10**	**p@15**	**p@5**	**p@10**	**p@15**
France vs Allemagne	0	0	0	0.6	0.5	0.466
La carte du monde	0.8	0.7	0.533	0.4	0.5	0.4
La 2ieme guerre mondiale	0.8	0.7	0.533	1	0.8	0.733
La bataille de Waterloo	1	0.6	0.466	0.8	0.7	0.533
Moyenne taux-précision	**0.65**	**0.5**	**0.383**	**0.7**	**0.625**	**0.533**
User 6	**p@5**	**p@10**	**p@15**	**p@5**	**p@10**	**p@15**
Atlas	0	0	0	0.8	0.6	0.533
La seconde guerre mondiale	0.4	0.6	0.533	0.8	0.8	0.666
Moyenne taux-précision	**0.2**	**0.3**	**0.266**	**0.8**	**0.7**	**0.599**
User 7	**p@5**	**p@10**	**p@15**	**p@5**	**p@10**	**p@15**

voyage pas cher	0	0.2	0.266	0.6	0.8	0.733
Moyenne taux-précision	0	0.2	0.266	0.6	0.8	0.733
User 8	**p@5**	**p@10**	**p@15**	**p@5**	**p@10**	**p@15**
Bon plan	0	0.1	0.066	1	0.8	0.8
Destination de rêve	0.6	0.7	0.666	1	0.8	0.666
Hôtel Algérie 5 étoile	1	0.8	0.733	1	0.8	0.666
Moyenne taux-précision	**0.233**	**0.533**	**0.688**	**1**	**0.8**	**0.710**
User 9	**p@5**	**p@10**	**p@15**	**p@5**	**p@10**	**p@15**
Apple	0.2	0	0	0.6	0.7	0.533
production techniques	0	0	0	0.4	0.6	0.6
Fertilizer Organic Potato	0.6	0.5	0.533	0.8	0.8	0.666
Moyenne taux-précision	**0.266**	**0.166**	**0.177**	**0.6**	**0.7**	**0.599**
User 10	**p@5**	**p@10**	**p@15**	**p@5**	**p@10**	**p@15**
Allowed fertilizer in organic	0.8	0.5	0.4	0.8	0.6	0.6
Organic food	0.4	0.3	0.333	0.6	0.8	0.666
Type of pesticide	0	0.2	0.2	0.6	0.7	0.6
Moyenne taux-précision	**0.4**	**0.333**	**0.311**	**0.666**	**0.7**	**0.622**

Tableau 3. 1*: Résultats expérimentaux des requêtes qui acceptent la reformulation pour chaque utilisateur.*

La première constatation au vu de ces résultats est que notre modèle est efficace et présente des améliorations de performances significatives comparativement au modèle classique (requête sans reformulation) aux points de précision *p@5, p@10 et p@15.*

Lorsqu'on analyse les taux de précision pour chaque requête, on remarque que le degré d'amélioration change d'une requête à une autre. Ceci dépend probablement, de la correspondance entre les centres d'intérêts de l'utilisateur simulés et la requête.

Nous notons que l'utilisation de notre profil social d'utilisateur a apporté une amélioration de la précision par rapport aux approches classique, notamment au niveau des 5 premiers documents les plus susceptibles d'être examinés par l'utilisateur. Cette amélioration est due à la manière avec laquelle nous avons conçu les 10 profils utilisateurs, ainsi qu'à l'ensemble des requêtes envoyées.

Et dans une deuxième partie, nous présentons le tableau (***Tableau 3.2***) des résultats des requêtes qui n'acceptent pas des reformulations à cause d'absence des termes de requête dans le centre d'intérêts de l'utilisateur concerné ou celles de ses amis, qui signifie que l'utilisateur a des nouveaux besoins ou des nouveaux intérêts.

	Sans reformulation de requête			***Avec reformulation de requête***		
User 6	**p@5**	**p@10**	**p@15**	**p@5**	**p@10**	**p@15**
Atlas	0	0	0	0.8	0.6	0.533
La seconde guerre mondiale	0.4	0.6	0.533	0.8	0.8	0.666
La propagande nazie	0.8	0.6	0.6	-	-	-
Quelle est la capitale de la Turquie ?	0.8	0.7	0.533	-	-	-
Moyenne taux-précision	**0.5**	**0.475**	**0.416**	**0.4**	**0.35**	**0.299**
User 8	**p@5**	**p@10**	**p@15**	**p@5**	**p@10**	**p@15**
Bon plan	0	0.1	0.066	1	0.8	0.8
Destination de rêve	0.6	0.7	0.666	1	0.8	0.666
Sites et monuments touristique	0.6	0.4	0.266	-	-	-
Hôtel Algérie 5 étoile	1	0.8	0.733	1	0.8	0.666
Moyenne taux-précision	**0.55**	**0.5**	**0.432**	**0.75**	**0.6**	**0.533**
User 10	**p@5**	**p@10**	**p@15**	**p@5**	**p@10**	**p@15**
Allowed fertilizer in organic	0.8	0.5	0.4	0.8	0.6	0.6
Organic food	0.4	0.3	0.333	0.6	0.8	0.666
Type of pesticide	0	0.2	0.2	0.6	0.7	0.6
Hydrponic	1	0.6	0.533	-	-	-
Moyenne taux-précision	**0.55**	**0.4**	**0.366**	**0.5**	**0.525**	**0.466**

Tableau 3. 2*: Résultats expérimentaux des requêtes qui n'acceptent pas la reformulation pour quelques utilisateurs.*

Nous notons que les résultats de précision obtenus sont proches dans les 2 scenarios et il n'y a pas une nette amélioration, cela indique que notre système est moins performent lorsque l'utilisateur à des nouveaux besoins en information.

Exemples :

Voici quelques exemples d'améliorations des résultats obtenus après la reformulation de requête apportés dans notre approche de personnalisation :

- ✓ *Exemple 1* : Les premiers résultats de la requête « bon plan » envoyé par l'utilisateur 8 qui s'intéresse au tourisme ne satisfaire pas ses besoins.

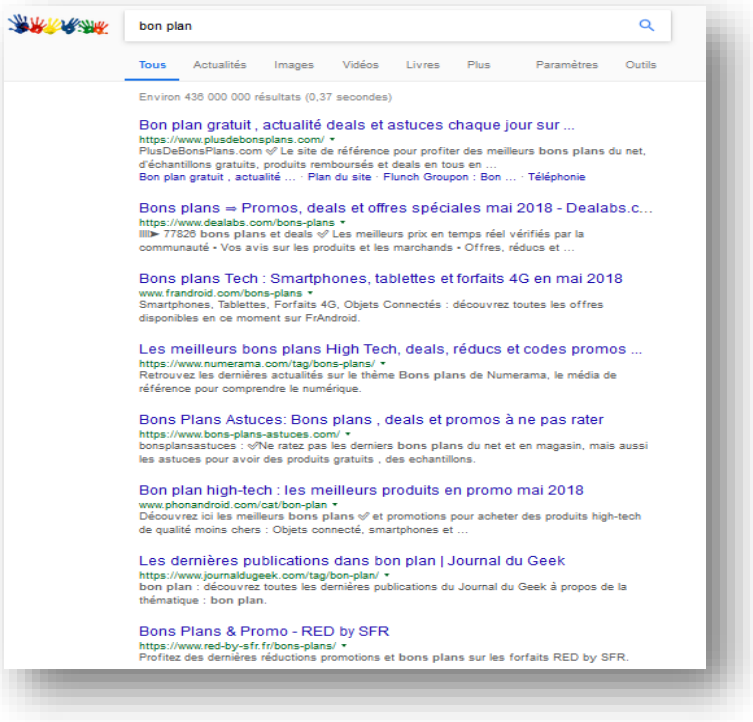


***Figure 3. 7** : représentation des résultats de la requête "bon plan" avant la reformulation.*

Mais après la reformulation de la requête par l'ajout du terme « voyage » extrait à partir du centre d'intérêts de l'utilisateur7 (qui est l'un des amis de l'utilisateur8), les résultats ont complétement changé et l'utilisateur été satisfait.

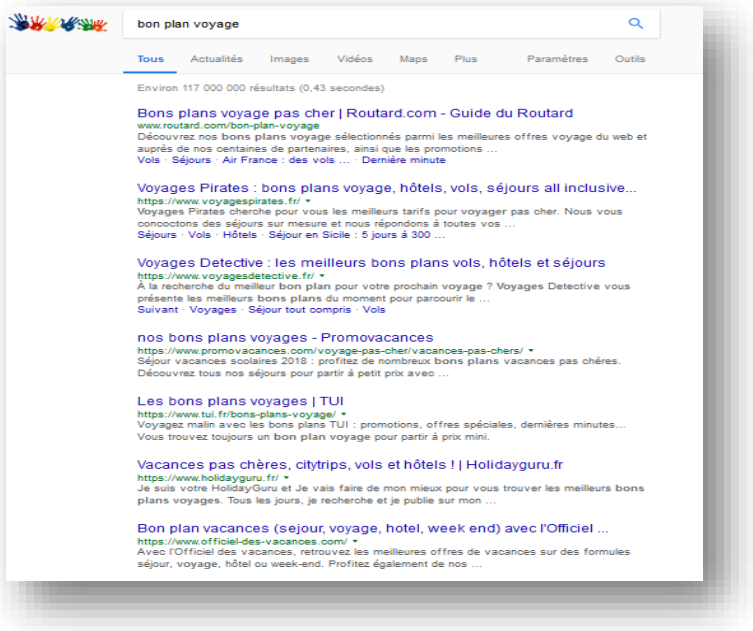


***Figure 3. 8** : représentation des résultats de la requête " bon plan " après la reformulation.*

- ✓ *Exemple 2* : l'utilisateur5 a envoyé la requête « France vs Allemagne » au moteur de recherche afin d'obtenir des résultats qui satisfaire son besoin et les résultats été comme suit :

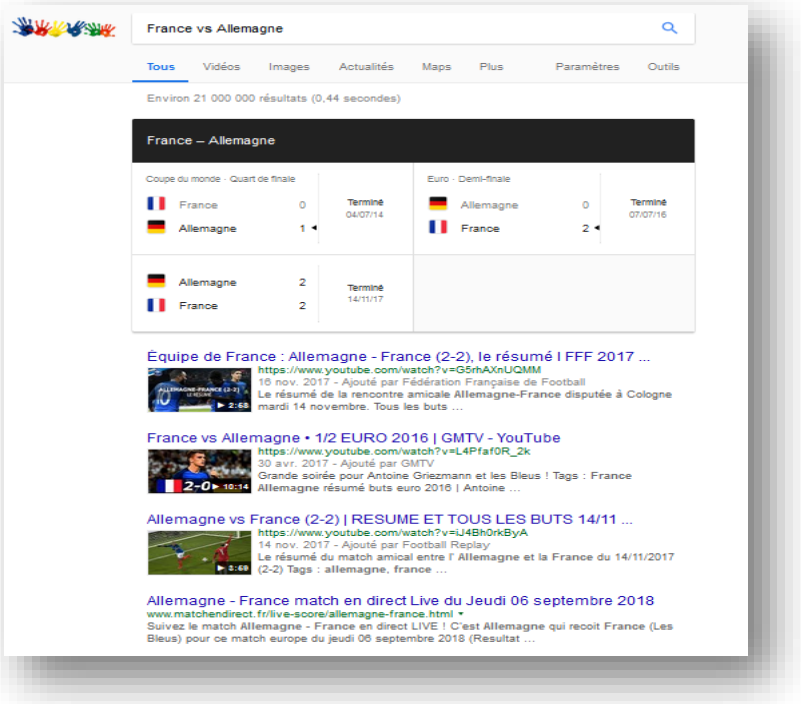


***Figure 3. 9** : représentation des résultats de la requête " France vs Allemagne " avant la reformulation.*

Nous remarquons que les premiers résultats retournés par le moteur de recherche, concerne tout un autre domaine diffèrent des intérêts de l'utilisateur. Par contre après la reformulation de la requête, les résultats ont beaucoup mieux améliorés.

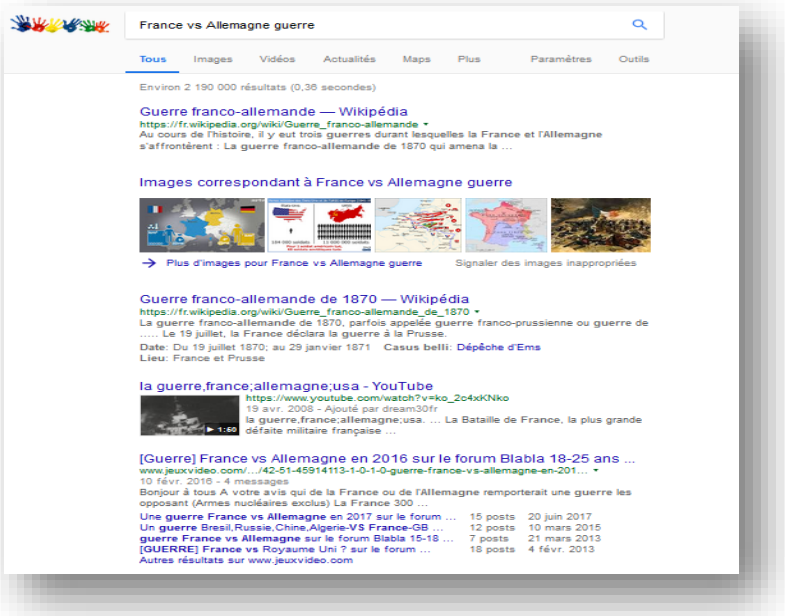


***Figure 3. 10** : représentation de résultats de la requête " France vs Allemagne " après reformulation.*

Et à la fin nous présentons le tableau (***Tableau 3.3)*** qui résume les résultats de notre expérimentation et qui présente la moyenne des précisions des résultats obtenus sur l'ensemble des requêtes envoyées par les utilisateurs (l'ensemble des requêtes qui acceptent la reformulation de requête) selon les 2 scenarios déjà expliqués.

	Scenario (A)	**Scenario (B)**	**Taux d'amélioration (B/A)**
P @5	0.231	0.733	0.502
P @10	0.295	0.695	0.4
P @15	0.326	0.560	0.238
Précision moyenne	**0.284**	**0.662**	**0.378**

Tableau 3. 3*: la moyenne des précisions des résultats obtenus sur l'ensemble des requêtes envoyées par les utilisateurs.*

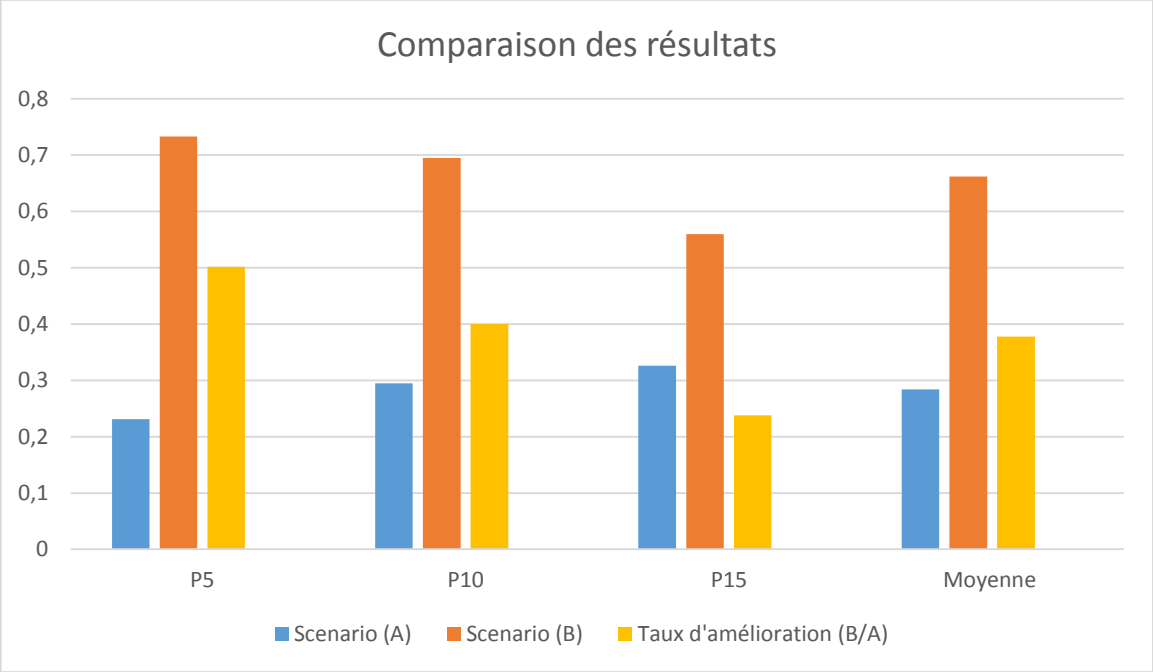


Figure 3. 11 *: comparaison des résultats des 2 scénarios.*

Les détails de la simulation montrent que nos résultats sont bons et améliore la pertinence des résultats retournés, dans le cas où la requête est courte et le profil est très riche. Et nos résultats sont peu intéressants dans le cas d'un nouveau profil ou d'un nouveau besoin en information.

A noter, que l'utilisation d'un cadre d'évaluation plus important que le nôtre en termes : du nombre d'utilisateurs, des tailles des profils, du nombre des requêtes, pourra donner des résultats mieux reflétant la réalité.

4.3. Implémentation

Comme l'expérimentation n'est pas implémentée intégralement, l'objectif de cette implémentation est de prouver l'applicabilité de notre solution proposé.

L'implémentation touche les phases suivantes :

- La reformulation de la requête.
- La mise à jour de profil d'utilisateur ainsi que la liste des amis.

4.3.1. Organisation physique des composantes de profil

Les différentes composantes du profil utilisateur sont modélisés à l'aide d'un MCD (Modèle Conceptuel des Données) de la méthode MERISE [Tardieu et al., 1983] comme suit :

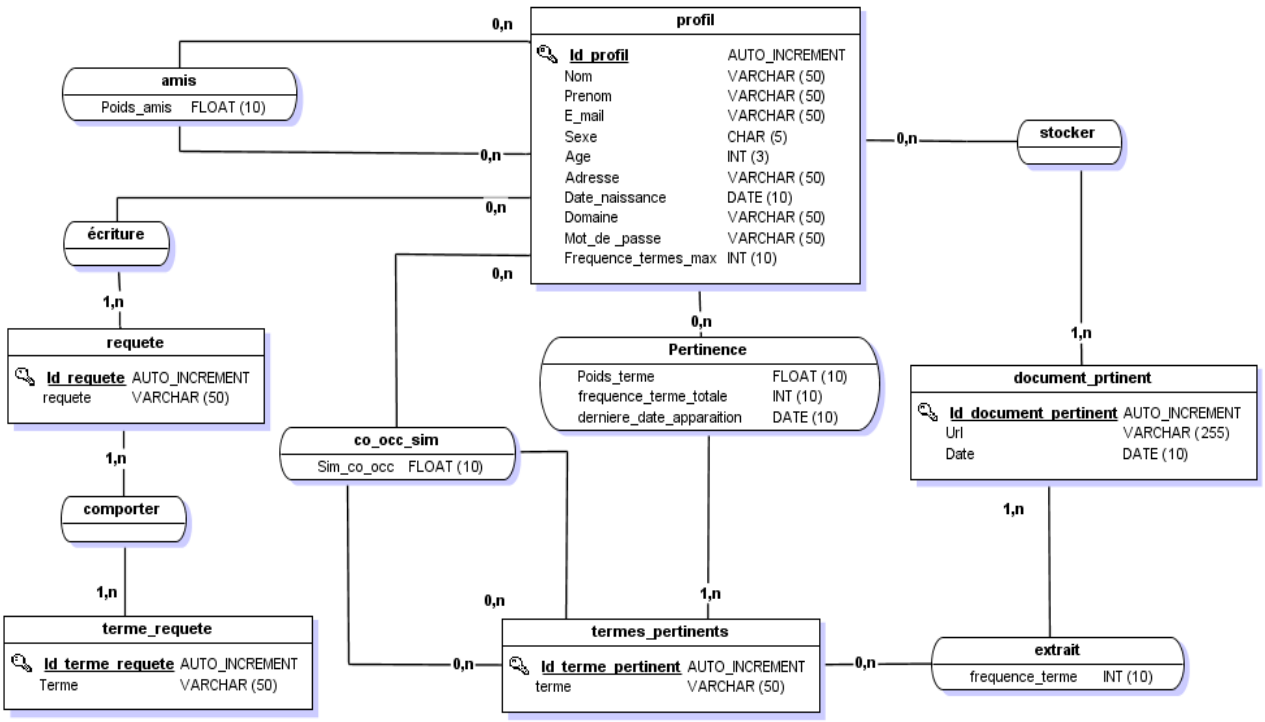


***Figure 3. 12** : MCD du profil utilisateur.*

4.3.2. Présentation de la base de données

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

4.3.2.1. Présentation de MySQL

MySQL est un serveur de bases de données relationnelles basé sur une bibliothèque de gestion de données prouvée depuis des nombreuses années, il utilise une architecture multithread et multi-utilisateurs qui lui permettent d'être très rapide et robuste. Aussi c'est un logiciel libre qui fonctionne sur plusieurs plateformes : *Linux, Open server, Windows XP*...etc.

4.3.2.2. Création des tables de base de données

Notre base de données se compose des tables suivantes (***Figure 3.13***) :

	Table	Action	Lignes	Type	Interclassement	Taille	Perte
☐	amis	Afficher Structure Rechercher Insérer Vider Supprimer	~5	InnoDB	utf8_general_ci	32 Kio	-
☐	comporter	Afficher Structure Rechercher Insérer Vider Supprimer	~135	InnoDB	utf8_general_ci	32 Kio	-
☐	co_occ_sim	Afficher Structure Rechercher Insérer Vider Supprimer	~1 406	InnoDB	utf8_general_ci	256 Kio	-
☐	documents_pertinents	Afficher Structure Rechercher Insérer Vider Supprimer	~6	InnoDB	utf8_general_ci	16 Kio	-
☐	ecriture	Afficher Structure Rechercher Insérer Vider Supprimer	~113	InnoDB	utf8_general_ci	32 Kio	-
☐	extrait	Afficher Structure Rechercher Insérer Vider Supprimer	~28	InnoDB	utf8_general_ci	32 Kio	-
☐	mots_vides	Afficher Structure Rechercher Insérer Vider Supprimer	~495	InnoDB	utf8_general_ci	16 Kio	-
☐	pertinence	Afficher Structure Rechercher Insérer Vider Supprimer	~75	InnoDB	utf8_general_ci	32 Kio	-
☐	profil	Afficher Structure Rechercher Insérer Vider Supprimer	~5	InnoDB	utf8_general_ci	16 Kio	-
☐	requete	Afficher Structure Rechercher Insérer Vider Supprimer	~137	InnoDB	utf8_general_ci	16 Kio	-
☐	stocker	Afficher Structure Rechercher Insérer Vider Supprimer	~4	InnoDB	utf8_general_ci	32 Kio	-
☐	termes_pertinents	Afficher Structure Rechercher Insérer Vider Supprimer	~63	InnoDB	utf8_general_ci	16 Kio	-
☐	terme_requete	Afficher Structure Rechercher Insérer Vider Supprimer	~135	InnoDB	utf8_general_ci	16 Kio	-
	13 tables	Somme	2 607	InnoDB	utf8_general_ci	544 Kio	0 o

Figure 3. 13 *: Les tables de la base de données.*

4.3.2.3. Liste des tables de la base de données

- ✓ *Table « amis » :* les relations d'amitié entre les utilisateurs sont toutes définies dans cette table, elle comporte les amis de chaque utilisateur avec leurs poids.
- ✓ *Table « comporter » :* représente la relation de chaque requête avec ses termes.
- ✓ *Table « co_occ_sim » :* contient les relations de cooccurrence entre les termes pertinents de chaque utilisateur.
- ✓ *Table « documents_pertinents » :* cette table sert à stocker les documents pertinents de chaque utilisateur.
- ✓ *Table « écriture » :* sert à stocker les requêtes écrites par chaque utilisateur.
- ✓ *Table « extrait » :* représente les termes extraits de chaque document comme suit (***Figure 3.14***) :

	Id_termes_pertinence	Id_document_pertinent	frequence_terme
Modifier Copier Effacer	1	5	50
Modifier Copier Effacer	1	28	22
Modifier Copier Effacer	2	5	112
Modifier Copier Effacer	3	2	70
Modifier Copier Effacer	4	2	38
Modifier Copier Effacer	5	4	34
Modifier Copier Effacer	6	4	10
Modifier Copier Effacer	7	4	19
Modifier Copier Effacer	8	2	598
Modifier Copier Effacer	9	2	48
Modifier Copier Effacer	10	2	17
Modifier Copier Effacer	11	5	19
Modifier Copier Effacer	12	2	25
Modifier Copier Effacer	13	1	27
Modifier Copier Effacer	13	28	55

Figure 3. 14 *: exemple de la table «extrait ».*

- ✓ *Table « mots_vides » :* contient les mots qui doivent êtres éliminer de la requête écrite par l'utilisateur avant de lancer la recherche.
- ✓ *Table « pertinence » :* représente le degré de pertinence des termes de chaque utilisateur, elle contient le poids de chaque terme, sa fréquence totale (la fréquence de terme dans tout l'historique des documents pertinents de l'utilisateur) et sa dernière date d'apparition.
- ✓ *Table « profil » :* cette table contient les informations personnelles qui identifier chaque utilisateur et représente comme suit (***Figure 3.15***) :

	Id_profil	Nom	Prenom	E_mail	Sexe	Age	Adresse	Date_naissance	Domaine	Mot_de__passe	Frequence_termes_max
Modifier Copier Effacer	1	TOUBAL SEGHIR	nassima	nassima.tsgh@gmail.com	femme	24	Médéa	1994-06-12	Informatique	azerty	512
Modifier Copier Effacer	2	BACHENE	leila	bl@gmail.com	femme	24	Médéa	1994-01-31	Informatique	123	235
Modifier Copier Effacer	4	TOUBAL SEGHIR	dounia	tsd@gmail.com	femme	24	Médéa	1994-06-13	Astronomie	789	55
Modifier Copier Effacer	5	BACHENE	amel	ba@gmail.com	femme	24	Médéa	1994-10-29	Astronomie	abcd	116
Modifier Copier Effacer	6	TOUBAL SEGHIR	yasmine	tyas@gmail.com	femme	24	Médéa	1994-10-18	histoire/géographie	mino	95

Figure 3. 15 *: exemple de la table «profil ».*

- ✓ *Table « requête » :* toutes les requêtes lancées par un utilisateur sont stockées dans cette table.
- ✓ *Table «stocker» :* représente l'historique des documents pertinents de chaque utilisateur.
- ✓ *Table « termes_pertinents » :* sert à stocker les termes pertinents des utilisateurs (***Figure 3.16)*** :



	Id_termes_pertinence	terme
Modifier Copier Effacer	1	informatique
Modifier Copier Effacer	2	logiciel
Modifier Copier Effacer	3	langage
Modifier Copier Effacer	4	programmation
Modifier Copier Effacer	5	java
Modifier Copier Effacer	6	eclipse
Modifier Copier Effacer	7	NetBeans
Modifier Copier Effacer	8	PHP
Modifier Copier Effacer	9	SQL
Modifier Copier Effacer	10	SGBD
Modifier Copier Effacer	11	application
Modifier Copier Effacer	12	algorithme
Modifier Copier Effacer	13	information

***Figure 3. 16** : exemple de la table «termes_pertinents ».*

- ✓ *Table « terme_requête » :* sert à stocker les termes des requêtes écrites par les utilisateurs.

4.3.3. Environnement de développement de l'application

Pour réaliser cette application, nous avons besoin de manipuler certains langages et logiciels :

PHP :

Nous avons choisi le langage *PHP* (PHP Hypertext Processor) qui est un langage de programmation web coté serveur et *open source*.

PHP est un outil riche et puissant pour créer des pages web dynamiques et interactives, il couvre pratiquement tous les domaines en rapport avec les applications web.

Wamp Server :

Wamp est un acronyme signifiant (*Windows*, *Apache*, *MySQL* et *PHP*). Afin de pouvoir utiliser WAMP, nous avons un logiciel du nom de **«** *WampServer* **»** qui nous en donne la possibilité. *Wamp server* est une plateforme de développement web et tests internes. Il inclut également MySQL et PHP, deux des technologies les plus utilisées pour créer des sites web dynamique, ainsi que *PHPMyAdmin* pour l'administration Web des bases *MySQL*.

4.3.4. Présentation des interfaces

La figure (***Figure 3.17***) présente la page d'inscription des utilisateurs afin qu'ils puissent créer des profils pour effectuer leurs recherches classiques ou personnalisées :

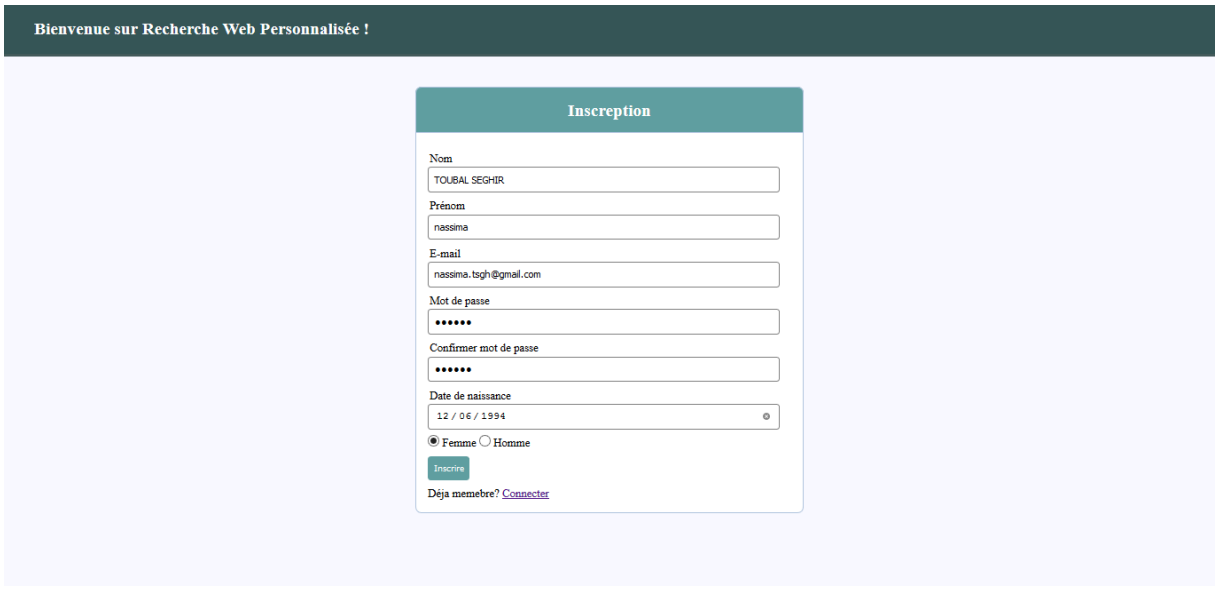


Figure 3. 17 *: Page d'inscription.*

Une fois l'utilisateur est inscrit, il peut voir toutes les dimensions de son profil et lancer des recherches après s'être identifié. La figure (***Figure 3.18***) ci-dessous présente la page d'identification :

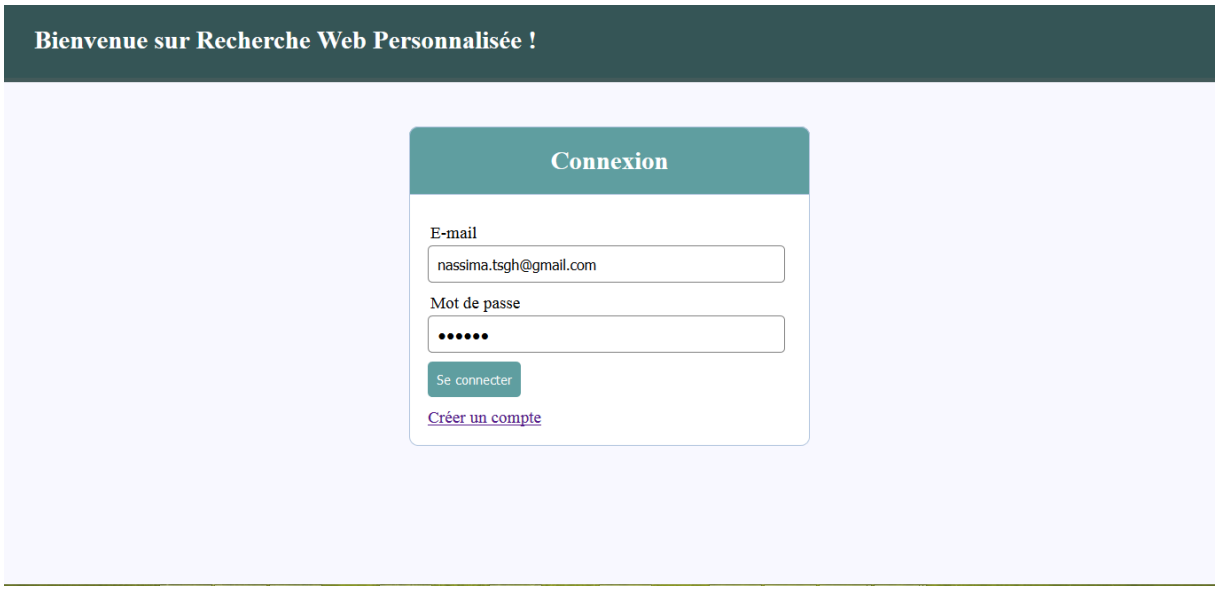


Figure 3. 18 *: Page d'identification.*

Dès que l'utilisateur est identifié, la page de recherche va s'apparaitre (***Figure 3.19***). Dans cette page l'utilisateur peut lancer sa requête soit en mode de recherche classique ou personnalisée :

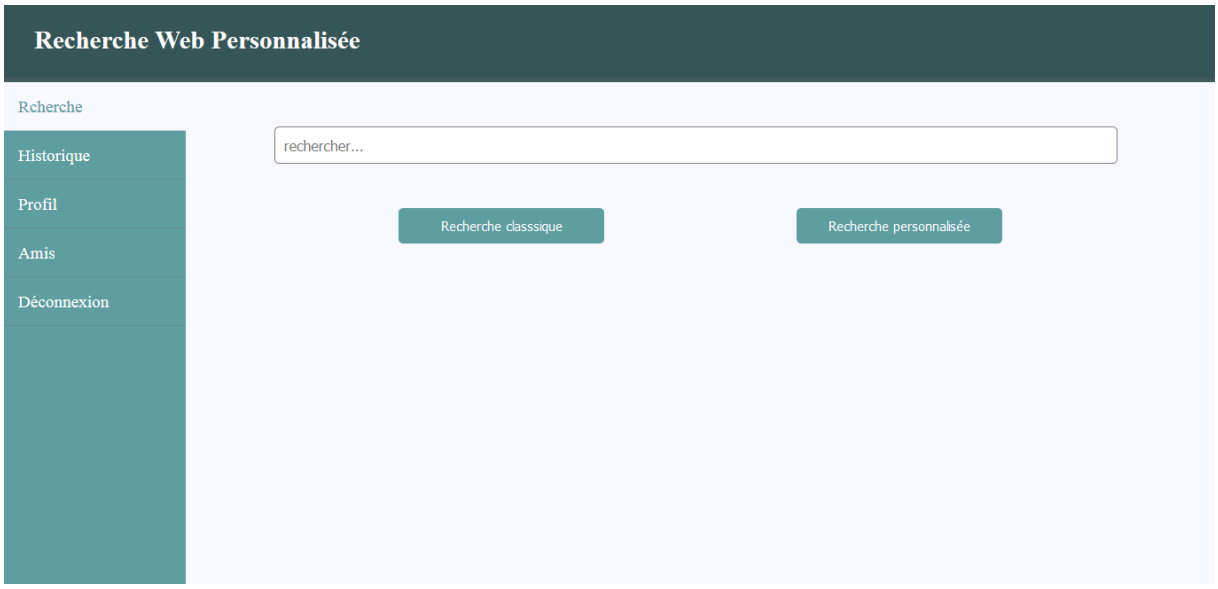


***Figure 3. 19** : Page de recherche utilisateur.*

Lorsqu'un utilisateur saisi sa requête et lance la recherche en mode classique il va se rediriger vers la page des résultats du moteur de recherche Google comme la montre la figure (***Figure 3.20***) suivante :

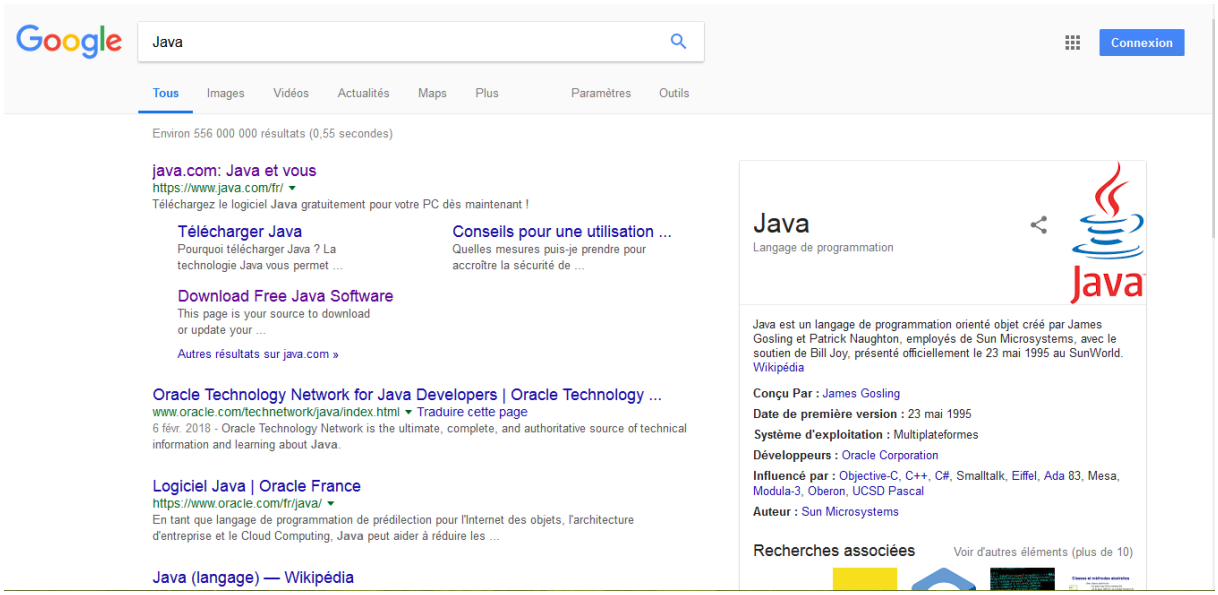


***Figure 3. 20** : Résultats de recherche en mode classique.*

Mais s'il choisit de lancer la recherche en mode personnalisé, tout un processus est suivi pour reformuler la requête puis l'envoyer au moteur de recherche afin d'obtenir des résultats selon son besoin. Les figures (***Figure 3.21, Figure 3.22***) montrent les étapes suivies par le processus de recherche :



Figure 3. 21 *: L'étape N°1 de la reformulation de la requête.*

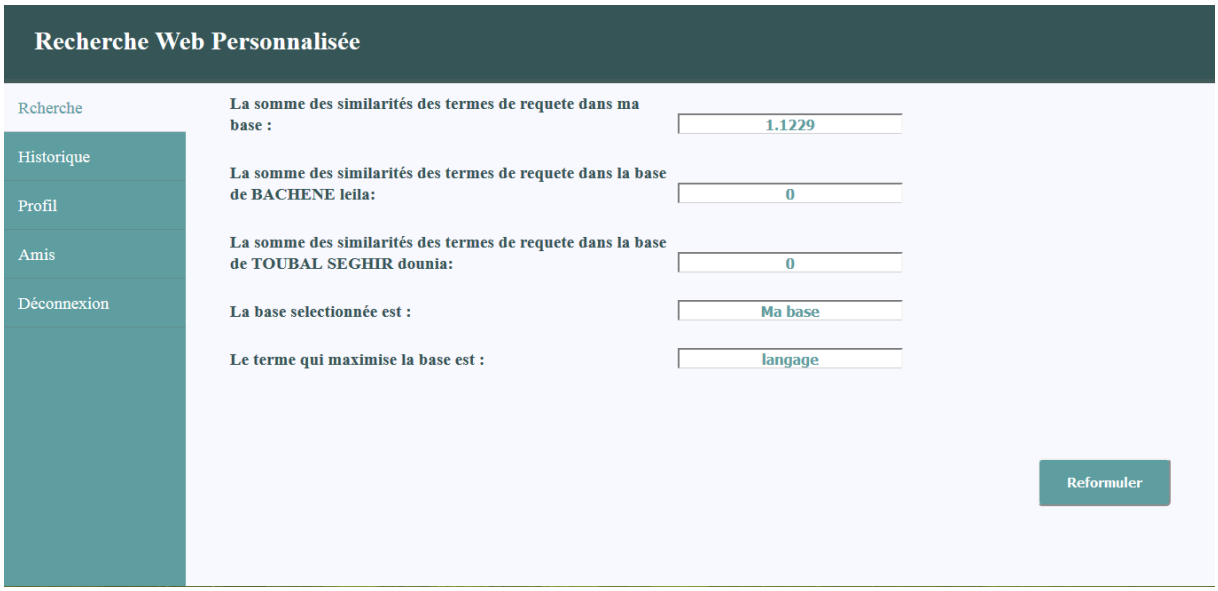


Figure 3. 22 *: L'étape N°2 de la reformulation de la requête.*

Lorsque l'utilisateur clique sur le bouton Reformuler, la requête va se reformuler est envoyer au moteur de recherche Google (***Figure 3.23***) :

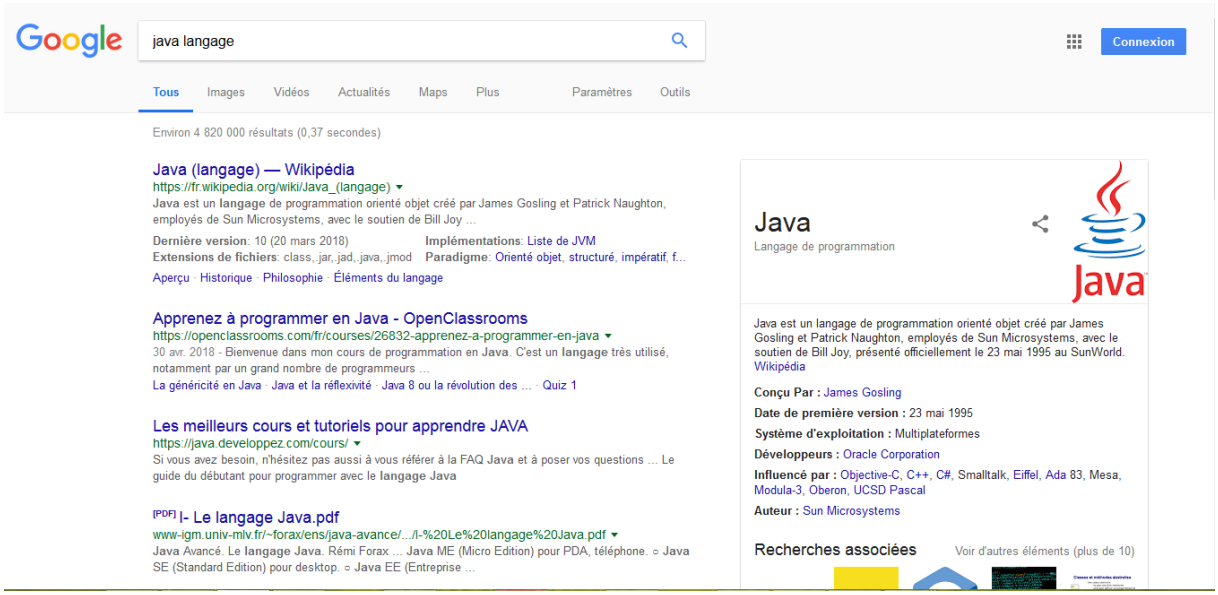


Figure 3. 23 *: Résultats de recherche après la reformulation.*

Un utilisateur peut consulter son historique de recherche (***Figure 3.24***), sa liste d'amis proches (***Figure 3.25***) ainsi que ses informations personnelles et son centre d'intérêt (***Figure 3.26***) :



Figure 3. 24 *: Historique de recherche d'un utilisateur.*



Figure 3. 25 *: La liste des amis proches.*



Figure 3. 26 *: Profil d'un utilisateur.*

5. Conclusion

Dans ce chapitre nous avons présenté le cadre conceptuel de notre approche qui est basée sur un profil d'utilisateur social qui s'intéresse aux intérêts de l'utilisateur ainsi que ses relations d'amitié, pour réaliser un système de recherche d'information personnalisé. Aussi nous avons présentées dans ce travail les relations de cooccurrence entre les termes de l'historique pertinent de recherche eux-mêmes dans le but de valoriser ces relations afin de pouvoir sélectionner le profil le plus adéquate pour guider la recherche de l'utilisateur selon ses intérêts, ses amis ainsi que sa requête.

Cette approche est proposée dans l'objectif d'améliorer la pertinence des résultats de recherche retournés, par la reformulation de la requête initiale et l'adaptation des résultats retournés aux intérêts de l'utilisateur.

Ensuite nous avons présentées une méthode d'évaluation des performances de notre système SRIP. Les résultats obtenus montrent que plus que le profil est riche, plus la requête est mieux reformulée et plus les résultats sont pertinents.

A la fin nous avons passées à l'implémentation où nous avons présentées le modèle de profil, la base de données et les différentes interfaces de notre application.

Conclusion et perspectives

Dans ce travail, et dans un contexte de recherche d'information personnalisée, nous avons traité la problématique d'exploitation des données sociales de l'utilisateur pour améliorer la pertinence des résultats de recherche.

Nous avons proposé une approche de personnalisation des recherches d'information basée sur la conception et l'utilisation d'un profil social d'utilisateur.

Le profil utilisateur combine le contexte thématique représenté par les termes pertinents extraits à partir de l'historique de recherche et le contexte social représenté par les liens d'amitié avec les autres utilisateurs de site social.

Les résultats de l'expérimentation ont affiché que plus le profil d'utilisateur est riche, plus sa contribution à guider les nouvelles recherches est utile, et plus les résultats de recherche sont pertinent.

Ce travail ouvre voie vers d'autres perspectives :

- ✓ Introduire la sémantique dans les liens d'amitiés sur le réseau social.
- ✓ Etudier l'évolution dans le temps du profil d'utilisateur et sa capacité de représenter les intérêts de l'utilisateur.

Références

[Benjabeur, 2010] : L. Benjabeur, L. Tamin. and M. Boughanem., "*Personalization and Fusion of Heterogeneous Information*" (RIAO 2010), France. A social model for Literature Access: Towards a weighted social network of authors. International Conference on Adaptivity

[Boubekeur, 2008] : F. Boubekeur. "*Contribution à la définition de modèles de recherche d'information flexibles basés sur les CP-Nets*", thèse de doctorat en informatique, Université Paul Sabatier, 2008.

[Bouhini, 2014] : Chahrazed Bouhini. « *Impact des réseaux sociaux sur le processus de recherche d'information. Réseaux sociaux et d'information* » [cs.SI]. Université Jean Monnet - Saint-Etienne, 2014. Français. <NNT : 2014STET4027>. <tel-01528583>.

[Boughanem, 2008] : M. Boughanem and J. Savoy, editors. « *Recherche d'information états des lieux et perspectives* »,Hermès Science Publications, 2008.

[Bouramoul, 2011] : A. Bouramoul. « *Recherche D'information Contextuelle et Sémantique sur le Web* », Université MANTOURI de Constantine, 2011.

[Bouzeghoub et Kostadinov 2004] : Bouzeghoub, M. et D. Kostadinov (2004). « *Une approche multidimensionnelle pour la personnalisation de l'information* ». Rapport PRiSM, Versailles, France.

[Buidguaguen, 2011] : O. Buidguaguen. « *Accès contextuel à l'information dans un environnement mobile : approche basée sur l'utilisation d'un profil situationnel de l'utilisateur et d'un profilde localisation des requêtes* ». *Phd thesis, Univ Toulouse III - Paul Sabatier, France 2011.*

[Canut et al., 2016] : Canut, Marie-Françoise and On-At, Sirinya and Péninou, André and Sèdes, Florence « *Construction du profil social de l'utilisateur dans un contexte dynamique : application d'une méthode de pondération temporelle* ». (2016) Ingénierie des Systèmes d'Information, vol. 21 (n° 2). pp. 65-94. ISSN 1633-1311.

[Cleverdon, 1967]: C. Cleverdon. "*The cranfield test on index language devices"*. ASLIB Proceedings, 1967.

[Daoud et al., 2008] : Mariam Daoud, Lynda Tamine-Lechani, Mohand Boughanem, and Bilal Chebaro. « *Construction des profils utilisateurs à base d'ontologie pour une recherche d'information personnalisée* ». In 5ème Conférence en Recherche d'Information et Applications (CORIA), mars 2008.

[Daoud, 2009] : M. DAOUD, « *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* », Université Paul Sabatier de Toulouse, thèse de doctorat de l'université Paul Sabatier, décembre 2009.

[Gowan, 2003] : J. Mc Gowan. "*A multiple model approach to personnalised information access"*. Master's thesis, Faculty of science, University college Dublin, February 2003.

[Hedin, 2004] : S. Hedin , M. Hoestlandt, S. Lenouvel, S. Verraest. "*L'évaluation des systèmes de recherche d'informations*". 2004.

Références

[Ingwersen, 1992]: P. Ingwersen. « *Information retrieval interaction* ». London, Taylor Graham, 1992.

[Ingwersen, 1994]: P. Ingwersen. "*Polyrepresentation of information needs and semantic entities: elements of a cognitive theory for information retrieval interaction*". In Proceedings of the Seventeenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 1994.

[Ingwersen, 1996]: P. Ingwersen. "*Cognitive perspectives of information retrieval interaction: Elements of a cognitive theory"*. Journal of documentation, 52(1):3–50, 1996a.

[Ishioka, 2003] T. Ishioka. "*Evaluation of criteria for information retrieval*". Web Intelligence, IEEE/WIC International Conference on Volume, 2003.

[Jeon et al., 2010]: H.Jeon,T.Kim, J.Choi, "*Personalized information retrieval by using adaptive user profiling and collaborative filtering"*. Adv. Inf. Sci. Serv. Sci. 2(4) (2010)134–142.

[Kechid, 2009]: Samir Kechid, « *Intégration du modèle utilisateur dans un système de recherche d'information distribué* », 2009.

[Khazri et al., 2008]: Mohamed Khazri, Mohamed Tmar, Mohamed Abid and Mohand Boughanem, (2008), « *Proposition pour l'intégration des réseaux petits mondes en recherche d'information* ». *Volume 11 - Numéro spécial CARI 2008 - Editeur : Marc Kokou Assogba – pp.69-81.ARIMA.*

[Korfhage, 1997]: Korfhage, R. (1997). "*Information storage and retrieval*". Wiley Computer Publishing 0-471-14- 338 3.

[Kostadinov, 2003]: D. Kostadinov. « *La personnalisation de l'information, définition de modèle de profil utilisateur »*. Rapport de dea. Master's thesis, Université de Versailles, France, 2003.

[Liu et al., 2004]: LIU F., YU C., MENG W., 2004, « *Personalized Web Search For Improving Retrieval Effectiveness* », *IEEE Trans. on Knowl. and Data Eng.*, *16*, 1, p. 28–40.

[Lobna, 2007] : H. Lobna. "*Reformulation de Requêtes par Réinjection de Pertinence dans les Documents Semi-Structurés*". 2007.

[Moulahi, 2015] : B. Moulahi. « *Définition et évaluation de modèles d'agrégation pour l'estimation de la pertinence multidimensionnelle en recherche d'information* », Recherche d'information [cs.IR]. Université Toulouse III Paul Sabatier, 2015. Français. <tel-01249652>

[Naouar, 2017] : Fatiha Naouar. « *Un modèle possibiliste de recherche d'information collaborative basée sur la réinjection de pertinence* », Université de SFAX, 2017.

[Panjarkar et al., 2014]: Panjarkar Anmol G., Shete Devakinandan S., Mr. S. A. Kahate, (2014), "personalized web search with user's profile in re-ranking". *International journal of science, engineering and technology (ijset) ISSN: 2348-4098. vol 2 issue 8, nov-dec 2014.*

[Rijsbergen, 1979]: C. van Rijsbergen. "*Information Retrieval. Butterworth & Co (Publishers)*

Ltd", second edition, London, 1979.

[Robertson, 1976]: S.Robertson & K.Sparck Jones, "*Relevance Weighting for Search Terms"*. Journal of the American Society for Information Science, Vol 27, N°3, 1976.

[Rocchio, 1971]: J.J. Rocchio. "*Relevance feedback in information retrieval. In The SMART retrieval system-experiments in automatic document processing"*, pages 313,323. Prentice Hall Inc, 1971.

[Salton, 1968]: Gerard Salton, Michael Lesk, "*Computer Evaluation of Indexing and Text Processing"*. J. ACM 15(1): 8-36 (1968).

[Salton, 1971]: G. Salton. "*The SMART Retrieval System: Experiments in Automatic Document Processing"*, Prentice-Hall Inc, NJ.1971.

[Salton, 1988]: Salton, G. and Buckley, C., "*Term Weighting Approaches in Automatic Text Retrival*", Information Processing and Management. 513-523, 1988.

[Saoud et Kechid, 2016]: ZakariaSaoud ,SamirKechid, "*integrating social profile to improve the source selection and the result merging process in distributed information retrieval"*, Information Sciences 336 (2016) 115–128. http://dx.doi.org/10.1016/j.ins.2015.12.012.

[Saracevic, 1997]: T. Saracevic. The strate_ed, "*model of information retrieval interaction: extension and applications"*. In *Proceedings of the 60th annual meeting of the American Society for Information Science*, pages 313.327, Medford, NJ, 1997.

[Shaw et al., 1994]. SHAW J.A., FOX E.A., SHAW J.A., FOX E.A., 1994, « *Combination of Multiple Searches* », *The Second Text REtrieval Conference (TREC-2*, p. 243–252.

[Sieg et al., 2004a] : A. Sieg, B. Mobasher, et R. Burke."*User's information context : Integrating user profiles and concept hierarchies"*. Dans *Proceedings of the 2004 Meeting of the International Federation of Classification Societies*, numéro 1, pages 28–40, 2004a. (Cité page 9).

[Tamine, 2008] : Lynda Tamine-Lechani. « *De la recherche d'information orientée sytème à la recherche d'information orientée contexte : Verrous, contributions et perspectives »*. Informatique [cs]. Université Paul Sabatier- Toulouse III, 2008. <tel-00355842>

[Tamine& al, 2005]: Lynda Tamine, Nesrine Zemirli, and Mohand Boughanem. « *Accès personnalisé à l'information: Proposition d'un profil utilisateur multidimensionnel »*. In International Symposium On Programming Systems (ISPS), Alger, 2005.

[Tamine et al., 2008]: L. Tamine-Lechani, M. Boughanem, et N. Zemirli. "*Personalized document ranking : exploiting evidence from multiple user interests for profiling and retrieval"*. to appear. *In Journal of Digital Information Management*,2008. (Cité pages 25, 26, 27, 33, 44 et 83).

[Tardieu et al., 1983] : Tardieu H., Rochfeld A., Coletti R., « *La méthode Merise »*, tome 1 : *Principes et outils*, éditions d'Organisation, 1983.

Références

[Tchuente, 2013] Tchuente, D. (2013). « *Modélisation et dérivation de profils utilisateurs à partir de réseaux sociaux : approche à partir de communautés de réseaux k-egocentriques* ».
Thèse de doctorat, Université Paul Sabatier, Toulouse, France.

[Tchuente et al., 2013]: Tchuente D., CANUT M.-F., JESSEL N., PENINOU A., SÈDES F., 2013, « *A community-based algorithm for deriving users' profiles from egocentrics networks: experiment on Facebook and DBLP* », *Social Network Analysis and Mining*, *3*, 3, p. 667 683.

[Toms et al., 2013] : Elaine Toms, Maristella Agosti, Norbert Fuhr and Pertti Vakkari. *(2013),* "Evaluation Methodologies in Information Retrieval". *In: agstuhl Reports, Vol. 3, Issue 10, pp. 92–126.*

[Yates, 1999]: R. B.Yates, R. Neto, "*Modern Inforation Retrival"*. ACM Press, Addison Wesley, 1999.

[Zemirli, 2004] : W.N.Zemirli, « *vers le développement d'un système de recherche d'information personnalisé intégrant le profil utilisateur* », Université Paul Sabatier- Toulouse III, 2004.

[Zemirli, 2008] : Nesrine Zemirli. « *Modèle d'accès personnalisé à l'information basé sur les Diagrammes d'Influence intégrant un profil utilisateur évolutif* ». Informatique [cs]. Université Paul Sabatier - Toulouse III, 2008. Français. <tel-00355844>