



HAL
open science

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

Mariam Daoud

► To cite this version:

Mariam 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. Informatique [cs]. Université Paul Sabatier - Toulouse III, 2009. Français. NNT : . tel-00442931

HAL Id: tel-00442931

<https://theses.hal.science/tel-00442931v1>

Submitted on 24 Dec 2009

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

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

THÈSE

Présentée devant

l'Université Paul Sabatier de Toulouse

en vue de l'obtention du

Doctorat de l'Université Paul Sabatier

Spécialité : **INFORMATIQUE**

Par

Mariam 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

Soutenue le 10 Décembre 2009, devant le jury composé de :

M. C. CHRISMENT	Professeur à l'Université Paul Sabatier, Toulouse	Président
M. E. GAUSSIER	Professeur à l'Université Joseph Fourier, Grenoble	Rapporteur
M. I. OUNIS	Reader à l'Université de Glasgow, Glasgow	Rapporteur
Mme. M. FRANCINE-MOENS	Professeur à l'Université catholique, Belgique	Examinatrice
M. B. CHEBARO	Matre de confrences à l'Université libanaise, Liban	Examinateur
M. M. BOUGHANEM	Professeur à l'Université Paul Sabatier, Toulouse	Directeur de thèse
Mme. L. TAMINE-LECHANI	HDR MCF à l'Université Paul Sabatier, Toulouse	Co-directrice de thèse

INSTITUT DE RECHERCHE EN INFORMATIQUE DE TOULOUSE

Centre National de la Recherche Scientifique - Institut National Polytechnique - Université Paul Sabatier
Université Paul Sabatier, 118 Route de Narbonne, 31062 Toulouse Cedex 04. Tel : 05.61.55.66.11

Resumé

Le but fondamental de la recherche d'information (RI) contextuelle consiste à combiner des sources d'évidences issues du contexte de la requête et du contexte de l'utilisateur dans une même infrastructure afin de mieux caractériser les besoins en information de l'utilisateur et améliorer les résultats de la recherche.

Notre contribution porte sur la conception d'un système de RI personnalisé intégrant la caractérisation du type de recherche véhiculé par la requête ainsi que la modélisation et construction de profils sémantiques des utilisateurs. Nous exploitons le contexte de la requête dans un mécanisme de prédiction du type de recherche en tant qu'informationnel lié à la recherche par contenu, navigationnel lié à la recherche des sites d'accueil et transactionnel lié à la recherche des services en ligne.

Notre approche de prédiction du type de recherche consiste à combiner des sources d'évidences issues des caractéristiques morphologiques de la requête d'une part et le contexte de la session défini par le type de recherche porté par les requêtes de la session d'autre part. Nous nous intéressons ensuite à exploiter, dans un premier temps, le profil de l'utilisateur à court terme dans un processus d'accès personnalisé à l'information. Le profil à court terme traduit généralement un centre d'intérêt de l'utilisateur construit sur la base d'une ou plusieurs sessions de recherche traitant un même besoin en information. Ce profil est construit en exploitant les documents jugés pertinents par l'utilisateur et une ressource sémantique, en l'occurrence une ontologie de base, servant à la représentation conceptuelle du profil. Dans un second temps, le profil à long terme de l'utilisateur est construit par agrégation des profils utilisateurs à court terme construits au cours des sessions de recherche de l'utilisateur. Nous exploitons le profil de l'utilisateur à long terme dans l'initialisation du profil dans une session de recherche traitant un besoin en information récurrent. Le profil de l'utilisateur créé ou qui évolue dans une même session de recherche est utilisé dans le ré-ordonnement des résultats de recherche des requêtes appartenant à la même session.

Nous avons validé notre approche de prédiction du type de recherche sur une collection de requêtes annotées de TREC par comparaison à un classifieur classique. En outre, vu qu'il n'existe pas des cadres d'évaluation standards d'accès personnalisé à l'information, plus particulièrement adaptés à la

personnalisation à court terme, nous avons proposé des cadres d'évaluation orienté-contexte basé sur l'augmentation des collections de la campagne TREC par des profils utilisateurs et des sessions de recherche simulés. Nous avons exploité ces cadres d'évaluation pour valider notre contribution dans le domaine. Nous avons également comparé expérimentalement notre contribution à une approche de personnalisation reconnue dans le domaine et avons montré que notre approche est à l'origine d'un gain de performance significatif.

Remerciements

Je tiens à remercier très sincèrement Messieurs les Professeurs Claude Chrisment et Gilles Zurfluh, responsables de l'équipe SIG, pour m'avoir accueillie au sein de leur équipe.

Je tiens à exprimer ma profonde gratitude à Monsieur Mohand Boughanem, Professeur à l'Université de Toulouse III, pour avoir encadré et dirigé mes recherches. Je le remercie pour m'avoir soutenu et appuyé tout au long de ma thèse. Ses précieux conseils, son exigence et ses commentaires ont permis d'améliorer grandement la qualité de mes travaux et de ce mémoire.

Je tiens à remercier vivement ma codirectrice de thèse, Madame Lynda Tamine-Lechani, Maître de conférences, HDR à l'université de Toulouse III pour avoir encadré et dirigé mes recherches. Je la remercie pour son souci constant de l'avancement de ma thèse et son suivi continu de mon travail, ses précieux conseils de tous ordres, sa disponibilité et sa confiance. Par sa bonne humeur et sa collaboration, elle m'a toujours encouragé et aidé à surmonter les difficultés. Son expérience et ses grandes compétences ont permis l'accomplissement de ce travail. Qu'elle trouve ici les marques de ma reconnaissance et de mon respect.

Je remercie très sincèrement Monsieur Eric Gaussier, Professeur à l'Université de Grenoble et Monsieur Iadh Ounis, Reader à l'Université de Glasgow pour avoir accepté d'être rapporteurs de ce mémoire, et pour l'honneur qu'ils me font en participant au jury. Merci également à Madame Marie Francine-Moens, Professeur à l'Université catholique de Louvain (Belgique), à Monsieur Claude Chrisment, Professeur à l'Université de Toulouse III, et à Monsieur Bilal Chebaro, Maître de conférences à l'université Libanaise d'avoir accepté de juger ce travail et de faire partie du jury. Je les remercie pour leur évaluation scientifique et leur travail de synthèse.

Cette thèse a été financée en partie par l'association EGIDE pendant les deux premières années. Que les responsables de l'EGIDE et Monsieur Bilal Chebaro, qui m'a assuré ce financement et m'a encouragé à continuer en thèse depuis mon Master de recherche, soient assurés de mes remerciements sincères pour avoir supporté mes travaux.

Mes remerciements vont de même à tous les membres de l'équipe SIG de l'IRIT pour leur aide et leur gentillesse. Plus particulièrement, je tiens à ex-

primer ma reconnaissance à Cécile Laffaire. Je la remercie pour ses aides dès le début de ma thèse, sa disponibilité et sa générosité pour faire avancer mes expérimentations.

Je remercie mes amis de l'équipe qui ont contribué à la finalisation des quelques tâches d'évaluation dans ce mémoire. Je remercie Ourdia, Ihab, Saad et Amjad pour leur collaboration et leur disponibilité.

Merci à mes amis Ghinwa, Maya, Rania, Norma, Rola et Imad pour leur encouragement continu et leur fidélité malgré la distance qui nous sépare.

Mouna! je n'ai pas pensé à la difficulté d'exprimer combien tu comptes pour moi. Tu es ma belle sœur ' mayno' qui ne m'a jamais laissé dans les moments difficiles et qui m'a toujours encouragé à atteindre les buts. Je n'oublierai jamais les secrets de nos allers retour à l'IRIT, les nuits blanches, les conférences et les moments de joie pendant toute la période qu'on avait vécu ensemble. Le plus important est que je n'ai jamais senti la solitude à l'étranger grâce à toi. Je ne trouve pas la fin entre ces lignes!! je dirai : tu resteras toujours ma troisième sœur pour toujours.

Je remercie du fond du cœur mes parents, mes sœurs Fatima et Nada et mon frère Mohamed qui n'ont jamais cessé de croire en moi pendant toutes mes années d'études et qui m'ont toujours encouragé à atteindre mes ambitions. Merci à ma sœur Nada, le jumeau de mon âme depuis mon enfance, qui m'a encouragé à confronter les difficultés de voyage pour continuer mes études et qui était toujours proche de moi malgré la distance qui nous sépare.

Enfin et surtout, je remercie avec grand amour mon fiancé Ali qui m'a encouragé à y aller de l'avant. Il était la lumière étincelle qui se fortifie jour après jour et qui animait mon chemin avec plein d'émotions, d'enthousiasme, du bonheur et d'ambitions. C'est la première fois que le goût du succès est accompagné par un bonheur complet.

Une partie de ce travail de thèse est réalisée dans le cadre du projet QUAERO financée par OSEO, l'association française pour l'innovation.

Table des matières

Introduction générale	1
I Accès contextuel à l'information : aperçu sur les modèles et les cadres d'évaluation en RI contextuelle	11
1 Concepts de base de la RI classique et émergence de la RI adaptative	12
1.1 Introduction	12
1.2 Les fondements de la recherche d'information	13
1.2.1 Notions de base	13
1.2.2 Mise en œuvre d'un SRI	15
1.2.2.1 Indexation	15
1.2.2.2 Interrogation	17
1.2.3 Aperçu des modèles de RI	18
1.2.3.1 Modèle booléen	18
1.2.3.2 Modèle vectoriel	19
1.2.3.3 Modèle probabiliste	19
1.3 De la RI classique à la RI adaptative	20
1.3.1 Facteurs d'émergence de la RI adaptative	21
1.3.2 La RI adaptative	22
1.3.2.1 Reformulation de requêtes	22
1.3.2.2 Désambiguïsation du sens des mots de la requête	24
1.3.2.3 Regroupement thématique des résultats de recherche	25
1.3.3 Limitations de la RI adaptative	26
1.4 Conclusion	27
2 Émergence de la RI contextuelle	28
2.1 Introduction	28
2.2 La notion du contexte : définitions et taxonomies du contexte multidimensionnel	29
2.2.1 Qu'est ce que le contexte en RI?	29

2.2.2	Le contexte : un concept multidimensionnel	31
2.2.2.1	Taxonomie de Fuhr 2000	31
2.2.2.2	Taxonomie de Cool 2001	32
2.2.2.3	Taxonomie de Ingerwersen et al. 2005	33
2.2.2.4	Taxonomie de Tamine et al. 2009	34
2.2.3	Le contexte : définition et exploitation dans le cadre de nos travaux de thèse	37
2.3	Accès contextuel à l'information guidé par le contexte de la tâche de recherche	37
2.3.1	Cadre et motivations	38
2.3.2	Techniques d'identification du contexte de la tâche	39
2.3.3	Techniques d'exploitation du contexte de la tâche de recherche dans le processus de RI	40
2.4	Accès contextuel à l'information guidé par le profil utilisateur	41
2.4.1	Cadre et motivations	42
2.4.2	La notion de profil utilisateur	42
2.4.3	Architecture fonctionnelle d'un système de RI personnalisé (SRIP)	43
2.4.4	Représentation et construction du profil utilisateur	45
2.4.4.1	Acquisition et collecte des sources d'information	45
2.4.4.2	Représentation et dérivation du profil utilisateur	48
2.4.4.2.1	Représentation basée sur l'historique de recherche	48
2.4.4.2.2	Représentation ensembliste	49
2.4.4.2.3	Représentation connexionniste	52
2.4.4.2.4	Représentation conceptuelle	55
2.4.5	Évolution du profil utilisateur	58
2.4.5.1	Évolution du profil utilisateur à court terme	58
2.4.5.2	Évolution du profil utilisateur à long terme	60
2.4.5.3	Approches de délimitation des sessions de recherche	61
2.4.5.4	Synthèse des approches de modélisation du profil utilisateur	64
2.4.6	Les modèles d'accès personnalisé à l'information	65
2.4.6.1	Modèles de reformulation de requêtes	66
2.4.6.2	Modèles d'appariement personnalisé requête-document	69
2.4.6.2.1	Modèles d'ordonnancement personnalisé basé-contenu	69
2.4.6.2.2	Modèles d'ordonnancement personnalisé basé-hyperliens	70

2.4.6.3	Modèle d'ordonnancement personnalisé basé-contenu	72
2.4.6.4	Synthèse des modèles d'accès personnalisé à l'information	73
2.5	Conclusion	75
3	Accès personnalisé à l'information : de l'évaluation orientée-laboratoire à l'évaluation orientée-contexte	76
3.1	Introduction	76
3.2	Les approches d'évaluation des SRI orientées-laboratoire	77
3.2.1	Le paradigme de <i>Cranfield</i>	77
3.2.2	Collection de test	78
3.2.2.1	Collection de documents	78
3.2.2.2	Collection de requêtes	80
3.2.2.3	Jugements de pertinence	81
3.2.3	Mesures d'évaluation	82
3.2.4	Protocole d'évaluation	85
3.2.5	Limites de l'évaluation orientée-laboratoire en présence du contexte	86
3.2.5.1	Limites des collections de test	86
3.2.5.2	La non-exhaustivité des mesures d'évaluation	87
3.3	Vers l'évaluation orientée-contexte pour une tâche d'accès personnalisé à l'information	88
3.3.1	Objectifs de l'évaluation orientée-contexte	88
3.3.2	Approches d'évaluation orientées-contexte	90
3.3.2.1	Évaluation de la qualité du profil utilisateur	90
3.3.2.2	Évaluation de l'efficacité du modèle de RI personnalisée	91
3.3.2.2.1	Approches d'évaluation orientées-contexte initiées par TREC	91
3.3.2.2.2	Approches d'évaluation par simulation des contextes	92
3.3.2.2.3	Approches d'évaluation par utilisation des contextes réels	93
3.3.3	Mesures d'évaluation orientées-contexte	93
3.3.3.1	Mesures d'évaluation de la qualité du profil	94
3.3.3.2	Mesures d'évaluation de l'efficacité de RI	95
3.3.3.2.1	Mesures d'évaluation orientées-système	95
3.3.3.2.2	Mesures d'évaluation orientées-tâche	97
3.4	Les approches d'évaluation orientée-contexte initiées par TREC	98
3.4.1	La tâche interactive de TREC	98
3.4.1.1	Aperçu des tâches interactives de TREC	98

3.4.1.2	Composantes expérimentales de la tâche interactive de TREC	99
3.4.2	La tâche HARD de TREC	101
3.4.2.1	Aperçu de la tâche HARD de TREC	101
3.4.2.2	Composantes expérimentales de la tâche HARD de TREC	102
3.4.3	Discussion	104
3.5	Évaluation des SRI personnalisés par simulation de contexte	104
3.5.1	La simulation du contexte	104
3.5.2	Composantes expérimentales de l'évaluation par simulation de contexte	105
3.5.3	Discussion	108
3.6	Évaluation des SRI personnalisés par utilisation des contextes réels	108
3.6.1	Principe de l'évaluation	108
3.6.2	Composantes expérimentales de l'évaluation par utilisation des contextes réels	110
3.6.3	Discussion	113
3.7	Conclusion	113

II Conception et évaluation d'un modèle d'accès personnalisé à l'information basé-session dans des cadres d'évaluation orientés-contexte **115**

4	Prédiction contextuelle du type de besoin derrière la requête : proposition et évaluation	116
4.1	Introduction	116
4.2	Problématique et motivations	117
4.3	Prédiction du type du besoin derrière la requête	118
4.3.1	Estimation du type de besoin selon la morphologie de la requête	119
4.3.1.1	Le descripteur morphologique de la requête	119
4.3.1.2	Principe de l'estimation du type de besoin selon la morphologie	122
4.3.1.3	Exemple illustratif	123
4.3.2	Estimation du type de besoin de la requête selon le profil de la session	124
4.3.2.1	Principe de l'estimation du type de besoin selon le profil de la session	124
4.3.2.2	Exemple illustratif	126

4.3.3	Classification de la requête par combinaison de la morphologie et du profil de la session	126
4.3.3.1	Description de l'approche de combinaison	126
4.3.3.2	Exemple illustratif	127
4.4	Evaluation expérimentale de la prédiction du type de besoin	127
4.4.1	Description des ressources expérimentales	128
4.4.1.1	Collection de test	128
4.4.1.2	Outils et ressources d'extraction des caractéristiques de la requête	129
4.4.1.3	Outil de classification des requêtes par apprentissage supervisé basé sur TiMBL	130
4.4.2	Protocole d'évaluation	130
4.4.3	Mesures de performance	132
4.4.4	Résultats de la classification	133
5	Construction et évolution du profil utilisateur à base d'ontologie pour une recherche d'information personnalisée	135
5.1	Introduction	135
5.2	Problématique et motivations	136
5.3	Architecture du système	138
5.4	Définitions de base et notations	140
5.5	Définition du profil sémantique d'un utilisateur de SRI : approche basée sur la récurrence des sujets véhiculés par les sessions de recherche	141
5.5.1	Représentation sémantique du profil utilisateur	142
5.5.1.1	Représentation de la ressource sémantique : l'ontologie de l'ODP	142
5.5.1.2	Le profil comme une liste de concepts	145
5.5.1.3	Le profil comme un graphe de concepts	145
5.5.2	Principe de délimitation des sessions de recherche	147
5.5.2.1	Représentation conceptuelle de la requête	147
5.5.2.2	Mesures de délimitation des sessions de recherche	149
5.5.2.2.1	Mesure de corrélation basée-Kendall	149
5.5.2.2.2	Mesure de similarité basée-WebJaccard	150
5.5.3	Principe de définition du profil utilisateur à court et à long terme	150
5.5.3.1	Algorithme générique de définition du profil à court et à long terme	151
5.5.3.2	Illustration	153
5.6	Construction et évolution du profil de l'utilisateur à court terme	157
5.6.1	Construction du profil sémantique de la requête	157

5.6.2	Désambiguïsation du profil de la requête	158
5.6.2.1	Désambiguïsation par activation des liens hiérarchiques	158
5.6.2.2	Désambiguïsation par activation des liens de référence	159
5.6.3	Adaptation et évolution du profil utilisateur à court terme	162
5.6.3.1	Évolution du profil par combinaison linéaire des listes de concepts	162
5.6.3.2	Évolution du profil par combinaison des graphes de concepts	162
5.7	Construction et évolution du profil utilisateur à long terme . . .	164
5.7.1	Principes généraux	164
5.7.2	Aperçu des métriques de distance basé-graphe	166
5.7.2.1	Distance basée sur le plus grand sous-graphe commun (<i>mcs</i>)	166
5.7.2.2	Distance basée sur le plus petit super-graphe (<i>MCS</i>)	167
5.7.2.3	Distance basée sur la combinaison du <i>mcs</i> et <i>MCS</i>	167
5.7.2.4	Distance d'édition	167
5.7.3	Vers une mesure de distance sémantique entre graphes de profils	168
5.7.3.1	Motivations	168
5.7.3.2	Extension sémantique du plus grand sous-graphe commun <i>mcs</i>	170
5.7.3.3	La mesure de distance sémantique combinant <i>MCS</i> et <i>mcs</i>	171
5.8	Exploitation combinée du profil utilisateur à court terme et à long terme dans la personnalisation du processus de recherche .	171
5.9	Conclusion	173
6	Evaluation basée sur un cadre d'évaluation issu de TREC Adhoc :	
	mise en œuvre et résultats	174
6.1	Introduction	174
6.2	Le cadre d'évaluation issu de TREC <i>ad-hoc</i>	175
6.2.1	Collection de test	176
6.2.1.1	Requêtes	176
6.2.1.2	Collection de documents	177
6.2.2	Les sessions de recherche	177
6.2.3	Le profil utilisateur	178
6.2.4	Protocole d'évaluation	178
6.2.4.1	Protocole d'évaluation de la qualité du profil . .	178
6.2.4.2	Protocole d'évaluation de la RI personnalisée .	180

6.2.5	Métriques d'évaluation	180
6.2.5.1	Métriques d'évaluation de la qualité du profil	181
6.2.5.2	Métriques d'évaluation de l'efficacité de recherche	181
6.3	Évaluation de la qualité du profil	181
6.4	Évaluation de l'efficacité de recherche la personnalisée à court terme	183
6.4.1	Modèle de recherche classique basé sur BM25	183
6.4.2	Évaluation basée sur le profil comme une liste de concepts	183
6.4.2.1	Étude de l'impact du nombre de documents pertinents dans la construction du profil	184
6.4.2.2	Étude de l'impact du paramètre de réordonnement des résultats	184
6.4.2.3	Étude de l'impact de la profondeur du profil	186
6.4.2.4	Évaluation de l'efficacité de la recherche personnalisée	187
6.4.3	Évaluation basée sur le profil représenté par un graphe de concepts	189
6.4.3.1	Étude de l'impact du nombre de documents pertinents et du nombre de concepts du profil	190
6.4.3.2	Étude de l'impact du paramètre de réordonnement des résultats	191
6.4.3.3	Évaluation de l'efficacité de la recherche personnalisée	192
6.5	Évaluation de l'efficacité de la recherche la personnalisée à long terme	195
6.5.1	Évaluation de la précision d'identification des profils à long terme	196
6.5.2	Évaluation de l'efficacité de la recherche personnalisée à long terme	198
6.6	Conclusion	200
7	Évaluation basée sur un cadre d'évaluation issu de TREC HARD 2003 : propositions, mise en œuvre et résultats	202
7.1	Introduction	202
7.2	Le cadre d'évaluation issu de TREC HARD 2003 : un cadre dédié à une recherche personnalisée orientée-session	203
7.2.1	Collection de test	204
7.2.1.1	Collection de requêtes	204
7.2.1.2	Collection de sous-requêtes	205
7.2.1.2.1	Principe de génération des sous-requêtes et de simulation des sessions de recherche	206

7.2.1.2.2	Collection de sous-requêtes générées de la source <i>pertinence-qrels</i>	207
7.2.1.2.3	Collection de sous-requêtes générées de la source <i>pertinence-qrels-Top50</i>	208
7.2.1.3	Collection de documents	211
7.2.1.4	Les jugements de pertinence	212
7.2.2	Les sessions de recherche	212
7.2.3	Le profil utilisateur	212
7.2.4	Stratégie d'évaluation	213
7.2.4.1	Stratégie d'évaluation de la qualité du profil	213
7.2.4.2	Stratégie d'évaluation de l'efficacité de la recherche personnalisée	213
7.2.5	Métriques d'évaluation	214
7.3	Évaluation de la qualité du profil	215
7.3.1	Évaluation des profils générés de la source <i>pertinence-qrels</i>	215
7.3.2	Évaluation des profils générés de la source <i>pertinence-qrels-Top50</i>	217
7.4	Évaluation de l'efficacité de la recherche personnalisée	218
7.4.1	Modèle de recherche classique basé sur BM25	219
7.4.2	Évaluation du mécanisme de délimitation des sessions de recherche	219
7.4.2.1	Analyse de corrélation requête-profil	220
7.4.2.2	Évaluation de la précision de la délimitation des sessions de recherche	223
7.4.3	Étude des paramètres de personnalisation	226
7.4.3.1	Étude de l'impact du nombre de concepts utilisés dans le score personnalisé	226
7.4.3.2	Étude de l'impact du paramètre de réordonnement des résultats	228
7.4.4	Évaluation de l'effet de la propagation du score sur la recherche personnalisée	230
7.4.5	Évaluation de l'efficacité de la personnalisation	232
7.4.5.1	Évaluation sur la séquence des sous-requêtes de test	233
7.4.5.2	Évaluation sur l'ensemble des requêtes principales de test	238
7.5	Une évaluation comparative avec l'approche de <i>Gauch</i>	241
7.5.1	Aperçu de l'approche de (Gauch et al. 03)	241
7.5.2	Évaluation comparative et résultats	242
7.6	Conclusion	245

Conclusion générale	246
Bibliographie	252

Liste des tableaux

2.1	Classification des comportements utilisés comme mesures de pertinence implicite par Oard et Kim [OK01]	49
2.2	Synthèse des différents modèles de représentation du profil utilisateur	65
2.3	Techniques d'accès personnalisé à l'information	75
3.1	Les composantes expérimentales des tâches interactives de TREC	100
3.2	Composantes expérimentales de la tâche HARD de TREC . . .	102
3.3	Composantes expérimentales de l'évaluation par simulation des contextes	106
3.4	Composantes expérimentales de l'évaluation par utilisation des contextes réels	111
4.1	Exemple des caractéristiques d'une requête transactionnelle . . .	123
4.2	Choix des collections de requêtes	129
4.3	Propriétés de la collection WT10G	130
4.4	Résultats de classification des requêtes de test	134
5.1	Les premiers concepts associés à la requête <i>encephal mosquito spray</i>	154
5.2	Les premiers concepts associés à la requête <i>state encephal nile</i> .	155
5.3	Les premiers concepts associés au profil de la requête <i>state encephal nile</i>	156
5.4	Les premiers concepts associés à la requête <i>earthquakes state</i> . .	156
6.1	Domaines de TREC choisies pour la simulation des profils utilisateurs	177
6.2	Caractéristiques de la collection de test TREC 1	177
6.3	Exemple du profil sémantique obtenu pour le domaine <i>Environment</i> à juger par des utilisateurs réels	179
7.1	Un exemple de sous-requêtes générées de la source pertinence-qrrels	208
7.2	Caractéristiques de la collection de test TREC HARD 2003 . . .	212

Table des figures

1	Les moteurs de recherche d'information sur le Web	2
1.1	Processus en U de la RI	15
1.2	Désambiguïsation du sens des mots de la requête <i>Java</i> sur <i>Google</i>	24
2.1	Dimensions du contexte multidimensionnel de [Fuhr 2000]	32
2.2	Le concept du contexte multidimensionnel en RI	34
2.3	Architecture fonctionnelle d'un SRIP	45
2.4	L'historique de recherche utilisé dans <i>Google psearch</i> qui sauvegarde les requêtes et les résultats sélectionnés par l'utilisateur	50
2.5	Représentation du profil utilisateur et du document dans le système <i>Wifs</i>	53
2.6	Exemple d'un profil utilisateur représenté par un hiérarchie de concepts	54
2.7	Exemple d'un profil utilisateur représenté par un graphe de termes reliés par des opérateurs logiques	55
2.8	Représentation textuelle du contexte de recherche sur <i>Yahoo Y!Q</i>	66
2.9	Les sessions de requêtes sont à la base des corrélations entre les termes des requêtes, les éléments contextuels et les termes des documents	68
3.1	Topic 705 de la tâche Terabyte de TREC	80
3.2	Technique de <i>Pooling</i> pour la création des collections de test larges	82
3.3	Forme générale de la courbe de précision-rappel d'un SRI	84
4.1	Architecture générale de notre modèle	118
4.2	Exemple d'une séquence des sessions de type informationnel, navigationnel et transactionnel	133
5.1	Architecture générale de notre modèle	139
5.2	L'interface de l'ontologie de domaine de l'ODP	143
5.3	Une portion d'un profil utilisateur représenté sous forme d'un graphe issu de l'ODP.	146

5.4	Désambiguïisation du profil de la requête par activation des liens hiérarchiques	159
5.5	Évolution du profil de l'utilisateur à court terme	163
5.6	Extension sémantique du graphe par activation des liens de référence	170
5.7	(a) les deux graphes n'ont pas des liens de références entre eux, (b) les deux graphes ont des liens de référence entre eux	172
6.1	Protocole d'évaluation basée sur la stratégie de la cross-validation appliqué sur TREC adhoc	180
6.2	Evaluation de la qualité des profils graphes construits sur la collection TREC adhoc	182
6.3	Variation de l'efficacité de la recherche personnalisée en fonction du nombre de documents pertinents pour le profil conceptuel ensembliste sur la collection TREC Adhoc	185
6.4	Variation de l'efficacité de la recherche personnalisée en fonction du paramètre de réordonnancement des résultats pour le profil représenté par une liste de concepts sur la collection TREC Adhoc	186
6.5	Variation de l'efficacité de la recherche personnalisée en fonction de différentes représentations du profil sur la collection TREC Adhoc	187
6.6	Gain de performance de la recherche personnalisé comparativement à la recherche classique pour le profil conceptuel ensembliste sur la collection TREC Adhoc	188
6.7	Gain de performance calculé par domaine de la recherche personnalisé comparativement à la recherche classique pour le profil conceptuel ensembliste sur la collection TREC Adhoc	189
6.8	Variation de la recherche personnalisée à P10 en fonction des paramètres de construction du profil graphe sur la collection TREC Adhoc	191
6.9	Variation de la précision MAP de la recherche personnalisée en fonction des paramètres de construction du profil graphe sur la collection TREC Adhoc	191
6.10	Variation de l'efficacité la recherche personnalisée en fonction du paramètre de réordonnancement des résultats pour le profil graphe sur la collection TREC Adhoc	192
6.11	Évaluation de la performance du modèle de RI personnalisée en termes de Top-n précision moyenne et Top-n rappel moyen en utilisant le profil graphe sur TREC adhoc	193
6.12	Evaluation du gain de performance de la recherche personnalisée en termes de Top-n précision moyenne et Top-n rappel moyen en utilisant en utilisant 25 documents pertinents sur TREC Adhoc	194

6.13	Evaluation de gain de performance de la recherche personnalisée en termes de la précision P10 et la MAP par domaine de TREC Adhoc	194
6.14	Taux de la corrélation conceptuelle entre profils de domaines . . .	197
6.15	Évaluation de la précision d'identification des profils à long terme adéquats aux requêtes issus des domaine de TREC Adhoc	198
6.16	Évaluation de l'efficacité de la RI personnalisée à long terme sur TREC Adhoc	199
6.17	Évaluation de l'efficacité de la RI personnalisée à long terme par domaine sur TREC Adhoc	200
7.1	Principe de génération des sous-requêtes d'une même requête de la collection TREC HARD 2003 selon la <i>source pertinence-qrels-Top50</i>	207
7.2	Taux de recouvrement de pertinence des requêtes et des sous-requêtes en termes du nombre de documents pertinents	209
7.3	Taux de non-recouvrement moyen des sous-requêtes en termes du nombre de documents différents sur les Top-20 et Top-50 documents	209
7.4	Taux de recouvrement et de pertinence moyen et de non-recouvrement de documents des sous-requêtes générées à partir de la source <i>pertinence-qrels-Top50</i>	210
7.5	Comparaison du taux de recouvrement de pertinence moyen des requêtes, des sous-requêtes générées à partir de la source <i>pertinence-qrels</i> et des sous-requêtes générées à partir de la source <i>pertinence-qrels-Top50</i> retournés par le système	211
7.6	Phase de test du protocole d'évaluation appliqué sur TREC HARD 2003	214
7.7	Qualité des profils construits pour les requêtes principales de la collection TREC HARD en utilisant la source <i>pertinence-qrels</i> .	216
7.8	Qualité des profils construits pour les requêtes principales de la collection TREC HARD en utilisant le profil pertinence de 10 documents	217
7.9	Phase d'apprentissage du protocole d'évaluation appliqué sur TREC HARD 2003	220
7.10	Analyse de corrélations requête-profil selon les mesures Kendall, WebJaccard et cosinus calculées le long de la séquence des sous-requêtes d'apprentissage	222
7.11	Précision de classification des sessions selon les mesures de <i>Kendall</i> et <i>WebJaccard</i> et <i>Cosinus</i> en fonction de la variation du seuil de corrélation	225

7.12	Efficacité de la recherche personnalisée en fonction du nombre de concepts considérés dans le calcul du score personnalisé en utilisant la source pertinence-qrels sur la collection TREC HARD	227
7.13	Efficacité de la recherche personnalisée en fonction du nombre de concepts considérés dans le calcul du score personnalisé en utilisant la source pertinence-qrels-Top50 sur la collection TREC HARD	227
7.14	Efficacité de la recherche personnalisée en fonction du paramètre de réordonnement en utilisant la source pertinence-qrels documents sur la collection TREC HARD	229
7.15	Efficacité de la recherche personnalisée en fonction du paramètre de réordonnement en utilisant la source pertinence-qrels-Top50 sur la collection TREC HARD	230
7.16	Validation du processus de propagation des scores lors de la construction du profil sur la collection TREC HARD	232
7.17	Comparaison de la recherche personnalisé à la recherche classique en termes de la précision moyenne et rappel moyen sur la séquence des sous-requêtes de test	235
7.18	Comparaison de la recherche personnalisé à la recherche classique en termes de la précision moyenne et rappel moyen sur la séquence des sous-requêtes de test	237
7.19	Comparaison de la recherche personnalisé à la recherche classique en termes de la précision moyenne et rappel moyen sur la séquence des sous-requêtes de test	237
7.20	Comparaison de la recherche personnalisé à la recherche classique en termes de la précision moyenne et rappel moyen sur la séquence des requêtes principales de test	239
7.21	Comparaison de la recherche personnalisé à la recherche classique en termes de la précision moyenne et rappel moyen sur les requêtes principales de test	240
7.22	Comparaison de la performance de notre approche de personnalisation à l'approche de (gauch et al. 03) en termes de la précision moyenne et rappel moyen sur la séquence des requêtes principales de test	243
7.23	Comparaison du pourcentage de gain de performance obtenu par notre approche de personnalisation et l'approche de (Gauch et al. 03) sur l'ensemble des requêtes principales de TREC HARD 2003	244

Introduction générale

Contexte et problématique

La recherche d'information (RI), est une branche en informatique qui s'intéresse à l'acquisition, l'organisation, le stockage et la recherche des informations. C'est l'ensemble de procédures et techniques permettant de sélectionner, parmi un ensemble de documents, les informations (documents ou parties de documents) pertinentes en réponse à un besoin en information exprimé par l'utilisateur à travers une requête.

Le domaine de la RI remonte au début des années 1950, peu après l'invention des ordinateurs ; il est né pour automatiser la RI dans les bibliothèques. La première problématique en RI portait sur l'indexation des documents qui consiste à établir des représentations des documents, à travers la construction d'index, dans le but d'en récupérer des informations. Au sens large, la RI inclut deux aspects : l'indexation des collections de documents et l'interrogation du fonds documentaire ainsi construit. L'indexation des documents consistent tout d'abord à extraire à partir du texte des documents, des descripteurs qui consistent souvent en l'ensemble de termes (souvent lemmatisés) qui apparaissent dans les documents formant ainsi l'index, et puis représenter le document par un vecteur dans l'espace de termes. Une fois l'index est créé, la recherche des documents en réponse à une requête de l'utilisateur est ainsi possible via des modèles d'appariement requête-document. Les premiers systèmes de RI utilisés par des libraires sont fondés sur des modèles de recherche booléens. Le développement du système SMART par Salton [158] à la fin des années 1960 (1961-1965) a permis de remonter des modèles élaborés de RI, tels que le modèle vectoriel, le modèle probabiliste, etc. La mise au point de différentes méthodes d'indexations et de différents modèles de RI passe par une phase d'évaluation qui a pour but de mesurer leurs efficacité/performance en réalité. Le projet de *Cranfield* dirigé par *Cleverdon* (1957-1967), était le premier à mettre en place un paradigme d'évaluation utilisé encore aujourd'hui pour l'évaluation des systèmes de RI (SRI). L'évaluation selon ce paradigme est vigoureusement contrôlée qui suppose l'utilisation des métriques permettant de comparer deux méthodes de recherche sur la base d'une collection de test constituée d'un en-

semble de documents et de requêtes. Ces requêtes sont évaluées par des experts afin de déterminer les réponses souhaitées (les documents pertinents du point de vue de l'utilisateur). Finalement, pour mesurer la performance d'un SRI, les résultats d'une recherche automatique retrouvés par le système selon un modèle de RI sont comparés avec les réponses souhaitées. La pertinence d'un document pour une requête au niveau système s'exprime sous forme d'un score calculé selon le modèle d'appariement requête-document. La pertinence d'un document du point de vue de l'utilisateur est estimé selon des critères liés principalement à son besoin en information.

L'explosion des ressources d'information disponibles et leur hétérogénéité à travers l'essor du web dans les années 1990 a propulsé la RI au premier plan. Suite à cette explosion, plusieurs moteurs de recherche ont apparu sur le marché permettant l'accès à l'information à grande échelle (*Lycos, Yahoo!* et *Altavista*) (cf. Figure 1). Cette explosion apporte de problèmes à la RI auxquels elle ne s'est jamais confrontée : collection gigantesque, dynamique et changeante, surabondance de l'information, données multimédia, données réparties, multilinguisme, etc. Ces problèmes ont remis en cause les modèles classiques de RI ; en effet, en présence des documents non textuels (image, son, vidéo, etc.), les méthodes classiques d'indexation et de recherche en RI sont surtout destinées aux textes, et ne sont pas directement applicables à d'autres médias. C'est ainsi que de nouvelles technologies de l'information sont définies : la RI multimédia, la RI multilingue et les systèmes de RI répartis.

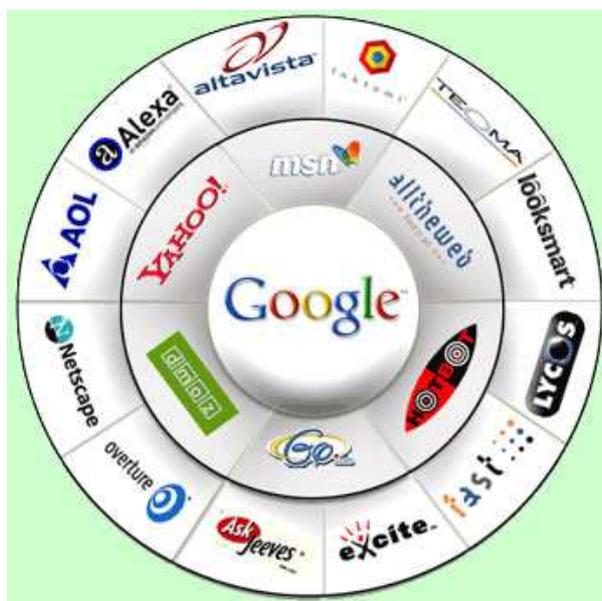


FIG. 1 – Les moteurs de recherche d'information sur le Web

Nous nous intéressons au problème de la surabondance de l'information qui a également remis en cause les modèles classiques de RI. ceux ci se ca-

ractérisent par un processus d'accès à l'information dépendant seulement de la disponibilité de l'information et de critères de sélection par le contenu. Selon cette hypothèse de recherche, plus le volume de l'information est grand, plus le système renvoie de documents à l'utilisateur qui se trouve ainsi face à une surcharge informationnelle qui le désoriente où il prend la charge de sélectionner ce qui est pertinent de ce qui ne l'est pas. Cette problématique a motivé les recherches au développement des systèmes question-réponse (QA) dont le but est d'identifier directement la réponse à une requête utilisateur plutôt que de proposer un ensemble de documents dans lesquels l'utilisateur peut trouver une réponse (ou l'information recherchée). Dans le but de clarifier le besoin en information de l'utilisateur, les chercheurs sur QA ont appliqué des techniques d'analyse du langage naturel (TAL). Le défi majeur qui se pose revient à la capacité de la machine à comprendre le besoin en information de l'utilisateur. Ceci est toujours le problème clé à surmonter.

Même avec le développement des systèmes QA idéaux, la surabondance de l'information reste une problématique en présence de l'ambiguïté des requêtes des utilisateurs, la diversité de leurs besoins en information et de leurs situations de recherche. Les modèles classiques en RI sont basés sur une approche généraliste qui répond invariablement les utilisateurs en renvoyant une même liste de résultats pour deux utilisateurs ayant émis la même requête et ayant pourtant des besoins en informations et des préférences de recherche différentes. Cette problématique est d'autant plus aigüe suite à la prolifération des technologies mobiles d'accès à l'information (les agendas personnels, la téléphonie mobile, etc.) qui font de la localisation géographique de l'utilisateur un élément à prendre en compte dans le processus de filtrage ou de sélection de l'information.

Les premières tentatives proposées permettant de pallier cette problématique s'apparentent à la RI adaptative. Le but de la RI adaptative est d'adapter le processus de RI de manière à retourner de l'information pertinente répondant aux caractéristiques et aux préférences de recherche de l'utilisateur. Les techniques dérivées de ce cadre concernent la reformulation de requête par réinjection de pertinence, l'expansion des requêtes par désambiguïsation, développées durant les années 1970 [147] ou le regroupement thématique des résultats, développées durant la période 1960 à 1970 [24]. Malgré l'efficacité de la recherche obtenue par ces techniques, elles sont généralement limitées par la rétroaction explicite de la part des utilisateurs. En outre leur efficacité est relativement dépendante du niveau de familiarité de l'utilisateur avec le sujet de recherche et ne seront efficaces qu'après plusieurs itérations de recherche. Outre ces limitations techniques, ces approches ne permettent pas d'adapter le processus de recherche à des besoins en information évolutifs tels que les centres d'intérêts ou à un contexte d'interactions avec le système permettant de reconnaître les utilisateurs et leurs préférences dans des conditions d'utilisation du système différentes.

Suite aux limitations des techniques de la RI adaptative et dans le but d'améliorer la performance des SRI, plusieurs études ont été menées dans le but de mieux cerner la notion de pertinence du point de vue de l'utilisateur et d'identifier les différents facteurs ayant un impact sur cette notion et par conséquent sur la performance des SRI. Ces études confirment que la pertinence n'est pas une relation isolée entre un document et une requête ; elle est définie en fonction du contexte dans lequel la recherche est effectuée. Les premières études effectuées dans ce sens [33, 94, 83] se sont focalisées sur l'impact de la dimension cognitive du contexte, définie par les centres d'intérêts de l'utilisateur et ses préférences de recherche sur la notion de pertinence. Ces différentes études ont permis de définir clairement la notion du contexte comme étant l'ensemble des facteurs impliqués durant l'interaction utilisateur-système, ainsi que de mener des réflexions avancées autour des taxonomies de ses différentes dimensions. C'est ainsi qu'une nouvelle direction de recherche basée sur la RI contextuelle fut apparaître. Selon [16], la RI contextuelle est définie comme suit : "*Combine search technologies and knowledge about the query and user context into a single framework in order to provide the most appropriate answer for a user's information needs*".

Selon les différentes définitions du contexte dans la littérature, le contexte de recherche est lié globalement à des dimensions dépendantes de l'utilisateur, de la requête et de l'environnement de recherche. Les problématiques majeures posées par la RI contextuelle concernent principalement la modélisation du contexte, l'évolution du contexte au cours des sessions de recherche de l'utilisateur, l'exploitation du contexte dans le modèle de RI et finalement l'évaluation d'un modèle d'accès contextuel à l'information.

Les modèles développés en RI contextuelle se différencient entre eux par la définition et la modélisation du contexte prise en compte dans le processus de RI. Parmi ces modèles, nous citons particulièrement ceux qui répondent au contexte défini par le type de besoin derrière la requête (informationnel, navigationnel ou transactionnel), les modèles qui répondent au contexte défini par des critères et des préférences qualitatifs de l'information (fraîcheur, genre, etc.), les modèles qui répondent au contexte cognitif de l'utilisateur, défini par ses centres d'intérêts, et qui s'apparentent à la RI personnalisée et les modèles qui répondent au contexte physique défini par localisation géographique de l'utilisateur et qui s'apparentent à la RI mobile.

En s'intéressant à la RI personnalisée, le contexte cognitif de l'utilisateur est modélisé dans une structure informationnelle, appelée *profil*. Les approches développées dans ce domaine se distinguent par la technique de modélisation du profil de l'utilisateur et de son exploitation dans la chaîne d'accès à l'information. La modélisation du profil utilisateur repose sur des techniques et des outils permettant non seulement de capturer et de le représenter mais aussi de gérer son évolution de manière dynamique au cours du temps. Les premières

approches de représentation du profil étaient basés sur des modèles de sacs de mots puis se sont évolués pour supporter des modèles connexionistes révélant des relations de corrélations entre les centres d'intérêts constituant le profil ou conceptuels basés sur l'utilisation des ontologies et des taxonomies de domaine prédéfinies. Les techniques de dérivation du profil utilisateur peuvent être explicites nécessitant une rétroaction explicite de la part de l'utilisateur ou implicites souvent basées sur l'exploitation de l'historique de recherche et d'interaction de l'utilisateur permettant de dégager ses centres d'intérêts et ses préférences. Quant à l'exploitation du profil dans le processus de RI, les premières techniques développées s'inscrivent dans le cadre du filtrage d'information [23] et de recommandation [121, 139] utilisant le profil utilisateur dans la dissémination de l'information. Les techniques les plus élaborées se basent sur l'intégration du profil utilisateur dans l'une des phases du processus de RI, notamment la reformulation de requêtes [165, 109], l'appariement requête-document [180, 19] ou la présentation des résultats [67, 164, 123].

L'émergence de la RI personnalisée a permis non seulement de réviser la chaîne d'accès à l'information mais aussi de remettre en cause le paradigme d'évaluation classique des SRI (paradigme de *Cranfield*). Les limites majeures de ces cadres d'évaluation sont principalement liées d'une part à la collection de test qui n'intègre pas le contexte de recherche comme étant une source d'évidence pour l'évaluation de pertinence des documents, et d'autre part au protocole d'évaluation qui ne tient pas en compte des interactions de l'utilisateur avec le système en considérant que son besoin en information et sa perception de pertinence comme des notions statiques. On distingue deux types d'évaluations adaptés à la RI personnalisée : l'évaluation par simulation de contextes et l'évaluation par utilisation des contextes réels. Dans ces deux types d'évaluations, les cadres proposés sont fortement liés au SRI personnalisé à évaluer quant à la spécification des données exploitées pour la construction du profil utilisateur et la mesure d'évaluation en question.

Contribution

Les travaux présentés dans ce mémoire se situent dans le contexte précis de l'accès personnalisé à l'information. Plus précisément, nos contributions dans le domaine porte sur quatre volets : (1) la définition du contexte selon deux dimensions ; le type de besoin (informationnel, navigationnel ou transactionnel) et le profil utilisateur, (2) la modélisation du contexte de recherche, plus précisément, l'identification du type de besoin et la construction et l'évolution du profil utilisateur à court et à long terme, (3) l'exploitation du profil utilisateur dans le personnalisation du processus de RI, (4) la proposition d'un nouveau cadre d'évaluation basé-session qui permet d'évaluer l'efficacité de re-

cherche personnalisée basé-session.

Plus précisément, les grandes lignes de nos contributions sont les suivantes :

1. **Définition du contexte** : le contexte de recherche dans le cadre de nos travaux de recherche est défini par le type de besoin derrière la requête et le profil utilisateur à court et à long terme. Pour chacun des éléments du contexte ainsi défini, nous donnons notre motivation pour le choix de cet élément et les défis concernant sa modélisation comme suit :

- (a) Le type de besoin derrière la requête est considéré comme le premier élément contextuel à identifier dans le but d'exploiter les sources d'évidence les plus appropriées à chaque type de besoin servant ainsi à l'amélioration de la performance de recherche. Les défis majeurs à ce niveau consistent à définir au mieux des caractéristiques spécifiques à chaque type de besoin ainsi que de résoudre le problème de l'ambiguïté du type de besoin.
- (b) Les centres d'intérêts de l'utilisateur caractérisant son profil constituent l'élément contextuel le plus important qui impactent la performance de recherche informationnelle. L'exploitation du profil de l'utilisateur permet de désambigüiser la recherche, de répondre à des besoins en information récurrents de l'utilisateur, et d'élucider 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.

2. **Modélisation du contexte de recherche** :

- (a) Prédiction du type de besoin derrière la requête : la prédiction du type de besoin derrière la requête consiste à exploiter non seulement les caractéristiques linguistiques et morphologiques de la requête mais aussi le contexte de la session courante défini par le type de besoin porté par les requêtes précédentes. Ceci permet en effet d'estimer au mieux le type de besoin et d'intégrer la notion du contexte dans la désambiguïtation des types des requêtes en fonction des requêtes précédentes.
- (b) Modélisation du profil de l'utilisateur :
Nous définissons une représentation sémantique du profil de l'utilisateur basée sur l'exploitation d'un sous-ensemble de documents jugés pertinents implicitement par l'utilisateur et une ontologie de domaines Web. Notre objectif est de dégager à partir des interactions de l'utilisateur avec le système, une représentation conceptuelle de ses centres d'intérêts persistants et à court terme. A cet

effet, la modélisation du profil utilisateur dans notre approche est caractérisée par :

- la représentation sémantique du profil de l'utilisateur selon un ensemble de concepts sémantiquement liées via l'ontologie de référence utilisée.
- un procédé de construction du profil de l'utilisateur basé sur la désambiguïsation des concepts associés par l'extraction des concepts pertinents reliés entre eux via les liens sémantiques de l'ontologie.
- un procédé d'évolution du profil utilisateur à court terme qui repose sur l'augmentation de l'importance des concepts récurrents dans la session. Ce procédé est conjointement lié à un mécanisme de délimitation conceptuelle des sessions de recherche qui permet de scruter le changement des centres d'intérêts au cours du temps.
- un procédé d'évolution du profil à long terme basé sur la récurrence des centres d'intérêts à court terme construits au cours des sessions de recherche de l'utilisateur.

3. Exploitation du profil utilisateur dans le processus de RI :

le profil utilisateur est exploité dans le réordonnement des résultats de recherche des requêtes traitant du même besoin en 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 court terme est exploité dans le réordonnement des résultats des requêtes émises et groupées dans la même session que le profil.

4. Évaluation des performances d'un processus d'accès personnalisé à l'information basé-session :

En l'absence de cadre d'évaluation standard pour la RI personnalisée, nous proposons deux cadres d'évaluation par simulation des contextes. Nous ré-utilisons dans un premier temps un cadre d'évaluation issu de la collection TREC Adhoc [180] qui, par la présence d'annotations prédéfinies des requêtes par le type de besoin, présente l'avantage d'évaluer l'efficacité de la recherche personnalisée dans des sessions de recherche simulées par les domaines de la collection. Nous proposons dans un deuxième temps un nouveau cadre d'évaluation issu de TREC HARD 2003 qui intègre des sessions de recherche et des profils utilisateurs simulés en l'absence d'annotations préalables des requêtes par des domaines d'intérêts. Un aperçu de ces cadres d'évaluation est le suivant :

- (a) Le cadre d'évaluation issu de TREC *Ad-hoc* : ce cadre est élaboré selon une approche d'évaluation par simulation des contextes. Il est basé sur l'augmentation de la collection TREC par des profils utilisateurs simulés par des domaines d'intérêt prédéfinis dans la collection. Dans ce cadre, l'évaluation de l'efficacité de la recherche personnalisée est basée sur la cross-validation appliquée sur l'ensemble des requêtes d'un même domaine de TREC représentant ainsi une session de recherche simulée.

- (b) Le cadre d'évaluation issu de *TREC HARD 2003* : ce cadre est élaboré selon une approche d'évaluation par simulation des contextes qui intègre des situations de recherche proches des situations réelles. L'objectif de ce cadre d'évaluation consiste à évaluer l'efficacité de la recherche personnalisée selon un scénario de recherche intégrant un basculement des centres d'intérêts de l'utilisateur sur des séquences des sessions de recherche simulées. Une session de recherche est simulée par un ensemble de sous-requêtes générées à partir d'une même requête principale de la collection et considérées comme étant de différentes reformulations possibles du même sujet de recherche. La construction des profils utilisateurs sur les sessions de recherche simulées est appliquée en parallèle avec le mécanisme de délimitation des sessions de recherche et l'évaluation de la recherche personnalisée en respectant la séquence des requêtes simulées.

Organisation de la thèse

Ce mémoire est constitué de deux parties principales. La première partie regroupe trois chapitres.

Le chapitre 1 présente les notions et concepts de base de la RI ainsi que les facteurs d'émergence de la RI adaptative. L'architecture générale d'un système de RI y est présentée ainsi que les principaux modèles qui sont à la base de la majorité des systèmes actuels. On y aborde ensuite les limites de la RI classique dans un environnement hétérogène et volumineux ainsi que les facteurs d'émergence de la RI adaptative. Les techniques de base de la RI adaptative ainsi que ses limites dans le cadre de l'adaptation du processus de RI à la notion du contexte multidimensionnel sont ensuite présentées .

Le chapitre 2 traite de l'émergence de la RI contextuelle selon deux voies de recherche ; la RI contextuelle guidée par le type de besoin informationnel, navigationnel ou transactionnel et la RI contextuelle guidée par le profil utilisateur, connu sous le nom de RI personnalisée. Selon la RI contextuelle guidée par le type de besoin, les différentes approches de prédiction du type de besoin derrière la requête ainsi que les différentes stratégies de recherche adaptées à chacun des types de besoin sont décrits. Selon la RI personnalisée guidée par le profil de l'utilisateur, les différents modèles de représentation du profil, les différents procédés de construction et d'évolution du profil, ainsi que les différents modèles d'accès personnalisé à l'information intégrant le profil dans l'une des phases du processus de recherche sont abordés.

Le chapitre 3 présente le principe du paradigme d'évaluation classique

de *Cranfield* et ses limites dans l'évaluation des SRI personnalisés. Ensuite, l'émergence des trois grandes tendances des approches d'évaluation orientées-contexte ainsi que des mesures d'évaluation orientées-contexte sont décrites. Plus particulièrement, il s'agit des approches d'évaluation orientées-contexte initiées par TREC, les approches d'évaluation par simulation de contexte et les approches d'évaluation par utilisation de contextes réels.

La deuxième partie présente notre contribution dans le domaine de la RI personnalisée.

Le chapitre 4 présente notre approche de prédiction du type de besoin derrière la requête ainsi qu'une évaluation expérimentale de cette approche. Le principe de l'estimation du type de besoin selon les caractéristiques de la requête et selon le contexte de la session y est décrit. Concernant l'évaluation expérimentale, un protocole d'évaluation intégrant des séquences de sessions simulées de différents types de requêtes ainsi que les résultats de l'évaluation y sont présentés.

Le chapitre 5 présente notre contribution à la définition d'un profil utilisateur sémantique à court et à long terme pour une recherche d'information personnalisée basée-session. Deux modèles de représentation sémantiques du profil et le principe de délimitation des sessions de recherche sont présentés. Un algorithme général de définition du profil à court et à long terme, accompagné d'un exemple illustratif du déroulement du procédé de construction du profil dans une session, y est détaillé. Ensuite, la technique de construction et d'évolution du profil à court et à long terme y est décrite. Enfin, l'exploitation du profil dans le processus de recherche y est décrit.

Le chapitre 6 présente le cadre d'évaluation inspiré de *TREC adhoc* ainsi que l'évaluation de notre approche de personnalisation selon ce cadre en utilisant la collection de la tâche *adhoc* de TREC 1. L'évaluation expérimentale selon ce cadre d'évaluation concernent principalement la qualité des profils appris, l'étude des différents paramètres de notre modèle ainsi que l'efficacité de la recherche personnalisée.

Le chapitre 7 présente notre contribution à la définition d'un nouveau cadre d'évaluation basé-session, inspiré de *TREC HARD*. Ce chapitre présente également les résultats expérimentaux obtenus pour l'évaluation de notre approche de personnalisation selon ce cadre en utilisant la collection de la tâche *HARD* de TREC. En plus des aspects évalués selon le cadre issu de *TREC Adhoc*, ce chapitre présente également une évaluation expérimentale comparative.

En conclusion, nous dressons le bilan de nos travaux réalisés dans le cadre de l'accès personnalisé à l'information. Nous introduisons ensuite les perspectives

de ces travaux selon la personnalisation de la RI et les cadres d'évaluation des SRI personnalisés.

Première partie

Accès contextuel à l'information : aperçu sur les modèles et les cadres d'évaluation en RI contextuelle

Chapitre 1

Concepts de base de la RI classique et émergence de la RI adaptative

1.1 Introduction

La recherche d'information (RI) est une discipline de recherche qui intègre des modèles et des techniques dont le but est de faciliter l'accès à l'information pertinente pour un utilisateur ayant un besoin en information. Le besoin en information de l'utilisateur est souvent formulé en langage naturel par une requête décrite par un ensemble de mots clés. Pour une requête utilisateur, un système de RI permet de retrouver un sous-ensemble de documents susceptibles d'être pertinents, à partir d'une collection de documents, en réponse à cette requête.

L'essor du web a remis la RI face 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. La limite majeure de la plupart des moteurs de recherche classiques est qu'ils retournent la même liste des résultats pour une même requête soumise par des utilisateurs étant dans des contextes et/ou des situations de recherche pourtant différents et par conséquent ayant des besoins en information différents. Les études [36] montrent que la problématique de ces systèmes réside en partie dans le fait qu'ils sont basés sur une approche généraliste qui considère que le besoin en information de l'utilisateur est complètement représenté par sa requête et délivrent alors des résultats en ne tenant compte que des critères de sélection par contenu et de la disponibilité des sources d'informations. Les premières techniques développées en RI dans le but de pallier cette problématique s'apparentent à la RI adaptative. Il s'agit du développement des techniques de refor-

mulation de requêtes [147], de désambiguïsation du besoin derrière les requêtes [165] ou du regroupement thématique des résultats de recherche (Vivisimo¹, Exalead², Grouper [202]).

Ce chapitre traite les concepts de base de la RI classique ainsi que l'émergence de la RI adaptative. La section 1.2 présente tout d'abord les fondements de la RI classique. La section 1.3 décrit l'évolution de la RI classique à la RI adaptative. Plus particulièrement, nous abordons les problématiques majeures de la RI classique en présence du contexte, ensuite nous présentons l'orientation des travaux vers la RI adaptative pour laquelle nous donnons un aperçu des techniques développées ainsi que les limitations inhérentes. La dernière section conclut le chapitre.

1.2 Les fondements de la recherche d'information

Un système de recherche d'information (RI) est un système qui permet de retrouver, à partir d'une collection de documents, les documents susceptibles d'être pertinents à un besoin en information d'un utilisateur exprimé sous forme d'une requête. Dans cette définition, il y a trois notions clés que nous allons présenter dans la suite : document, requête, pertinence.

1.2.1 Notions de base

– *Document*

On appelle document toute unité d'information qui peut constituer une réponse à un besoin en information/requête d'un utilisateur. Un document peut être un texte, un morceau de texte, une image, une bande vidéo, etc.

– *Requête*

Une requête est une formulation du besoin d'information d'un utilisateur. Elle peut être vue comme étant une description sommaire des documents ciblés par la recherche. Pour une recherche documentaire donnée, l'utilisateur doit soumettre une requête au moteur de recherche dans laquelle il spécifie les mots clés représentant son besoin en information.

– *Pertinence*

Selon les premières définitions de la pertinence [156], *la pertinence est la correspondance entre un document et une requête ; une mesure de l'infor-*

¹<http://clusty.com/>

²<http://www.exalead.com/search/>

mativité du document à la requête ; un degré de relation (chevauchement, etc.) entre le document et la requête ; etc. La notion de pertinence est le critère primaire pour l'évaluation des systèmes de recherche d'informations. Le processus de jugement de la pertinence de l'information est basé sur le degré de similitude de la représentation de la requête avec le contenu du document retrouvé par le système.

Cependant, de nombreuses études ont été menées [21, 140] autour de la notion de pertinence et montrent que la pertinence n'est pas une relation isolée entre le document et la requête et qu'elle est définie par un ensemble de critères et de préférences qui varient selon les utilisateurs. Ces critères sont des facteurs qui déterminent la pertinence accordée à l'information retrouvée par l'utilisateur dans un contexte de recherche précis. Les facteurs qui affectent les jugements de pertinence font l'objet de recherche de plusieurs études [40, 140, 21]. Nous citons les critères définis par [21] et groupés dans sept catégories : (1) le contenu informationnel des documents ; (2) le niveau d'expertise et de connaissances de l'utilisateur ; (3) les croyances et préférences de l'utilisateur ; (4) autres informations liées à l'environnement ; (5) les sources des documents ; (6) les documents comme des entités physiques ; et (7) la situation de l'utilisateur.

Compte tenu de ces facteurs, il existe plusieurs "pertinences" possibles entre un document et un besoin, nous en citons les quatre (4) les plus importantes [178] :

1. *pertinence algorithmique* : cette pertinence (appelée aussi pertinence système) est traduite par une mesure algorithmique basée sur le calcul de la pertinence de l'information par rapport à la requête en utilisant des caractéristiques des requêtes d'une part et des documents d'autre part. C'est le seul type de pertinence qui est indépendant du contexte ;
2. *pertinence thématique* : cette pertinence est traduite par la *topicalité* et est définie par le degré de couverture du document du thème évoqué par le sujet de la requête. C'est le seul type de pertinence appliqué dans les campagnes d'évaluation tels que TREC.
3. *pertinence cognitive* : c'est la pertinence liée au thème de la requête, selon la perception ou les connaissances de l'utilisateur sur ce même thème ; cette pertinence est caractérisée par une dynamique qui permet d'améliorer la connaissance de l'utilisateur via l'information renvoyée au cours de sa recherche.
4. *pertinence situationnelle* : cette pertinence est définie par l'utilité ou l'usabilité de l'information jugée relativement à la tâche de recherche exprimée par le besoin en information selon la perception de l'utilisateur. C'est une pertinence potentiellement dépendante du contexte de recherche et est vue comme étant une pertinence dynamique.

Il est à noter qu'un SRI idéal doit supporter un modèle de recherche

d'information qui rapproche la pertinence algorithmique calculée par le système aux jugements de pertinence donnés par des vrais utilisateurs.

1.2.2 Mise en œuvre d'un SRI

Le but fondamental d'un SRI est de sélectionner l'ensemble de documents pertinents répondant au besoin en information de l'utilisateur. La réalisation d'un tel système consiste principalement à mettre en œuvre un processus clé (processus en U de la RI) illustré dans la 1.1. Ce processus consiste en deux principales phases : l'indexation et l'interrogation.

1. L'indexation consiste à extraire et représenter le contenu des documents de manière interne sous forme d'index. Cette structure d'index permet de retrouver rapidement les documents contenant les mots clés de la requête.
2. L'interrogation est l'interaction d'un utilisateur final avec le SRI, une fois les documents sont représentés sous forme interne d'index. Suite à une requête utilisateur, le système calcule la pertinence de chaque document vis à vis de la requête utilisateur selon une mesure de correspondance du modèle de RI, et retourne la liste des résultats à l'utilisateur.

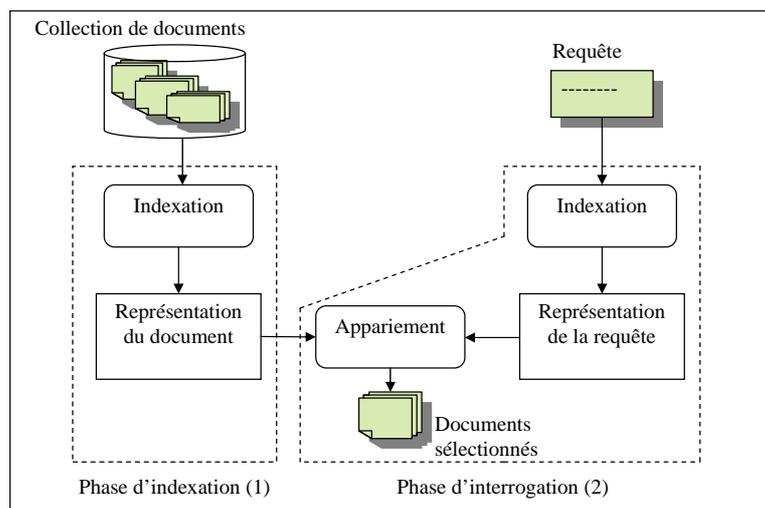


FIG. 1.1 – Processus en U de la RI

1.2.2.1 Indexation

L'indexation recouvre un ensemble de techniques visant à transformer les documents (ou requêtes) en substituts ou descripteurs capables de représenter leur contenu (Salton & McGill, 1983, 3.3.A). Ces descripteurs forment le langage d'indexation représenté selon une structure souvent basé sur un ensemble de mots clés ou groupes de mots représentant le contenu textuel du document.

Dès lors, l'indexation consiste à détecter les termes les plus représentatifs du contenu du document. Différents modes d'indexation existent en RI : l'indexation manuelle, automatique ou semi-automatique.

- *Indexation manuelle* : lors de l'indexation manuelle, un expert dans le domaine choisit les termes qu'il juge pertinents dans la description du contenu sémantique du document. Ce type d'indexation permet d'avoir un vocabulaire d'index contrôlé ce qui permet d'accroître la consistance et la qualité de la représentation obtenue. Toutefois cette approche est subjective d'une part car elle dépend des connaissances de l'opérateur et d'autre part inapplicable pour une collection volumineuse.
- *Indexation automatique* : Ce type d'indexation ne fait pas intervenir d'expert. L'indexation automatique repose sur des algorithmes associant automatiquement des descripteurs à des parties de document. Dans le cas des documents textuels, chaque mot est potentiellement un index du document qui le contient.
- *Indexation semi-automatique* : c'est une combinaison des deux méthodes précédentes : un premier processus automatique permet d'extraire les termes du document. Cependant, le choix final des descripteurs est laissé au spécialiste du domaine, qui utilise un vocabulaire contrôlé sous forme de thésaurus ou de base terminologique³.

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

- simplification du texte : cette étape consiste à supprimer dans un premier temps des mots de fortes fréquences, généralement en référence à une liste de mots appelée anti-lexiques ou plus fréquemment *stoplist*.
- regroupement des formes morphologiquement liées par la racinisation : cette étape consiste à effacer les terminaisons (flexions de nombre (un nombre est un concept caractérisant une unité, une collection d'unités ou une fraction d'unité.), genre, conjugaison, déclinaison) afin de retrouver les racines des mots. Cette opération est appelée *stemming*. Ce procédé permet de relever les fréquences en cumulant les nombres d'occurrence des variations des mêmes mots.
- pondération des termes : cette étape est généralement basée sur des formules de pondération qui affecte à chaque terme un degré d'importance (une valeur de discrimination) 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), définies dans ce qui suit :

1. 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 que plus un terme est fréquent dans ce document, plus il

³Un thésaurus est une liste organisée de descripteurs (mots-clés) obéissant à des règles terminologiques propres et reliés entre eux par des relations sémantiques.

est important dans la description de celui-ci. Soit le document d_j et le terme t_i , alors la fréquence TF_{ij} du terme dans le document est souvent utilisée directement ou exprimée selon l'une des déclinaisons suivantes [126] :

$$TF_{ij} = 1 + \log(\#td_{ij}), TF_{ij} = \frac{\#td_{ij}}{\sum_k td_{kj}} \quad (1.1)$$

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

2. 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 :

$$IDF_i = \log \frac{|N|}{n}, IDF_i = \log \frac{|N - n|}{n} \quad (1.2)$$

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 :

$$TF * IDF = \log(1 + TF) * \log \frac{|N|}{n} \quad (1.3)$$

1.2.2.2 Interrogation

L'interrogation du système implique un processus d'interaction de l'utilisateur avec le SRI illustré dans la figure 1.1. Cette interaction comprend : (1) la formulation d'une requête par l'utilisateur traduisant son besoin en information, (2) la représentation de la requête sous forme interne selon le langage d'indexation défini, (3) la correspondance entre la requête et les documents par exploitation de l'index et la présentation des résultats.

Plus précisément, l'interrogation implique le scénario suivant : l'utilisateur exprime son besoin en information sous la forme d'une requête. Le système interprète la requête et crée son index qui sera compatible avec le modèle d'index des documents. Ensuite le système évalue la pertinence des documents par rapport à cette requête en utilisant une fonction de correspondance. Cette fonction exploite l'index généré dans la phase d'indexation dans le but de calculer un score de similarité (en anglais *Relevance Status Value*), notée $RSV(Q, D)$, entre

la requête indexée Q et les descripteurs du document D . Différents modèles de RI ont été proposés dans la littérature et tentent de formaliser la pertinence en partant des modèles naïfs basés sur l'appariement exact vers des modèles plus élaborés basés sur l'appariement rapproché.

Le résultat est une liste de documents (pondérés ou non) généralement triée par ordre de valeur de correspondance décroissante, c'est-à-dire du plus pertinent au moins pertinent, et présenté à l'utilisateur. Celui-ci apporte son jugement sur les documents renvoyés par le système selon des critères liés à son besoin en information et au contexte dans lequel la recherche est effectuée.

1.2.3 Aperçu des modèles de RI

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. Il existe un grand nombre de modèles de RI textuelle développés dans la littérature. Ces modèles ont en commun le vocabulaire d'indexation basé sur le formalisme mots clés et diffèrent principalement par le modèle d'appariement requête-document. Le vocabulaire d'indexation $V = \{t_i\}, i \in \{1, \dots, n\}$ est constitué de n mots ou racines de mots qui apparaissent dans les documents. Selon [18], un modèle de RI est défini par un quadruplet $(D, Q, F, R(q,d))$: où

- D est l'ensemble de documents
- Q est l'ensemble de requêtes
- F est le schéma du modèle théorique de représentation des documents et requêtes
- $R(q,d)$ est la fonction de pertinence du document d à la requête q

Nous présentons dans la suite les principaux modèles de RI : le modèle booléen, le modèle vectoriel et le modèle probabiliste.

1.2.3.1 Modèle booléen

Le modèle booléen [154] est basé sur la théorie des ensembles. Dans ce modèle, les documents et les requêtes sont représentés par des ensembles de mots clés. Chaque document est représenté par une conjonction logique des termes non pondérés qui constitue l'index du document. Un exemple de représentation d'un document est comme suit : $d = t_1 \wedge t_2 \wedge t_3 \dots \wedge t_n$.

Une requête est une expression booléenne dont les termes sont reliés par des opérateurs logiques (OR, AND, NOT) permettant d'effectuer des opérations d'union, d'intersection et de différence entre les ensembles de résultats associés à chaque terme. Un exemple de représentation d'une requête est comme suit :

$q = (t_1 \wedge t_2) \vee (t_3 \wedge t_4)$. La fonction de correspondance est basée sur l'hypothèse de présence/absence des termes de la requête dans le document et vérifie si l'index de chaque document d_j implique l'expression logique de la requête q . Le résultat de cette fonction est donc binaire décrite comme suit : $RSV(q, d) = \{1, 0\}$.

1.2.3.2 Modèle vectoriel

Dans ces modèles [155], la pertinence d'un document vis-à-vis d'une requête est définie par des mesures de distance dans un espace vectoriel. Le modèle vectoriel [155] représente les documents et les requêtes par des vecteurs d'un espace à n dimensions, les dimensions étant constituées par les termes du vocabulaire d'indexation. L'index d'un document d_j est le vecteur $\vec{d}_j = (w_{1,j}, w_{2,j}, w_{3,j}, \dots, w_{n,j})$, où $w_{k,j} \in [0, 1]$ dénote le poids du terme t_k dans le document d_j . Une requête est également représentée par un vecteur $\vec{q} = (w_{1,q}, w_{2,q}, w_{3,q}, \dots, w_{n,q})$, où $w_{k,q}$ est le poids du terme t_k dans la requête q . La fonction de correspondance mesure la similarité entre le vecteur requête et les vecteurs documents. Une mesure classique utilisée dans le modèle vectoriel est le cosinus de l'angle formé par les deux vecteurs :

$$RSV(q, d_j) = \cos(\vec{q}, \vec{d}_j) \quad (1.4)$$

Plus deux vecteurs sont similaires, plus l'angle formé est petit, et plus le cosinus de cet angle est grand. A l'inverse du modèle booléen, la fonction de correspondance évalue une correspondance partielle entre un document et une requête, ce qui permet de retrouver des documents qui ne satisfont la requête qu'approximativement. Les résultats peuvent donc être ordonnés par ordre de pertinence décroissante.

1.2.3.3 Modèle probabiliste

Ce modèle est fondé sur le calcul de la probabilité de pertinence d'un document pour une requête [146] [157] [127]. Le principe de base consiste à retrouver des documents qui ont en même temps une forte probabilité d'être pertinents, et une faible probabilité d'être non pertinents. Etant donné une requête utilisateur Q et un document D , il s'agit de calculer la probabilité de pertinence du document pour cette requête. Deux possibilités se présentent : R , D est pertinent pour q et \bar{R} , D n'est pas pertinent pour q . Les documents et les requêtes sont représentés par des vecteurs booléens dans un espace à n dimensions. Un exemple de représentation d'un document d_j et une requête q est le suivant : $d_j = (w_{1,j}, w_{2,j}, w_{3,j}, \dots, w_{n,j})$, $q = (w_{1,q}, w_{2,q}, w_{3,q}, \dots, w_{n,q})$. Avec $w_{k,j} \in [0, 1]$ et $w_{k,q} \in [0, 1]$. La valeur de $w_{k,j}$ (resp. $w_{k,q}$) désigne si le terme t_k apparaît ou non dans le document d_j (resp. q). Le modèle probabiliste évalue la pertinence du document d_j pour la requête q . Un document est sélectionné si la probabilité

que le document d soit pertinent, notée $p(R/D)$, est supérieure à la probabilité que d soit non pertinent pour q , notée $p(\bar{R}/D)$ où R est l'événement de pertinence et \bar{R} est l'événement de non pertinence. Le score d'appariement entre le document D et la requête Q , noté $RSV(Q,D)$ est donné par :

$$RSV(Q, D) = \frac{P(R/D)}{P(\bar{R}/D)}$$

Ces probabilités sont estimées par les probabilités conditionnelles selon qu'un terme de la requête est présent, dans un document pertinent ou dans un document non pertinent. Cette mesure de similarité requête document peut se calculer par différentes formules.

Ce modèle a donné lieu à de nombreuses extensions. Il est à l'origine du système OKAPI. Le modèle Okapi BM25 a été développé par (Robertson et al., 1994) dans lequel le calcul des poids d'un terme dans un document intègre des aspects relatifs à la fréquence locale des termes, leur rareté et la longueur des documents :

$$w_{td} = \log\left(\frac{N - df + 0.5}{df + 0.5}\right) \times \frac{(K_1 + 1) * TF}{K_1 \times ((1 - b) + b \times \frac{dl}{avgdl}) + tf} \quad (1.5)$$

où $dl = moy_{di \in D}(dli)$ et k_1, b sont des paramètres qui dépendent de la collection ainsi que du type des requêtes.

1.3 De la RI classique à la RI adaptative

La recherche d'information classique se base principalement sur le calcul de la pertinence du document selon des critères de sélection par le contenu et de la disponibilité de l'information ou alors elle peut également exploiter la structure des liens entre les documents afin de retourner une liste de résultats en réponse à une requête utilisateur. 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 de recherche. Par exemple, la requête "gouvernement" est une requête locale qui doit renvoyer des résultats concernant le gouvernement du pays dans lequel est localisé l'utilisateur. Par exemple, cette requête doit renvoyer en *France* le site du gouvernement français. La requête "Java magazines" est une requête ambiguë du fait qu'elle est liée à différentes thématiques de recherche, telles que des le langage de programmation *Java*, l'île de *Java* ou une magazine ayant le nom *Java*. Pour ce type de requêtes, le système doit renvoyer des résultats dépendant des centres d'intérêts de l'utilisateur lors de la recherche.

L'accès à une information pertinente adaptée à des critères spécifiques de l'utilisateur et au contexte de recherche fait l'objet de la RI adaptative. Nous abordons dans la suite de cette section les facteurs d'émergence de la RI adaptative,

les techniques développées en RI adaptative ainsi que leurs limitations de couvrir la notion large du contexte.

1.3.1 Facteurs d'émergence de la RI adaptative

Les facteurs d'émergence de la RI adaptative sont principalement liés à la prolifération des ressources d'information hétérogènes, la diversité et l'ambiguïté des besoins en information des utilisateurs ainsi que l'apparition de la RI mobile basée sur l'adaptation du processus de RI au contexte géographique de l'utilisateur.

A. Volume de l'information

L'accès à une information pertinente dans un environnement de recherche qualifié par la prolifération des ressources hétérogènes (données structurées, documents textuels, composants logiciels, images) est un défi réel pour la RI classique.

Le volume de l'information disponible sur le *Web* ainsi que le nombre d'utilisateurs sont toujours en croissance vertigineuse. En 2000, la taille du *Web* est estimée à 7 million sites uniques, à 50% d'augmentation comparativement à l'an 1999. Une étude en 2005 [71] a estimé la taille du web à 11.5 billion pages. Plus récemment, selon une étude statistique couvrant plusieurs moteurs de recherche, la taille du Web est estimée au minimum entre 20 et 40 billion de pages web sur l'année 2009⁴. Notons aussi que le nombre d'utilisateurs est estimé en 2007 à 1 billion. En outre, les utilisateurs n'utilisent habituellement que quelques mots (environ 3 d'après [142], [88]) pour décrire, le document recherché. Face au volume de l'information, la problématique majeure des SRI classiques est qu'ils proposent en réponse une liste massive de documents, ayant tous des estimations de pertinence comparables indépendamment du contexte de recherche de l'utilisateur. En effet, les requêtes exprimées par un *sac de mots clés* souffrent d'une méthode d'évaluation naïve basée sur l'appariement requête-document. La sélection de l'information consiste à considérer que tout document contenant les mots de la requête dans n'importe quel ordre et à n'importe quel endroit de son contenu est potentiellement pertinent. Par la suite, plus le volume des ressources disponibles est important, plus la liste des résultats retournés par le système est importante. Par conséquent l'utilisateur se trouve face à une surcharge informationnelle qui le désoriente, et dans laquelle il prend la charge de distinguer ce qui est pertinent de ce qui ne l'est pas.

B. Diversité et expressivité des besoins en information

Les SRI classiques sont conçus pour servir un public le plus large possible et ne

⁴<http://www.worldwidewebsize.com/>

considèrent que la requête dans la représentation du besoin en information de l'utilisateur. La problématique de l'expressivité du besoin en information est d'autant plus accentués que les requêtes utilisateurs sont courtes ou ambiguës. La longueur moyenne des requêtes des utilisateurs sur le *Web* est estimée à quelques mots (environ 3 mots d'après [142], [88]). Dans ce cas, les SRI renvoient à une même requête soumise par des utilisateurs différents, la même liste de résultats qui correspond à des divers centres d'intérêts ayant des estimations de pertinence comparables. Les études [36] ont montré que l'approche généraliste des outils disponibles en RI sont à l'origine des problématiques évoqués des SRI classiques et leur incapacité à discriminer les utilisateurs en fonction de leurs centres d'intérêts ou de leurs préférences de recherche.

C. L'apparition de la RI mobile

L'émergence de nouveaux environnements de recherche tels que la RI mobile, les agendas personnels, la téléphonie mobile, etc. font du contexte géographique de l'utilisateur une source d'information importante permettant d'améliorer la précision de la recherche [194, 70]. En effet la RI mobile impose des contraintes physiques généralement liées à la localisation géographique de l'utilisateur, à l'instant de soumission de la requête et aux dispositifs dont il dispose. Ces préférences ne sont pas prises en compte par les SRI classiques que dans la mesure où le langage de requête permet de les exprimer et dans la mesure où l'utilisateur les formule systématiquement dans ses requêtes.

1.3.2 La RI adaptative

Les premières approches visant à adapter le processus de RI au contexte d'utilisation du système s'inscrivent dans le cadre de la RI adaptative [147] [53], [202], [50], [106]. Le but fondamental de ces approches consiste à exploiter des informations additionnelles, autres que la requête, extraites des interactions de l'utilisateur avec le système, dans le but d'améliorer la recherche. Les techniques développées en RI adaptative se focalisent principalement sur l'assistance à la formulation des requêtes des utilisateurs ou alors l'assistance à la navigation. Nous citons les techniques de reformulation de requête par réinjection de pertinence, des techniques de désambiguïsation du sens des mots de la requête ou des techniques de regroupement thématique des résultats de recherche [202].

1.3.2.1 Reformulation de requêtes

Le but des techniques de reformulation de requête par réinjection de pertinence [147], [58], [175] est de générer une nouvelle requête mieux adaptée au besoin en information de l'utilisateur. La reformulation de requête consiste

à ajouter de nouveaux termes à une requête initiale ou alors repondérer les poids de ses termes dans le but de cibler la recherche vers les documents pertinents. La génération de la nouvelle requête s'effectue soit par reformulation automatique ou par reformulation interactive.

La reformulation automatique de requête consiste à générer une nouvelle requête sans rétroaction explicite de l'utilisateur en exploitant les premiers documents retournés par le système en réponse à la requête initiale (*blind feedback*). La reformulation interactive des requêtes consistent à générer une nouvelle requête en exploitant un sous-ensemble de documents jugés pertinents par l'utilisateur et retournés en réponse à la requête après chaque itération feedback.

Différents techniques de reformulation de requêtes ont été introduites dans différents modèles, nous citons le modèle vectoriel et le modèle probabiliste.

- Reformulation de requêtes dans le modèle vectoriel :

Selon le modèle vectoriel, la reformulation de requêtes est souvent basée sur l'utilisation de l'algorithme de rocchio [147] où la nouvelle requête est générée en tenant en compte la distribution des termes dans les documents pertinents et non pertinents selon la formule suivante :

$$Q_n = \alpha Q_i + \frac{\beta}{n_p} \sum_{n_p} D_p - \frac{\gamma}{n_{np}} \sum_{n_{np}} D_{np} \quad (1.6)$$

Où :

Q_n : vecteur de la nouvelle requête,

Q_i : vecteur de la requête initiale,

D_p (resp. D_{np}) : vecteur d'un document pertinent (resp. non pertinent),

n_p (resp. n_{np}) : nombre de documents jugés pertinents (resp. non pertinents),

α, β et γ : des paramètres constants, $\alpha + \beta + \gamma = 1$.

- reformulation de requêtes dans le modèle probabiliste :

Selon le modèle probabiliste, la reformulation de requêtes est proposé dans [146] où la génération d'une nouvelle requête est fonction de la distribution des termes dans les documents pertinents et non pertinents donnée selon la formule suivante :

$$w_i = \frac{\frac{r_i}{R-r_i}}{\frac{n_i-r_i}{(N-n_i)-(R-r_i)}} \quad (1.7)$$

Où :

w_i : Poids du terme t_i dans la requête,

r_i : Nombre de documents pertinents contenant le terme t_i ,

R : Nombre de documents pertinents pour la requête,

n_i : Nombre de documents contenant le terme t_i ,

N : Nombre de documents dans la collection.

1.3.2.2 Désambiguïsation du sens des mots de la requête

Ces techniques consistent à aider l'utilisateur d'exprimer mieux son besoin en information et orienter la recherche vers les documents portant l'intention de recherche de l'utilisateur. Elles permettent à l'utilisateur de saisir le vrai sens évoqué par les termes de la requête et l'exploiter dans des techniques d'extension des langages de requêtes. La plupart de ces techniques se basent sur l'exploitation des interfaces de clarification interactives à base d'ontologie [165]. La figure 1.2 montre un exemple d'une requête ambiguë (*jaguar*) soumise sur *Google* et les catégories sémantiques issues de la hiérarchie de catégories sémantiques de *Google*⁵ en réponse à cette requête.

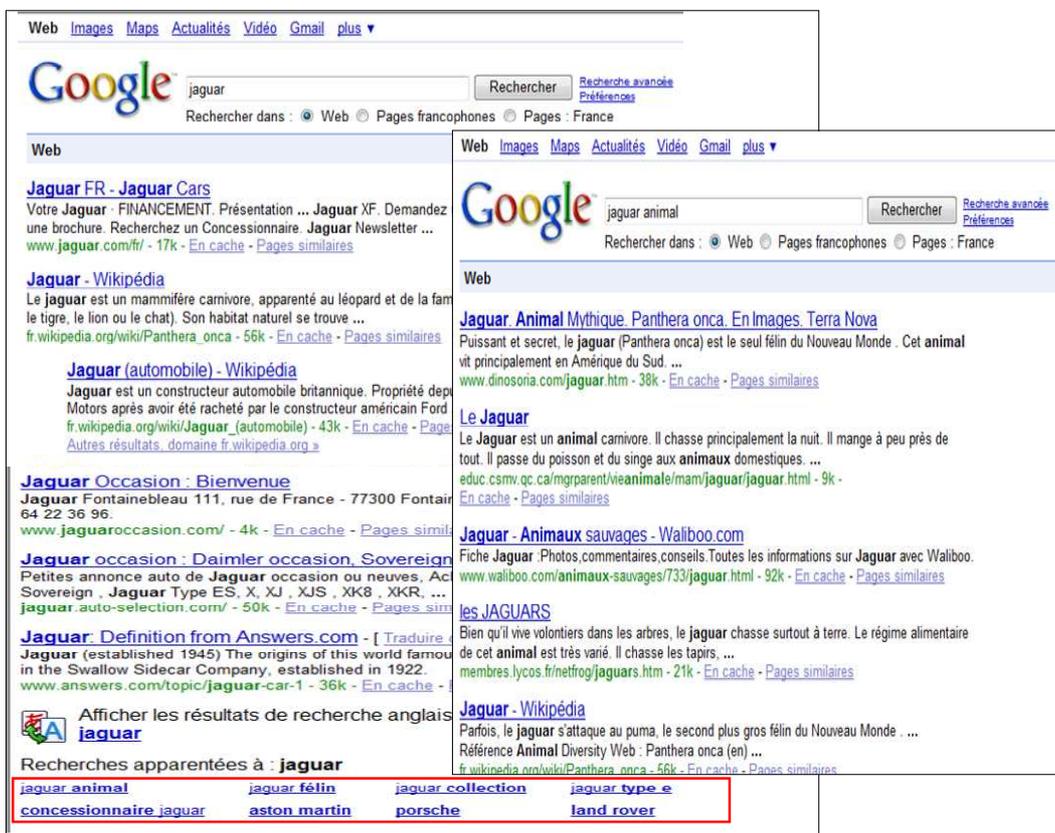


FIG. 1.2 – Désambiguïsation du sens des mots de la requête *Java* sur *Google*

L'adaptation du processus de recherche à des critères spécifiques de l'utilisateur par désambiguïsation du besoin en information derrière la requête est adopté dans plusieurs approches dans le domaine de la RI adaptative [54, 148, 118]. Ces approches se basent sur la définition des paramètres mesurables à partir de la requête ou à partir du profil des premiers documents

⁵<http://directory.google.com/>

retournés par le requête. Dans ce sens, un nouveau type de réinjection de pertinence qui est le profil de requête est exploité afin de détecter des besoins en information divers et des critères de qualité de l'information derrière la requête, tels que le thème de recherche et l'information récente.

L'étude dans [148] comporte la génération des profils temporels et thématiques des requêtes où le profil temporel de la requête est défini par la distribution des N premiers documents retournés pour la requête en fonction de leur date de création. Cette approche permet à l'utilisateur d'affiner sa requête en choisissant d'une part le sujet d'intérêt correspondant à la requête et d'autre part de choisir ses dates préférées identifiées à partir du profil temporel de la requête afin de générer une requête ciblant les résultats estampillés selon ces dates.

1.3.2.3 Regroupement thématique des résultats de recherche

Face à la croissance du web et les difficultés rencontrées par les moteurs de recherche classiques pour satisfaire les besoins en information de l'utilisateur, les techniques de clustering/regroupement des résultats de recherche dans des catégories (Grouper [202], Vivisimo⁶, Kartoo⁷, Exalead⁸, etc.) ou des approches de répertorisation du web dans des taxonomies de concepts (ODP⁹, Yahoo concept hierarchy, Google directory) ont été développées pour une accessibilité et navigation plus simple.

Les techniques de clustering sont basées sur le regroupement thématique (*clustering*) des résultats de recherche dans des catégories ou clusters à la place d'une liste de résultats paginés. Ces techniques sont basées sur le fait qu'un document qui est pertinent à une requête, a probablement une similarité avec d'autres documents qui sont peut-être aussi pertinents. Ce regroupement permet de mettre les documents similaires ensemble et avoir une idée assez générale et globale des résultats retournés et ensuite une accessibilité et une navigation plus simple. Ainsi lors d'une recherche, les utilisateurs disposeraient d'une classification virtuelle (navigation hiérarchique) des documents retournés.

C'est le cas de Ask.com utilisant Vivisimo¹⁰, Exalead, Grouper [202] utilise un algorithme incrémental (STC) original et tient compte de l'ordre des termes dans les documents pour construire un index inversé sur les phrases communes à plusieurs documents pour construire les classes.

Dans le même sens, plusieurs ontologies de domaines spécifiques ont été conçues et ce dans le but de faire asseoir une recherche conceptuelle permettant de simplifier la navigation à travers les catégories sémantiques de la hiérarchie utilisée. La conception de ces ontologies consistent en des techniques de répertorisation

⁶<http://clusty.com/>

⁷<http://www.kartoo.com/index.php3>

⁸<http://www.exalead.com/search/>

⁹<http://www.dmoz.org>

¹⁰<http://clusty.com/>

du web permettant de classifier manuellement les pages web dans des taxonomies des concepts hiérarchiques. Il existe plusieurs répertoires du Web édités par des êtres humains, nous citons l'Open Directory Project, Yahoo concept hierarchy ¹¹, Google directory ¹² et autres.

1.3.3 Limitations de la RI adaptative

Malgré le gain de performance apporté par les techniques développées en RI adaptative, elles présentent toutefois des limitations. Les limites sont principalement liées à la représentation limitée du contexte de l'utilisateur, le mode d'interaction explicite avec le système et à la dépendance de la familiarité de l'utilisateur avec le thème de recherche et le nombre d'itérations de recherche sur la performance de la RI adaptative.

- **Le contexte est peu connu** : les techniques de reformulation de requêtes en RI adaptative consistent à aider l'utilisateur à sélectionner les termes de la requête via des interfaces de clarification du besoin en information, ou à générer des requêtes ciblées par expansion automatique ou semi-automatique de la requête initiale. De ce fait, le contexte de recherche est limité au besoin en information véhiculé par les termes de la requête soumise à chaque itération de recherche. Le contexte dans lequel une recherche est effectuée (c'est à dire la tâche en cours, la situation géographique, etc.) est rarement utilisé pour interpréter la requête et la situer par rapport aux buts et connaissances préalables de chaque utilisateur ainsi que d'autres facteurs contextuels ayant un impact potentiel sur la performance de recherche [83].
- **Limitations de l'interaction explicite** : Les techniques de reformulation de la requête par réinjection de pertinence ou de clarification du besoin en information nécessitent une rétroaction explicite de la part de l'utilisateur. Par ailleurs, plusieurs études montrent que la majorité des utilisateurs se limitent à fournir la requête initiale et préfère des mécanismes d'amélioration qui fonctionnent sans demande explicite d'information complémentaire [88].
- **Impact de la familiarité de l'utilisateur avec le sujet de recherche** : la performance des techniques de reformulation de requête ou de clarification automatique du besoin en information derrière la requête est relativement dépendante de la familiarité de l'utilisateur avec le sujet de recherche d'une part et du nombre des itérations de recherche ayant un impact sur l'efficacité de la réinjection de pertinence d'autre part. Différentes études [79] ont montré que la familiarité de l'utilisateur avec le sujet de recherche et son niveau d'expertise [151, 198] influent

¹¹<http://dir.yahoo.com/>

¹²<http://directory.google.com/>

sur les performances de recherche. La sélection des documents pertinents utilisés comme sources de réinjection de pertinence dans la reformulation des requêtes diffère d'un utilisateur à un autre en fonction de sa familiarité avec le sujet de recherche. En outre, lorsque la reformulation de requête est guidée par un contrôle de pertinence, le système devient de plus en plus opérationnel avec l'augmentation du nombre des itérations de recherche. Ceci est accompagné par un risque de démotivation de l'utilisateur et cela d'autant plus que les qualités de rappel et précision risquent d'être faibles dans un premier temps.

1.4 Conclusion

Nous avons présenté au cours de ce chapitre les concepts de base de la RI classique ainsi que l'évolution de la RI classique à la RI adaptative. Nous avons cité les problématiques de la RI classique en présence du contexte, les facteurs d'émergence de la RI adaptative ainsi que les techniques développées en RI adaptative, notamment la reformulation des requêtes, la désambiguïsation du sens des mots des requêtes et le regroupement thématique des résultats. Compte tenu des limitations de la RI adaptative, les travaux se sont orientés vers la RI contextuelle dont le but est de répondre mieux aux besoins en informations de l'utilisateur selon une approche de RI non généraliste mais plutôt orientée-contexte. Le chapitre suivant présente l'essence des motivations des travaux vers la RI contextuelle.

Chapitre 2

Émergence de la RI contextuelle

2.1 Introduction

Compte tenu des limitations de la RI adaptative, les approches en RI se sont orientées vers une nouvelle génération des moteurs de recherche basés sur l'accès contextuel à l'information. L'objectif de la RI contextuelle est de mieux répondre aux besoins en information de l'utilisateur tout en intégrant le contexte de recherche dans la chaîne d'accès à l'information. Selon [16], la RI contextuelle consiste à exploiter le contexte de l'utilisateur ainsi que des connaissances liées à la requête dans le but de répondre mieux aux besoins en information de l'utilisateur. Cette définition est donnée comme suit : "*Combine search technologies and knowledge about query and user context into a single framework in order to provide the most appropriate answer for a user's information needs*".

Parmi les éléments contextuels les plus importants traités dans la littérature, nous citons les centres d'intérêts de l'utilisateur connu sous le contexte cognitif, sa tâche de recherche, ses préférences de recherche liées à la qualité de l'information retournée par le système (tel que la fraîcheur de l'information, la crédibilité de la source d'information, etc.), des préférences liées à la localisation géographique, au volume et au mode de présentation des résultats. La plupart des approches dans le domaine modélisent les centres d'intérêts de l'utilisateur dans une structure informationnelle, appelée profil, le contexte autour duquel s'articule la RI personnalisée. En RI contextuelle, la notion de contexte recouvre des dimensions plus larges englobant en plus du profil de l'utilisateur, le contexte de la requête, les préférences de recherche, le contexte d'interactions avec le système ainsi que le contexte temporel et géographique de recherche.

Ce chapitre traite de l'émergence de la RI contextuelle. Nous abordons les différentes définitions du contexte en RI, le concept du contexte multidimensionnel ainsi que la notion de contexte dans nos travaux de thèse. Nous nous intéresserons dans la section 2.3 à l'accès contextuel à l'information guidé par le contexte de la tâche de recherche, où nous présentons les différentes techniques d'identification d'exploitation de la tâche dans différents modèles de RI. Dans la section 2.4, nous mettons l'accent sur l'accès contextuel à l'information guidé par le profil utilisateur où nous abordons le cadre et les motivations de l'émergence de ce domaine, la notion de profil utilisateur, les différents modèles de représentation du profil utilisateur, les différentes techniques de construction et d'évolution du profil au cours des sessions de recherche ainsi que les différents modèles d'accès personnalisé à l'information. La section 2.5 conclut le chapitre.

2.2 La notion du contexte : définitions et taxonomies du contexte multidimensionnel

La RI se fait selon un processus d'interaction homme-machine (H-M) où plusieurs facteurs interviennent et influencent la perception de pertinence de l'information du côté utilisateur. Celui-ci apporte son jugement sur les documents renvoyés par le système selon des critères liés au contexte dans lequel la recherche est effectuée.

Nous présentons dans cette section ce qu'est le contexte en RI, plus particulièrement, nous abordons les directions de recherche en RI contextuelle en citant les différents éléments contextuels largement exploités dans la majorité des systèmes et le concept du contexte multidimensionnel ainsi que les différentes taxonomies proposées.

2.2.1 Qu'est ce que le contexte en RI ?

Plusieurs définitions du contexte proposées dans la littérature en RI contextuelle, diffèrent essentiellement par ses éléments constitutifs. Les facteurs contextuels les plus exploités sont les centres d'intérêts de l'utilisateur, la tâche de recherche en cours, les préférences de recherche liées à la qualité de l'information (fraîcheur de l'information, genre du document, crédibilité de la source de l'information, etc.), le temps de soumission de la requête, la localisation géographique de l'utilisateur, l'environnement de recherche, etc. Outre ces facteurs, on distingue sous l'angle de la dimension temporelle, le

contexte à court terme et le contexte à long terme. Le contexte à court terme inclut des éléments contextuels qui changent d'une recherche à une autre, tels que la localisation géographique de l'utilisateur, la nature de la tâche de recherche ou le type de besoin, etc. Le contexte à long terme inclut des éléments contextuels et des préférences de recherche persistants et évolutifs en même temps, tels que les centres d'intérêts et les préférences liées à la qualité de l'information.

Les premières approches en RI contextuelle se focalisent sur le contexte de l'utilisateur représenté par son profil. Ce contexte inclut les centres d'intérêts, buts et connaissances de l'utilisateur dégagés au cours de ses sessions de recherche. Il a été démontré [138] que les centres d'intérêts de l'utilisateur représentent l'élément contextuel le plus important qui permet de résoudre l'ambiguïté de recherche dans un système de recherche d'information textuelle. Les premiers systèmes développés sur la base de l'exploitation du profil utilisateur sont les systèmes de recommandation basés sur le filtrage collaboratif. Ces systèmes tels que GroupLens [107] exploitent le profil collaboratif lié à un groupe d'utilisateurs partageant des centres d'intérêts communs et persistants et retourne à l'utilisateur de l'information répondant aux critères du profil du groupe auquel il appartient. D'autres systèmes de recommandation personnels tels que LETIZIA [122] et JITTIR [143] servent comme assistants personnels pour le parcours du Web en exploitant un profil propre à chaque utilisateur. Ces systèmes sont capables de proposer pro-activement de l'information pertinente sans demande explicite par l'utilisateur. Nous citons également les systèmes de RI personnalisés qui se basent sur des techniques d'apprentissage du profil de l'utilisateur [123, 164, 67, 124] au cours de ses sessions de recherche selon un modèle de représentation riche dans le but d'améliorer la précision de recherche.

L'exploitation du contexte de la tâche de recherche fait l'objet d'une variété des approches en RI contextuelle [32, 150, 114, 96, 95] où le type de besoin derrière la requête est défini comme étant informationnel lié à la recherche du contenu informationnel de documents, navigationnel lié à la recherche des sites d'accueil des personnes, organisations ou autres ou transactionnel lié à la recherche des services en ligne. Ces systèmes visent à mettre en place des modèles de recherche guidés par la tâche de recherche en exploitant les sources d'évidences les plus pertinentes à chaque type de besoin.

Peu de travaux en RI contextuelle ont exploré la dimension du contexte liée aux préférences de qualité de l'information liée à la fraîcheur de l'information [118, 148], ou à la crédibilité de la source de l'information [74], sa cohérence, sa complétude, le genre de documents renvoyés [134], etc. Plus récemment, diverses études ont été menées dans le but de développer des moteurs de recherche sensitifs au contexte géographique de l'utilisateur [194], [70, 134]. La localisation géographique de l'utilisateur est une préférence de recherche qui peut prendre place même en l'absence d'environnements de recherche mobile, où l'utilisa-

teur préfère un ordre des résultats de recherche tenant compte d'une référence géographique. Cette préférence s'impose nécessairement dans un contexte de recherche où le but est d'accéder à des services tels que des achats en ligne, etc.

Il est difficile de trouver un SRI contextuel qui gère toutes les dimensions contextuelles en même temps. Les dimensions à considérer lors de la prise en compte du contexte peuvent être plus ou moins pertinentes selon la recherche effectuée.

2.2.2 Le contexte : un concept multidimensionnel

Les premières définitions du contexte en RI, rapportées dans la littérature, portent principalement sur des facteurs ou des caractéristiques liés à différentes entités impliquées dans le processus de recherche [85]. Ces facteurs traduisent ce que l'on appelle la *situation/contexte* de recherche. On distingue trois principales entités intervenant dans le processus de recherche : l'utilisateur, le système et l'environnement de recherche. Chacune de ces entités possède des caractéristiques ayant un impact sur la perception de pertinence de l'information chez l'utilisateur. Concernant l'utilisateur, on cite l'arrière plan (connaissances), ses buts et ses intentions de recherche, etc. Au niveau du système, on cite des caractéristiques liées au processus de recherche tels que le temps de réponse, le coût, etc. L'environnement de recherche, quant à lui possède des caractéristiques liées à sa nature qui peuvent être liées à des critères sociaux, organisationnels, etc.

Les travaux de Saracevic [159] et Ingerwersen [84] sont les premiers qui ont introduit la notion du contexte sans distinction avec la notion de situation de recherche. Le contexte y est défini selon un modèle cognitif par lequel on identifie des structures ou espaces cognitifs qui décrivent les intentions et les perceptions de l'utilisateur. Ces variables sont représentées par l'espace cognitif de l'utilisateur, l'environnement social ou organisationnel, les intentions et les buts de l'utilisateur ainsi que le système lui-même. Nous présentons dans la suite de cette section plusieurs modèles de définitions du contexte [66],[47],[85],[177] basées sur le concept du contexte multidimensionnel, faisant intervenir des dimensions cognitives dépendantes de l'utilisateur, spatio-temporelles dépendantes de l'environnement de recherche, et qualitatives dépendantes du support de l'information.

2.2.2.1 Taxonomie de Fuhr 2000

Un contexte multidimensionnel est défini dans [66] et ajoute aux caractéristiques de l'environnement de recherche déjà citées dans les définitions précédentes du

contexte, de nouvelles caractéristiques liées à l’aspect temporel du besoin en information d’une part et au type de recherche demandé d’autre part. Les trois principales dimensions retenues pour le contexte (Figure 2.1) sont le caractère social, l’application et le temps.

- la dimension sociale définit soit l’appartenance possible de l’utilisateur à un groupe ou à une communauté soit la non appartenance ce qui souligne un aspect individuel.
- la dimension *application* définit le but de la tâche accomplie en tant qu’une application workflow, une recherche *ad hoc* ou une résolution de problème.
- la dimension temps permet de définir le contexte temporel du besoin ; trois aspects temporels du besoin peuvent être identifiés : temps passé (*batch*), intention à court terme ou intention à long terme. Le contexte à court terme (*interactif*) est associé à des besoins et à des préférences instantanées de l’utilisateur alors que le contexte à long terme (personnalisation) traduit des besoins et des préférences persistants de l’utilisateur tout au long de diverses sessions de recherche. Une session de recherche peut être associée à une itération de recherche représentée par une requête soumise et ses résultats associés ou alors un ensemble d’itérations de recherche appartenant au même besoin en information.

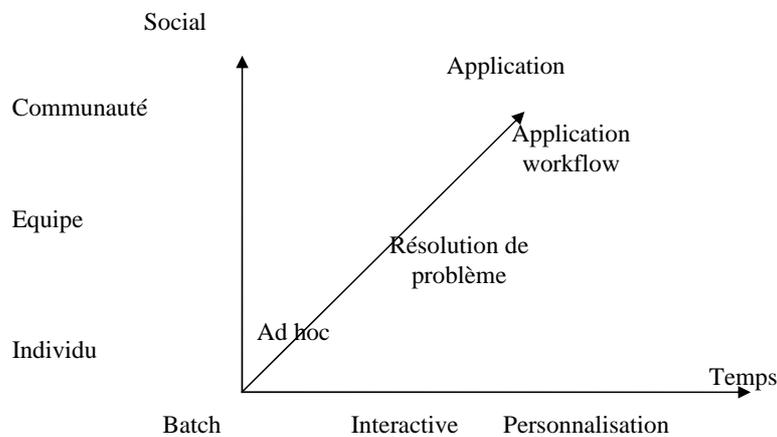


Fig 1 : Dimension d’un contexte [Fuh 2000]

FIG. 2.1 – Dimensions du contexte multidimensionnel de [Fuhr 2000]

2.2.2.2 Taxonomie de Cool 2001

Une tentative de définition plus large des aspects du contexte est apparue dans [47] dans le but de distinguer la notion de ”contexte ”de la notion de situation”. Cette définition fait intervenir la tâche accomplie et les activités courantes de recherche ainsi que la dimension géo-spatiale de recherche dépendant

du lieu et du temps de recherche. Plus particulièrement, plusieurs aspects du contexte sont définis [47] et sont liés à l'environnement cognitif, social et professionnel dans lesquels s'inscrivent des situations liées à des facteurs tels que le lieu, le temps et l'application en cours. Dans une définition plus élaborée du contexte, Cool et Spink [48] définissent des niveaux contextuels considérés les plus significatifs en RI afin de dissocier les entités intervenant dans le processus de recherche :

1. le premier niveau concerne 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,
2. le deuxième niveau concerne la RI liée aux connaissances de l'utilisateur, ses buts et ses intentions de recherche,
3. le troisième niveau concerne l'interaction utilisateur-système et met en évidence l'impact des situations ou de l'environnement sur la rétroaction ou les jugements de pertinence de l'utilisateur,
4. Le dernier niveau concerne le niveau de requête ou le niveau linguistique du contexte ; ce niveau explore la performance du SRI dans l'interprétation des requêtes des utilisateurs et leur habilité à les désambigüiser.

2.2.2.3 Taxonomie de Ingerwersen et al. 2005

Plus récemment, Ingwersen et Järvelin [85] ont développé une infrastructure contextuelle cognitive dans le but d'étudier les dimensions du contexte qui ont un impact sur le processus de RI. L'infrastructure proposée consiste en 9 classes/dimensions des contextes dépendantes l'une de l'autre. On cite :

1. la dimension de la tâche de travail naturelle dans une organisation et concerne les caractéristiques d'intérêts liées à la tâche,
2. la dimension de la tâche de recherche et concerne les caractéristiques spécifiques à la tâche elle-même,
3. la dimension de l'utilisateur qui concerne ses centres d'intérêts et ses connaissances. Elle concerne aussi le type de besoin dans une tâche de recherche ainsi que sa perception vis à vis de cette tâche.
4. les caractéristiques de la collection (genre des documents, etc.),
5. les caractéristiques du système telles que le principe de représentation des documents et des besoins ainsi que les méthodes d'appariement.

L'infrastructure propose aussi des caractéristiques d'interaction qui s'intègrent à la dimension d'accès à l'information. Elle définit également des variables temporelles concernant la durée des interactions (interactions à court terme, à long terme ou à base de session), le mode d'interaction (oral, iconique, pointage), etc.

Même si les auteurs ne convergent pas vers une même définition du contexte

on retrouve toutefois des dimensions descriptives commune telles que l’environnement cognitif, le besoin en information, et l’interaction liée à la recherche d’information.

2.2.2.4 Taxonomie de Tamine et al. 2009

Une taxonomie détaillée des dimensions du contexte est explorée dans [177]. Nous résumons dans la figure 2.2, les cinq dimensions associés.

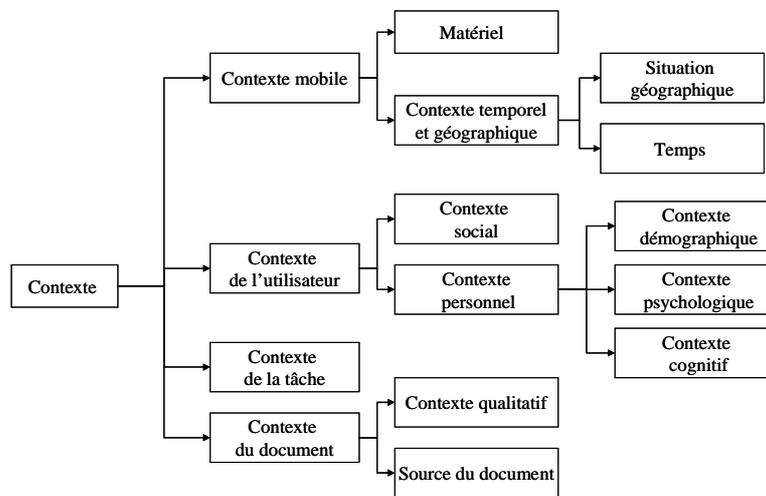


FIG. 2.2 – Le concept du contexte multidimensionnel en RI

A. Contexte mobile

Le contexte mobile regroupe l’environnement matériel de recherche ainsi que la situation géographique de l’utilisateur. L’environnement matériel de recherche est devenu une composante importante du fait de l’apparition de nouveaux environnements de recherche d’une part (tels que la RI mobile, agendas personnels, téléphonie mobile, etc.) et l’hétérogénéité des ressources disponibles sur le Web. D’autre part, ce contexte définit des caractéristiques physiques du processus d’interaction liées aux capacités et aux moyens matériels dont dispose l’utilisateur : le terminal disponible, l’état du réseau auquel l’utilisateur est connecté (bande passante, débit, etc.) ainsi que d’autres facteurs mesurés par des capteurs d’environnement. Ce contexte implique des contraintes physiques ayant un impact potentiel sur la perception de pertinence de l’utilisateur vis-à-vis de l’information retournée.

C. Contexte de l’utilisateur

Le contexte de l’utilisateur est la dimension centrale en RI contextuelle ; elle est largement abordée dans le domaine de la RI personnalisée. Cette dimension

contient deux sous-dimensions liées respectivement au contexte personnel de l'utilisateur et son environnement social.

- **contexte personnel** : ce contexte est associé aux sous-dimensions suivantes :
 - **contexte démographique** : ce contexte désigne des préférences personnelles comme le langage préféré des résultats de recherche (Google personalized, Yahoo) ou le genre du document utilisés dans le but d'adapter les résultats de recherche aux préférences de l'utilisateur [81, 65].
 - **contexte psychologique** : ce contexte désigne des caractéristiques affectives de l'utilisateur tels que *l'anxiété* et la *frustration* qui influencent son comportement de recherche ainsi que ses jugements de pertinence [26, 103].
 - **contexte cognitif** : c'est la sous-dimension la plus abordée dans le domaine, et désigne le niveau d'expertise de l'utilisateur, ses centres d'intérêts qui peuvent être à court ou à long terme. Les centres d'intérêts de l'utilisateur expriment le domaine d'expertise de l'utilisateur qui délimite le périmètre d'exploration de sa recherche. Ces centres d'intérêts servent à réduire virtuellement la collection de documents en ne retenant que ceux qui intéressent l'utilisateur dans sa recherche. Les centres d'intérêts sont souvent modélisés dans une structure informationnelle, appelé profil, investi en RI personnalisée.
- **contexte social** : ce contexte désigne une communauté d'utilisateurs comme des amis, des collègues ou des voisins. Selon la dimension sociale, l'adaptation du processus de RI consiste à retourner de l'information qui répond aux préférences de la communauté des utilisateurs plutôt que de répondre aux préférences d'un seul utilisateur [112, 167]. Le contexte social est souvent exploité dans les systèmes de recommandations [116] basé sur le filtrage collaboratif, utilisé sous le nom du profil collaboratif. Actuellement, plusieurs travaux [98], [97], [166] [201] exploitent le contexte social en RI. Plus particulièrement, le contexte social dans [166] est divisé en un contexte social personnel de l'utilisateur (amis et communautés à laquelle il appartient) et un autre communautaire contenant le rôle de l'utilisateur et son identité dans différents communautés.

D. Tâche de recherche

La tâche de recherche est définie dans la plupart des études [32, 119, 197] par l'ensemble des activités de recherche liées à un même besoin en information de l'utilisateur. Le contexte de la tâche désigne des aspects liés au type du besoin en information de l'utilisateur ou son intention de recherche. Nous distinguons deux aspects contextuels liés aux tâches de recherche : un aspect global défini par le type de besoin informationnel, navigationnel ou transactionnel de la tâche et un aspect spécifique associant un domaine d'intérêt à la tâche. Le contexte de la tâche est introduit du fait que la nature de la tâche influence le comportement de recherche de l'utilisateur [119] ainsi que la performance du système [197, 32].

Concernant le contexte de la tâche défini par le type de besoin en information, des études [32] sur la taxonomie des types de besoins sur le Web ont distingué deux types de besoins qui dépassent la recherche documentaire : la recherche des pages d'accueil des sites Web et la recherche des services en ligne. L'identification du type de recherche réalisée par l'utilisateur comme élément contextuel sert, selon plusieurs études, à améliorer la performance de système par l'exploitation des sources d'évidence autres que le contenu des documents et l'adoption des stratégies de recherche appropriées à chaque type de recherche. Les types des requêtes identifiées sont au nombre de trois :

- la requête informationnelle qui s'inscrit dans le cadre de la recherche documentaire classique,
- la requête navigationnelle dont l'intention est d'accéder à des sites d'accueil,
- la requête transactionnelle dont l'intention est d'accéder à des services en ligne.

La classification de la nature des tâches dans des domaines d'intérêts fait l'objet de différents études menées. Nous donnons comme exemple à ce deuxième type les tâches de recherche suivantes : (1) Academic Research, (2) News and Weather, (3) Shopping and Selling. L'approche proposée dans [119] définit des aspects spécifiques d'une tâche selon des facettes génériques et des facettes communes. Les facettes génériques d'une tâche concernent la source de la tâche, la durée, l'action et le but. Les facettes communes incluent des caractéristiques de la tâches et la perception de l'utilisateur dans la tâche.

E. Contexte du document

Le contexte du document est issu du principe de la poly-représentation développé par Ingwersen [82]. Le principe de la poly-représentation est basé sur l'hypothèse que les documents peuvent être classifiés dans un espace à variable multiples (structure du document, genre du document, style du contenu structure des hyperliens, etc) et que cette classification augmente la probabilité que ces documents soient utilisés dans certaines situations de recherche/ contexte de recherche. Trois sous-dimensions sont identifiés sous le contexte du document :

- La première dimension concerne la représentation du document tels que les éléments structurels, la forme, les citations et des métadonnées [186],
- La deuxième dimension concerne les caractéristiques de la source des données [200], telles que sa crédibilité, sa fiabilité, etc.
- La troisième dimension concerne la qualité de l'information. Vu que la mesure de pertinence de l'information ne dépend pas seulement du contenu des données mais aussi de sa qualité, la notion de pertinence est étendue à une notion plus générale liée à la qualité de l'information retournée. En général, la qualité de l'information concerne sa fraîcheur, sa précision, sa

cohérence, sa complétude, sa sécurité, etc.

2.2.3 Le contexte : définition et exploitation dans le cadre de nos travaux de thèse

Selon la taxonomie définie dans [177] du contexte multidimensionnel, nous considérons dans nos travaux de thèse deux dimensions définies par : (1) *le contexte de la tâche de recherche*, (2) *le contexte cognitif de l'utilisateur*. Ces deux dimensions sont largement étudiées et considérées les plus importantes dans l'amélioration de la précision de la recherche informationnelle. Nous définissons ainsi ces deux dimensions comme suit :

- **le contexte de la tâche de recherche** : ce contexte est défini par le type de besoin derrière la requête comme étant informationnel, navigationnel ou transactionnel. Il est défini dans le but de mettre en place une recherche guidée par le contexte de la tâche derrière la requête.
- **le contexte cognitif de l'utilisateur** : ce contexte traduit les divers centres d'intérêt de l'utilisateur qui peuvent être adressés au cours de ses sessions de recherche informationnelles. Ce contexte est défini dans le but de personnaliser les résultats de recherche, lever l'ambiguïté des termes des requêtes et répondre à des besoins en information récurrents.

Nous présentons dans la suite de ce chapitre une synthèse des travaux inscrits dans deux directions en RI contextuelle :

- l'accès contextuel à l'information guidé par la tâche de recherche où la nature de la tâche est définie comme étant informationnelle navigationnelle ou transactionnelle,
- l'accès contextuel à l'information guidé par le contexte cognitif de l'utilisateur, où le profil utilisateur fait le cœur de personnalisation.

Pour chacune de ces directions, nous présentons les techniques de modélisation du contexte de recherche ainsi que les différents modèles d'exploitation du contexte abordées dans quelques approches proposées en RI contextuelle.

2.3 Accès contextuel à l'information guidé par le contexte de la tâche de recherche

Nous présentons dans cette section quelques techniques développées en RI contextuelle où l'accent est mis sur le contexte de la tâche de recherche sur le Web. Plus précisément, nous abordons les points suivants :

- Le cadre et les motivations de la taxonomie des tâches de recherche selon

- le type de besoin derrière la requête en tant qu'informationnel, navigationnel ou transactionnel,
- les différentes techniques d'identification du contexte de la tâche,
- les différents modèles d'exploitation du contexte de la tâche dans le processus de RI.

2.3.1 Cadre et motivations

Vu la diversité de la nature des tâches de recherche et les types de besoins derrière les requêtes sur le Web, plusieurs études portent sur la taxonomie des requêtes des utilisateurs en trois types : les requêtes informationnelles associées à la recherche des ressources documentaires simple, les requêtes navigationnelles associées à la recherche des sites d'accueil des personnes ou organisation et les requêtes transactionnelles associées à la recherche d'un service négocié via le *Web*. Le but fondamental de l'identification du type de besoin derrière la requête est de mettre en place des stratégies et des modèles de recherche permettant d'exploiter des sources d'évidence les plus appropriées au dépistage de l'information pertinente à chaque type de besoin.

Broder [32] et Rose et Levinson [150] ont distingué trois types des tâches, la tâche navigationnelle, la tâche informationnelle et la tâche transactionnelle. Selon [32], la taxonomie des requêtes a été effectuée selon deux méthodes : la première se base sur l'analyse du fichier log d'un moteur de recherche et la deuxième se base sur les enquêtes des utilisateurs. Les résultats montrent que plus de 30% des requêtes sont transactionnelles et plus de 20% sont navigationnelles et le reste est informationnel. L'identification des ces trois tâches de recherche est à la base de l'évolution des moteurs de recherche vers l'intégration des modèles guidés par la tâche de recherche. Nous décrivons cette évolution en fonction du type de la tâche pris en compte par le modèle de recherche comme suit :

- La tâche de recherche informationnelle est supporté classiquement par la première génération des moteurs de recherche développés autour de 1995-1997. Il s'agit de la RI classique qui utilise le contenu des documents pour retrouver les réponses aux requêtes des utilisateurs. On donne comme exemple de moteur de recherche appartenant à cette génération AltaVista, Excite, WebCrawler, etc.
- La tâche de recherche navigationnelle fait l'objet de la deuxième génération des moteurs de recherche qui supporte des requêtes de contenu ainsi que des requêtes demandant des sites d'accueil. Il s'agit des modèles de recherche basés sur des algorithmes de calcul de pertinence des documents par utilisation des liens, textes d'ancrage et les données de clics. Les moteurs de recherche correspondants à cette deuxième génération sont développés en 1998-1999. *Google* a été le premier moteur de recherche

qui utilise l'analyse des hyperliens pour classer les résultats de recherche. *DirectHit* utilise les données de clics de navigation afin de répondre aux requêtes navigationnelles.

- A ce jour, les moteurs de recherche visent à supporter tout type de tâche y compris la tâche transactionnelle. Dans le but de répondre à ce dernier type de tâche, les modèles de recherche se basent sur l'analyse sémantique de la requête ou la détermination d'un contexte en se basant sur des serveurs de climats (yahoo¹), un serveur des cartes topographiques (Google maps²) ou à l'identification des pages transactionnelles contenant des hyperliens de transactions [95] (réservation des hôtels ou de vols, etc.).

2.3.2 Techniques d'identification du contexte de la tâche

La nature de la tâche est identifiée dans la plupart des études à partir du contexte de la requête. En effet, l'identification du type de la tâche derrière la requête se fait souvent par classification des requêtes selon les trois types soit manuellement ou automatiquement. On traduit le type de la tâche par le type de la requête.

L'identification automatique des types des requêtes des utilisateurs sur la *Web* se fait principalement grâce à l'exploitation des comportements de *clics* pour chaque type de requête [114], des caractéristiques linguistiques des termes de la requête ou des caractéristiques liées à la distribution des termes de la requête dans des hyperliens [96, 95].

Broder [32] et Rose et Levinson [150] adoptent une classification manuelle des requêtes en exploitant les enquêtes des utilisateurs et l'analyse des logs des requêtes. La classification manuelle est faite à l'aide d'une enquête de trois choix qui permettent de différencier la requête de navigation de la requête d'information et se base sur le nombre de votes associés au type de la tâche par les participants. Les résultats montrent que plus que 40% des requêtes correspondent à une tâche de recherche transactionnelle.

L'étude dans [114] repose sur une identification automatique des requêtes selon deux types informationnel et navigationnel sans aucune rétroaction explicite de la part de l'utilisateur et utilise deux caractéristiques servant à la classification qui sont les comportements de clics des utilisateurs et le taux d'utilisation des termes de la requête dans les textes des hyperliens. La distribution des clics désigne le nombre de clics accumulés par les utilisateurs sur un résultat de recherche et le nombre moyen de résultats cliqués pour une requête. La distribution des liens d'ancrage se base sur les pages cibles des liens qui ont un texte d'ancrage similaire aux termes de la requête. Cette étude se termine

¹weather.yahoo.com

²maps.google.com

par une évaluation des caractéristiques déjà citées dans la prédiction du type de la requête.

Dans ce même cadre [96] propose une méthode de classification des requêtes en deux types : navigationnelle et informationnelle. La classification est basée sur les paramètres suivants :

- Distribution des termes dans la requête dans les pages navigationnelles et informationnelles,
- Information mutuelle de chaque terme de la requête,
- Taux d'utilisation des termes de la requête dans les textes d'ancrages,
- Information POS (Part of speech tags)

Cette classification est utilisée pour choisir l'algorithme d'appariement le plus approprié pour chaque type de tâche et étudie l'effet sur la performance du système de recherche.

En outre, [95] étend la méthode de classification des requêtes web proposée dans [96] pour détecter les requêtes transactionnelles. Puisque la transaction négociée via le Web est habituellement mise en application par des hyperliens, 9 types d'hyperliens sont définis pour libeller les textes d'ancrage associés suivant l'action prévue avec l'objet (lire, visiter et télécharger). La classification est effectuée à l'aide du classifieur TiMBL à base d'expressions qui accepte en entrée un vecteur de caractéristiques de la requête et l'associe à une classe en utilisant des exemples de classification stockées au préalable utilisés pour la décision. Le vecteur des caractéristiques de la requête concerne le type de la requête (informationnelle ou navigationnelle selon [96]), un champ indiquant si la requête représente un nom de fichier, le premier terme de la requête, le dernier terme de la requête, un vecteur de scores calculés entre la requête et chacun des types d'hyperliens identifiés. L'exactitude de la méthode de classification de requêtes proposée est évaluée par de diverses expériences et celles-ci montrent une amélioration de 91% de classification des requêtes transactionnelles.

2.3.3 Techniques d'exploitation du contexte de la tâche de recherche dans le processus de RI

L'exploitation du contexte de la tâche de recherche en RI se fait par l'identification des sources d'évidences les plus pertinentes à chaque type de tâche dans le but d'améliorer la précision de recherche.

L'objectif fondamental de la classification des requêtes selon le type de recherche (informationnel, navigationnel et transactionnel) est de retourner des pages Web correspondantes au type de recherche derrière la requête en haut de la liste de résultats retournés. L'amélioration de la précision de recherche des requêtes informationnelles se basent sur le contenu textuel du document [96]. Tandis que la précision de recherche des requêtes navigationnelles et transactionnelles est basé sur l'exploitation des hyperliens et les textes d'ancrage

des pages Web dans lesquels les termes de ces deux types de requêtes sont fréquemment occurrents.

Pour les requêtes navigationnelles, la combinaison des informations de liens de la page web (PageRank) et des informations d'URL tels que la probabilité *a priori* du document de se classer comme une page d'accueil (*UrlPrior*) améliore la précision de la recherche pour la requête navigationnelle [96].

Des approches en RI utilisent la même méthodologie d'ordonnement des résultats de recherche pour la requête navigationnelle que pour la requête transactionnelle [86]. L'approche est basée sur une représentation multiple de la page Web : la première représentation est basée sur le contenu textuel de la page, la deuxième représentation est basée sur tous les textes d'ancrage et les textes encadrés par les balises *TITLE* dans les pages pointant vers la page courante. La troisième représentation est basée sur les textes encadrés par les balises *TITLE*, *H1* et *BIG* des pages pointant vers la page courante. Ensuite, le modèle de recherche s'appuie sur la combinaison pondérée des scores du document obtenus par chaque représentation. Ceci permet de dépister les pages d'accueil et les services en ligne. Les résultats de recherche peuvent être améliorés à condition que le type de recherche soit connu au préalable.

L'approche de classification des requêtes selon le type de recherche développée dans [95] ajoute au score d'appariement original du document, un score de service calculé selon le nombre d'hyperliens de type service disponibles dans le document. Alors que dans [120], l'amélioration de la précision de recherche pour une requête transactionnelle est effectuée par le calcul d'une probabilité *a priori* de la page en tant que page transactionnelle basée sur l'existence des caractéristiques transactionnelles telles que des hyperliens ayant des textes d'ancrage contenant des verbes transactionnels tels que *buy*, *download* etc. Dans [95], trois algorithmes d'appariement sont utilisés pour les trois types de requêtes. L'algorithme OKAPI [160] est appliqué pour les requêtes informationnelles. Pour les requêtes navigationnelles, l'algorithme PageRank et l'information des URL sont combinés avec le score OKAPI. Pour les requêtes transactionnelles, l'information des hyperliens de services est combinée avec le score OKAPI.

2.4 Accès contextuel à l'information guidé par le profil utilisateur

La RI contextuelle guidée par le profil utilisateur (connue sous le nom de la RI personnalisée) est une branche de la RI contextuelle dont le contexte prend une dimension cognitive et est défini par le profil de l'utilisateur.

Nous présentons dans cette section le cadre et les motivations de l'apparition de la RI personnalisée, la notion de profil utilisateur, l'architecture fonctionnelle d'un SRI personnalisé, les techniques de modélisation du profil utilisateur ainsi que les modèles d'accès personnalisé à l'information.

2.4.1 Cadre et motivations

La personnalisation est une discipline de recherche qui est apparue avec la notion du profil utilisateur vers les années 80 avec les assistants et les agents d'interface. Le but était de créer des applications personnalisées permettant de s'adapter à l'utilisateur [105].

En RI personnalisée, la dimension utilisateur est décrite par son profil représentant ses centres d'intérêts, ses connaissances et ses buts de recherche. Il a été démontré que le profil utilisateur est l'élément contextuel le plus important [138] qui permet d'améliorer la précision de la recherche. L'exploitation de cette dimension ne permet pas seulement de répondre à des requêtes ambiguës ou récurrentes et les interpréter en fonction du profil utilisateur mais aussi à enrichir et faire évoluer la représentation des connaissances de l'utilisateur ses centres d'intérêts et ses buts de recherche. Dans ce sens, la personnalisation intègre un processus de construction et d'évolution des centres d'intérêts de l'utilisateur au cours du temps.

En outre, plusieurs domaines d'application ont recours à la personnalisation avec des besoins différents. On cite le commerce électronique (e-commerce), la dissémination sélective de l'information, l'apprentissage assisté par ordinateur (e-learning), l'accès aux bibliothèques digitales, les systèmes d'informations mobiles (téléphonie mobile, agendas personnels, systèmes embarqués), configuration des logiciels, etc. Selon les domaines, la personnalisation consiste en l'une ou plusieurs des tâches suivantes : filtrer un flux d'information, guider la navigation dans un espace d'informations trop vaste, recommander un ensemble d'informations à l'utilisateur, ajuster les résultats d'une requête selon le profil, adapter l'interaction à la situation de l'utilisateur (interface, interaction). Quel que soit le domaine d'application, la notion de profil utilisateur est défini selon des dimensions liées à la finalité du système.

2.4.2 La notion de profil utilisateur

La notion de profil utilisateur est au cœur de la personnalisation en RI. A la différence de la notion du contexte qui couvre des éléments contextuels liés au système, à l'utilisateur ainsi qu'à des facteurs impliqués durant l'interaction

utilisateur-système, le profil utilisateur est défini par des éléments contextuels directement liés à l'utilisateur, tels que ses centres d'intérêts, ses préférences de recherche, etc. Le profil utilisateur fait partie du contexte cognitif de l'utilisateur selon la taxonomie du contexte multidimensionnel présenté dans [177]. On suppose que le profil a un caractère plus invariant comparativement au contexte de la tâche ou du contexte mobile même si les centres d'intérêts et les préférences de recherche évoluent au cours du temps.

Plusieurs définitions du profil ont été abordées dans la littérature en RI personnalisée. On distingue le profil cognitif [121, 116, 139], le profil qualitatif [74] et le profil multidimensionnel [108][179].

Le profil cognitif exploité dans plusieurs systèmes personnalisés [121, 116, 139] est analogue au contexte cognitif de l'utilisateur traité dans la taxonomie du contexte multidimensionnel dans [177] (cf. Sect.2.2.2.4 p.34).

Le profil qualitatif dans [74] est lié aux préférences de recherche de l'utilisateur quant à la qualité de l'information restituée par le système (fraîcheur, crédibilité des sources d'information, cohérence, etc.). Ces critères concernent le contexte du document qualitatif traité dans la taxonomie du contexte multidimensionnel dans [177] (cf. Sect.2.2.2.4 p.34).

Dans certaines études, le profil couvre en plus des centres d'intérêts et des préférences de l'utilisateur, des caractéristiques de l'environnement et du système, définissant ainsi un profil multidimensionnel [108][179]. Un profil multidimensionnel se rapproche de la notion du contexte multidimensionnel et qui couvre toutes les dimensions possibles impliqués dans l'interaction de l'utilisateur avec le système.

Bien que certains systèmes s'accordent sur la même définition du profil, les concepts et les formalismes diffèrent entre les systèmes ou les applications. En effet, les approches en RI personnalisée se distinguent par les techniques de modélisation et d'évolution du profil utilisateur en première phase et la personnalisation du processus de recherche par exploitation du profil utilisateur dans la chaîne d'accès à l'information en deuxième phase.

2.4.3 Architecture fonctionnelle d'un système de RI personnalisé (SRIP)

Le but fondamental d'un SRI personnalisé est de satisfaire les besoins en information de l'utilisateur en intégrant son profil dans la chaîne d'accès à l'information. L'architecture générale d'un SRIP est présentée dans la figure

2.3. Elle était introduite dans le cadre du projet *APMD*³ qui s'inscrit dans le cadre de l'accès personnalisé à de grands volumes d'informations. On distingue deux principales composantes : la modélisation et la gestion de l'évolution des profils utilisateur et l'accès personnalisé à l'information.

- La modélisation du profil repose sur des techniques et des outils permettant non seulement de représenter et construire le profil de l'utilisateur mais aussi de gérer son évolution de manière dynamique au cours du temps. Les défis majeurs présents dans cette phase concernent l'identification implicite des sources d'évidence utiles à une tâche de recherche courante et l'exploitation de ces informations dans la dérivation et l'évolution du profil.
 1. Représentation du profil utilisateur : le modèle du profil consiste à spécifier sous quelle forme les données du profil doivent être représentées. Les modèles de représentations peuvent être simples basés sur des mots clés ou complexes basés sur des ontologies de domaines ou des hiérarchies de concepts. La construction du profil utilisateur consiste à collecter et exploiter les données et sources d'information pertinentes pour les représenter. La collecte de ces sources d'information intègre la spécification du type des données pertinentes à collecter, le mode d'acquisition des données (explicite ou implicite). Le mode d'acquisition explicite des données est le plus facile à mettre en œuvre et permet à l'utilisateur de saisir manuellement des informations utilisées dans la construction du profil. En revanche, le mode d'acquisition implicite repose sur des techniques d'extraction des informations basées sur des mesures de pertinence implicite (fréquence de *clicks*, temps de lecture, etc) appliqués sur l'historique d'interactions de l'utilisateur.
 2. Évolution du profil utilisateur : le but fondamental de l'évolution du profil est d'affiner la connaissance que le système a de l'utilisateur en mettant à profit l'expérience passée lors de ses activités de recherche. Sous l'angle de la dimension temporelle, l'évolution du profil à long terme se fait par le déploiement de l'information issue des activités de recherche de l'utilisateur dans un processus similaire à la construction du profil. Cependant, l'évolution du profil à court terme ou du profil à long terme construit par agrégation des profils à court terme intègre des algorithmes et des techniques capables de détecter des changements possibles des centres d'intérêts au cours des sessions de recherche et par conséquent propager des changements au niveau de la représentation du profil utilisateur.
- L'accès personnalisé à l'information consiste à intégrer le profil utilisateur dans l'une des phases du processus de RI, notamment la reformulation de requêtes, l'appariement requête-document ou la présentation des résultats. Le but de cette phase est de renvoyer en haut de la liste de

³<http://apmd.prism.uvsq.fr/>

résultats présentés à l'utilisateur, ceux qui correspondent à ses centres d'intérêts et ses intentions de recherche.

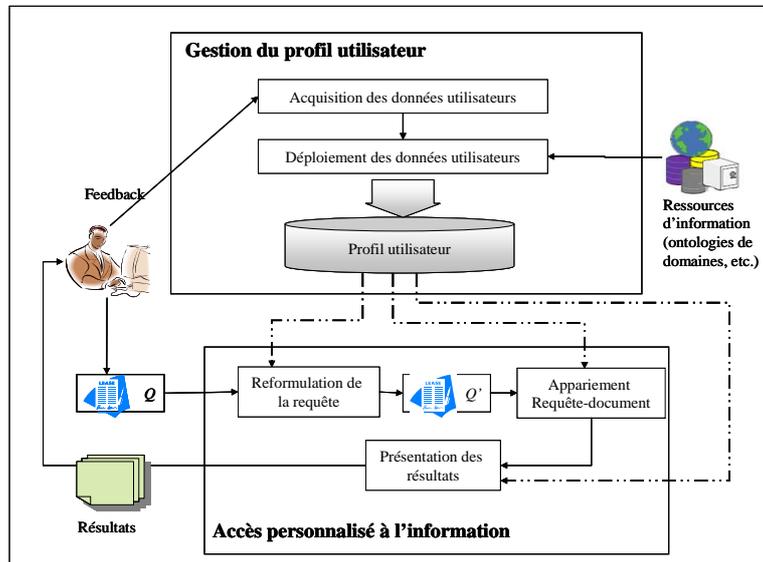


FIG. 2.3 – Architecture fonctionnelle d'un SRIP

2.4.4 Représentation et construction du profil utilisateur

La représentation du profil utilisateur consiste à représenter de manière interne ses centres d'intérêts. Indépendamment du modèle de représentation du profil, la construction du profil utilisateur repose sur des outils et des algorithmes permettant de collecter les sources d'information pertinentes servant à la dérivation du profil utilisateur en première phase, et des techniques de déploiement de ces sources d'information dans le but de le représenter en deuxième phase. Nous présentons dans cette section les techniques d'acquisition des sources d'information et les techniques de construction du profil selon plusieurs modèles de représentation possibles : la représentation basée sur l'historique de recherche, la représentation ensembliste, connexionniste ou conceptuelle.

2.4.4.1 Acquisition et collecte des sources d'information

L'acquisition des sources d'information et des données utilisateur est l'une des problématiques importantes à aborder dans le but de développer un SRI personnalisé efficace. On distingue deux modes d'acquisition des données utilisés dans la construction du profil utilisateur : acquisition explicite et acquisi-

tion implicite.

Acquisition explicite l'acquisition explicite des données utilisateur repose principalement sur les techniques de feedback explicite largement utilisées dans les systèmes de filtrage et de reformulation de requêtes par réinjection de pertinence. Ces techniques sont utilisées dans le cadre de la RI personnalisée dans le but de maintenir et faire évoluer le profil de l'utilisateur. L'acquisition explicite permet à l'utilisateur de saisir manuellement ses domaines d'intérêts ou alors juger de façon explicite les documents renvoyés pour une requête servant à la construction de son profil.

Ce type d'acquisition est utilisé dans plusieurs systèmes de RI personnalisée en ligne. *Google's Alerts* permet à l'utilisateur de saisir une requête profil servant comme critère de sélection de l'information. *Google personalized search* dans sa version 1.1 (2004) permet à l'utilisateur de saisir explicitement un ensemble de catégories/concepts qui représentent ses domaines d'intérêts. La recherche contextuelle mise en œuvre par yahoo dans sa version *Y!Q* [110] permet à l'utilisateur de sélectionner explicitement un texte d'une page en cours de lecture et qui sert à représenter un profil textuel lié à la recherche courante.

L'approche dans [165] requiert que l'utilisateur sélectionne parmi l'ensemble des concepts de l'ontologie de l'ODP une paire de concepts, un adéquat et un autre inadéquat au besoin en information, utilisés dans un processus de reformulation personnalisée de la requête.

Plusieurs autres systèmes en RI personnalisée adoptent ce mode d'acquisition. Liu et Yu [123] propose à l'utilisateur de sélectionner les concepts du profil issus de l'ODP, dans un processus de clarification du besoin derrière la requête et qui servent comme contexte de recherche. Koutrika et Ioannidis [109] et les systèmes ifWeb, Wifs et InfoWeb utilisent un feedback explicite de l'utilisateur permettant d'identifier les documents pertinents utilisés dans la dérivation du réseau de termes représentant le profil utilisateur.

Même si ces techniques servent à la construction contrôlée du profil utilisateur, cependant elles présentent des limitations. D'une part, l'acquisition explicite ne permet pas d'améliorer la représentation du modèle de l'utilisateur efficacement, plus particulièrement si les caractéristiques de l'interface d'acquisition explicite ne sont pas très bien assurées [192]. D'autre part, l'effort supplémentaire imposé à l'utilisateur à spécifier explicitement ses besoins est indésirable pour des utilisateurs Web. Ces limitations ont orienté les travaux vers des techniques d'acquisition implicite des données utilisateur dans le processus de construction du profil utilisateur.

Acquisition implicite Une approche alternative remplaçant l'acquisition explicite des besoins en information de l'utilisateur, consiste à développer des algorithmes d'acquisition implicite de ces besoins. L'acquisition implicite des données utilisateurs consiste à collecter les données de l'utilisateur en obser-

vant ses comportements et interactions avec le système durant les activités de recherche. L'avantage de cette approche est qu'elle ne nécessite aucune implication directe de l'utilisateur. Nous présentons dans cette section la collecte implicite des données utilisateurs exploitées dans la dérivation du profil utilisateur ainsi que les indicateurs de pertinence implicite utilisés dans cette collecte implicite.

– **Collecte implicite données utilisateurs :**

Le processus clé dans la plupart des approches de personnalisation consiste à exploiter des sources d'évidence issues de l'historique de recherche de l'utilisateur, afin d'inférer son profil. Letizia [121], WebMate [41], PersonalWebWatcher [132] et OBIWAN [67] construisent le profil en analysant les pages web visitées par l'utilisateur lors de sa recherche. D'autres sources d'information sont également exploitées, telles que les annotations dans Basar [185], les requêtes et ses résultats de recherche dans [144], Syskill and Webert [139], et Persona [182], les requêtes et les résumés textuels de ses résultats [163]. La combinaison des sources d'évidence multiples, telles que les pages web, les emails et les documents textes, est adoptée dans [57]. Comme source de personnalisation, plusieurs systèmes de recherche sur ordinateur, notamment, Copernic, Google Desktop Search, Mac Tiger, Windows Desktop Search, Yahoo! Desktop Search ou X1, et d'autres outils de recherche fournissent un accès simple à l'index de l'information créé, copié, visualisé par les utilisateurs. Microsoft's Stuff I've Seen (SIS) [184] fournit aussi ce type de personnalisation.

Après avoir collecté les données utilisateurs, ils seront exploités dans le but de dériver le profil utilisateur. Certaines études utilisent des techniques de *data mining* sur l'historique de recherche de l'utilisateur ou sur le fichier *log* d'un moteur de recherche [59]. D'autres études utilisent des machines d'apprentissage [195] des données utilisateurs dans le but d'inférer son profil et améliorer ainsi l'efficacité de la recherche.

– **Principe de l'acquisition implicite :**

Les techniques de collecte implicite des données utilisateurs se basent principalement sur les comportements observables de l'utilisateur durant ses interactions avec le système dans le but d'identifier ses intérêts et ses préférences. Un comportement observable de l'utilisateur constitue l'ensemble des actions qu'il effectue face aux résultats fournis par le système d'accès à l'information. D'après [137], toute interaction de l'utilisateur avec le système est considérée comme une estimation de son jugement d'intérêts.

Selon les études du domaine [136, 100], les mesures de pertinence implicite sont basées sur des comportements de l'utilisateur catégorisés par les types d'actions faites sur le document résultat. On cite la catégorie de comportements utilisateurs par la liste suivante : examiner, sélectionner, mettre en référence et annoter, etc. Bharat et al. [25] proposent l'observa-

tion des comportements de l'utilisateur lors de ses activités courantes. Ils identifient certains types d'événements tels que le défilement de page ou la sélection d'un article comme des indicateurs de pertinence sur le document. Oard et Kim [OK01] classifient les comportements observables de l'utilisateur selon deux axes, la catégorie du comportement et l'unité sur laquelle se porte l'observation. La catégorie du comportement (Examine, Retain, Reference and Annotate) traduit le but du comportement. La partie de l'objet sur laquelle l'action est appliquée (Segment, Object and Class) désigne la plus petite partie de l'objet ayant subi l'action. Cette classification est présentée dans le tableau 2.1. La classification des comportements dans [101] ajoute à cette dernière classification une cinquième catégorie, nommée *Create* et d'autres comportements à la catégorie *Examine* (*Scroll, Find, Query*). Cette catégorie décrit les comportements dans lesquels l'utilisateur s'engage pour créer de l'information (*Type, Edit, etc.*).

Salojarvi et al. [17] étudient les mouvements de l'utilisateur notamment la direction de son regard comme source de réinjection de pertinence implicite pour inférer ses centres d'intérêts. En outre, Joachims et al. [92] collecte des mesures implicites basée sur les données de *clics* passés fourni dans le fichier *log* des requêtes et qui font les données d'apprentissage des fonctions d'ordonnancement par apprentissage.

Plusieurs SRI personnalisés dédiés pour la recherche sur le web considèrent que le *clic* sur une page est une estimation de pertinence permettant de traduire un degré d'intérêt non négligeable de la page *cliquée*. Nous citons le système OBIWAN [67] qui exploite les pages web *cliqués* par l'utilisateur pour la requête comme les sources d'évidences principales pour la construction du profil utilisateur. L'étude faite dans [64] montre la corrélation positive entre les jugements de pertinence explicites donnés par des utilisateurs et des mesures implicites de pertinence en recherche sur le web.

2.4.4.2 Représentation et dérivation du profil utilisateur

2.4.4.2.1 Représentation basée sur l'historique de recherche La représentation du profil utilisateur basée sur l'historique de recherche consiste en l'ensemble des requêtes et des pages web visités ou cliqués de l'utilisateur (*Google My search history, Yahoo My Web*, [141]) ou l'ensemble des requêtes et les résumés textuels de ses résultats associés [181] accumulés au cours des sessions de recherche de l'utilisateur.

Après la première tentative de personnalisation de la recherche dans *Google personalized search 1.1*, *Google* propose une version récente⁴ basée sur la construc-

⁴<http://www.google.com/psearch>

Category behavior	Segment	Object	Class
Examine	View Listen	Select	
Retain	Print	Bookmark Save Delete Purchase	Subscribe
Reference	Copy- and-paste Quote	Forward Reply Link Cite	
Annotate	Mark up	Rate Publish	Organize

TAB. 2.1 – Classification des comportements utilisés comme mesures de pertinence implicite par Oard et Kim [OK01]

tion implicite d'un profil utilisateur basée sur l'historique de recherche. Le profil est ainsi représenté par requêtes et les sites Web que l'utilisateur a sélectionnés de l'ensemble des résultats retournés en réponse à la requête. La figure 2.4 montre un exemple d'un historique de recherche d'un utilisateur qui s'intéresse à la musique et aux langages de programmation. Suite à la mise en ligne de ce outil par Google, Yahoo propose un outil de recherche personnalisé, appelé *Yahoo My Web*, qui définit un profil utilisateur évolutif représenté par l'historique de recherche de l'utilisateur. Il est basé sur l'ensemble des pages Web visités par l'utilisateur ainsi que des pages annotées (*bookmark pages*).

Raghavan et Sever [141] représente un profil de l'historique de recherche servant comme une base de données des requêtes soumises précédemment ainsi que leurs résultats associés. Cette base de données est utilisé pour sélectionner les requêtes les plus similaires à une requête en cours d'évaluation. L'approche proposé par Shen et al. [181] représente le profil utilisateur par l'historique des requêtes passés et l'historique des données de clics sur les résultats sélectionnés par l'utilisateur en réponse à ces requêtes.

2.4.4.2.2 Représentation ensembliste La représentation ensembliste du profil utilisateur est basée sur un ensemble de mots clés (ou vecteurs de mots clés) pondérés représentés souvent selon le modèle vectoriel de Salton [158]. Ce type de représentation est le premier conçu pour modéliser le profil utilisateur. Les paquets de termes traduisent les centres d'intérêts de l'utilisateur. Le poids d'un terme est souvent calculé selon le schéma TF*IDF communément utilisé

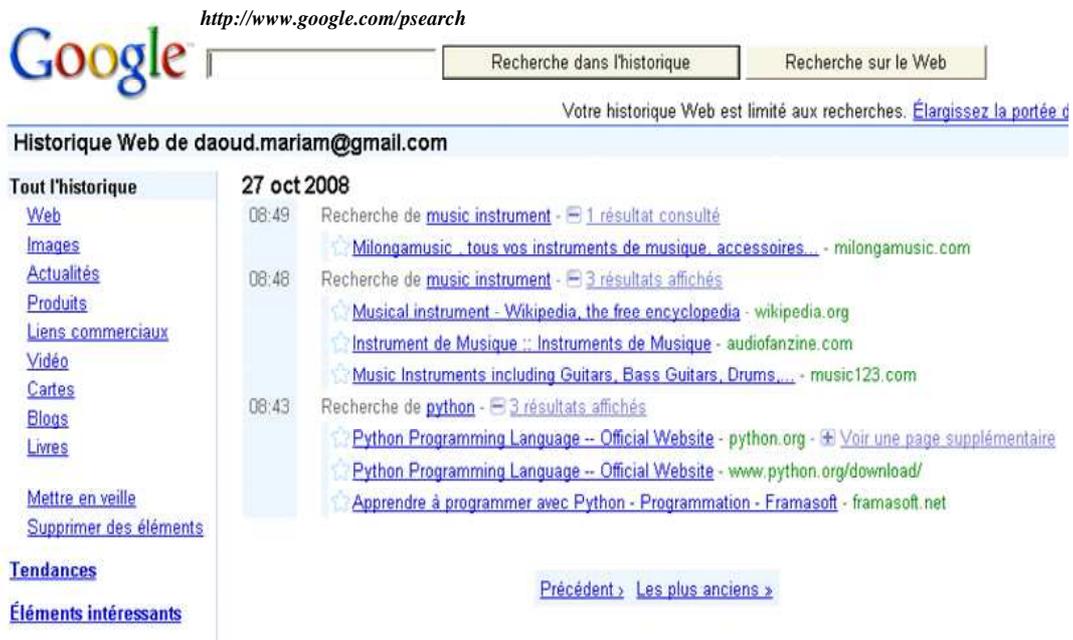


FIG. 2.4 – L'historique de recherche utilisé dans *Google psearch* qui sauvegarde les requêtes et les résultats sélectionnés par l'utilisateur

en RI [158] et représente le degré d'intérêt de l'utilisateur dans le profil. Nous citons trois sous-modèles de représentation ensemblistes du profil utilisateur : un ensemble de termes pondérés où chaque terme représente un centre d'intérêt possible de l'utilisateur, un vecteur de termes pondérés représentant un centre d'intérêt [122] ou un ensemble des vecteurs de termes pondérés dont chacun représente un centre d'intérêt [69, 165].

La construction d'un profil ensembliste repose sur des techniques d'extraction des termes à partir des documents pertinents, jugés implicitement ou explicitement par l'utilisateur. Plusieurs systèmes d'accès personnalisé à l'information adoptent ce type de représentation. Tel est le cas des portails web tels que *MyYahoo*, *InfoQuest*, *Anatagonomy* qui est un système personnalisé de consultation de nouvelles et de journaux en ligne [153], *Fab* [20] un système de recommandation de page web, *Letizia* [121], un système d'aide à la navigation, et *Syskill & Webert* [139] un système de recommandation, etc. *Google's Alerts* représente le profil utilisateur par une requête appelée *routing query* pour laquelle le système renvoie à l'utilisateur des informations par email à partir du Web et des sites des News. Le moteur de recherche le plus concurrent sur le Web *yahoo contextual search* (Y!Q) tient en compte un besoin de l'utilisateur à court terme représenté par un ensemble de termes issus d'un texte/paragraphe sélectionné à partir de la page en cours de lecture. *Google personalized search* dans sa version 1.1, représente un profil utilisateur par un ensemble de catégories représentés chacune par un terme ou un ensemble de termes saisis explicitement par l'utilisateur et utilisés dans l'amélioration de toute requête.

Dans les systèmes cités précédemment, l'utilisateur est modélisé par un mono-profil contenant un seul centre d'intérêt. Vu que l'utilisateur peut avoir plusieurs centres d'intérêts lors de sa recherche, plusieurs SRI personnalisés [69, 203] adoptent une représentation du profil basé sur un ensemble des vecteurs de termes traduisant des centres d'intérêts multiples de l'utilisateur. Parmi les systèmes ayant une représentation ensembliste du profil, nous citons les systèmes *WebMate* [41] et *Alipes* [199]. Dans *WebMate* [41], le profil est représenté par N vecteurs multiples, un par domaine d'intérêt où N est le nombre des centres d'intérêts dans le profil. Les vecteurs des N premiers documents, jugés pertinents, sont utilisés pour représenter les N centres d'intérêts. La construction du profil est basée ensuite sur la classification d'une nouvelle page web consultée par l'utilisateur dans le centres d'intérêts le plus similaires selon la mesure *cosinus* par combinaison des poids des termes. Par opposition au nombre fixé au préalable de centres d'intérêts par profil dans *WebMate*, le système *Alipes* [199] crée des nouveaux centres si la similarité entre un nouveau document à classer dans des centres d'intérêts déjà créés est toujours inférieure à un seuil de similitude prédéterminé. Si la similitude excède un seuil, le vecteur de document est combiné avec le centre d'intérêt le plus proche.

Dans *WebPersonae* [69], le profil utilisateur consiste en des centres d'intérêts représentés chacun par un vecteur de termes pondérés. Les documents jugés pertinents implicitement par l'utilisateur sont représentés par des vecteurs de termes pondérés selon le schéma de pondération classique TF*IDF, et subissent une technique de catégorisation basée sur l'algorithme k-means afin d'obtenir des classes ou groupes des documents. Le poids d'un terme dans un centre d'intérêt (représenté par un vecteur centroïde d'une classe) est calculé comme la moyenne de son poids dans l'ensemble de documents classifiés dans la classe représentant ce centre d'intérêt. L'ensemble des vecteurs centroides des classes obtenus constituent le profil à long terme de l'utilisateur. Ce système initialise le profil à court terme par comparaison aux centres d'intérêts déjà appris. En effet, les dernières pages visitées par l'utilisateur sont concaténées pour créer un vecteur de termes pondérés V . Le profil utilisateur à court terme est initialisé par le centre d'intérêt qui excède un seuil de similarité prédéterminé avec le vecteur V qui représente les activités de recherche courantes de l'utilisateur.

L'approche dans [203] se base sur une représentation vectorielle par mots clés pondérés des centres d'intérêts de l'utilisateur. L'approche de base consiste à représenter dans un premier temps la session de recherche selon une matrice document-termes avec leur poids associés, où les documents sont ceux jugés pertinents par l'utilisateur de manière implicite. Le coefficient d'importance d'un terme t dans un document d à l'instant s noté $CPT^s(t, d)$ est défini comme suit :

$$CPT^s(t, d) = \frac{w_{(t,d)}}{dl} * \sum_{t' \neq t, t' \in T(R_u^s)} EMIM(t, t') \quad (2.1)$$

$w_{(t,d)}$ est le poids du terme t dans le document d calculé selon le schéma classique $TF \star IDF$, dl est la longueur du document d , $EMIM(t, t')$ est le degré de corrélation entre les termes t et t' quantifié à l'aide de la mesure EMIM (Expected Mutual Information Measure). Celle-ci est appliquée sur l'ensemble de documents jugés pertinents par l'utilisateur à partir des sessions de recherche précédentes.

L'historique d'interactions de l'utilisateur est représenté par une matrice obtenue par agrégation des matrices des sessions de recherche précédentes. A partir de la matrice d'interaction, un vecteur de termes pondérés représentant le contexte courant est ainsi extrait, appelé contexte d'usage, et représente le centre d'intérêt courant de l'utilisateur. Les contextes d'usages construits sur plusieurs sessions de recherche définissent le profil utilisateur à long terme.

La représentation ensembliste du profil utilisateur apporte l'avantage de la simplicité de mise en œuvre. Cependant, même si les modèles de représentation ensembliste permettent de traduire une multiplicité des centres d'intérêts de l'utilisateur, cette représentation manque de structuration, de cohérence, des niveaux de généralité/spécificité et des relations de corrélation entre les divers centres d'intérêts de l'utilisateur.

2.4.4.2.3 Représentation connexionniste La représentation connexionniste du profil utilisateur consiste à représenter les centres d'intérêts de l'utilisateur par un réseau de nœuds pondérés dont chaque nœud représente un concept traduisant un centre d'intérêt de l'utilisateur. Cette représentation permet de résoudre les failles de la représentation ensembliste par la mise en place des relations de corrélation sémantiques entre les centres d'intérêts du profil. En effet, la richesse sémantique dans cette représentation permet de résoudre le problème de la polysémie des termes inhérents à la représentation ensembliste, l'incohérence possible entre les centres d'intérêts et l'identification d'un profil adéquat au sujet de la requête via les relations sémantiques.

Plusieurs SRI personnalisés adoptent ce type de représentation [22, 129, 68, 109]. La construction d'un profil utilisateur connexionniste consiste non seulement à extraire des termes à partir des documents pertinents de l'utilisateur, mais à intégrer ces termes dans un réseau de nœuds. La construction de tels profils nécessitent la création des relations de corrélation sémantiques entre les nœuds du réseau.

IfWeb [22] est un assistant personnel à la navigation, recherche et filtrage des documents adapté aux besoins spécifiques de l'utilisateur. Le profil utilisateur y est représenté par un réseau sémantique dont les nœuds correspondent aux termes (concepts) extraits des documents ainsi que des descriptions textuelles données par l'utilisateur comme des exemples positifs ou négatifs. Un arc reliant

deux termes dans le réseau traduit une relation de co-occurrence de ces deux termes dans un même document. La corrélation sémantique entre les nœuds du réseau permet de résoudre le problème de polysémie inhérente à l'appariement basé mots clés.

Le système *Wifs* décrit dans [129] est capable de filtrer les pages Web HTML et les documents textuels retournés par le moteur de recherche ALTAVISTA⁵ pour une requête donnée. Ce système intègre un profil utilisateur représenté par un réseau de termes reliés par des relations de co-occurrences. Par opposition au système *ifWeb*, les termes du réseau dans *wifs* sont issus à partir d'une base de données (TDB), créée préalablement par des experts qui sélectionnent les termes les plus pertinents pour représenter un domaine.

La figure 2.5 (a) illustre une description simplifiée d'un profil utilisateur hypothétique. Dans cette représentation, les termes actifs, nommées planètes T_1, T_2, \dots, T_n , sont contenus à la fois dans les documents jugés pertinents par l'utilisateur et dans la base de données (TDB) et les termes satellites t_1, t_2, \dots, t_m sont les termes contenus dans les documents pertinents, et qui n'existent pas dans TDB, mais qui co-occurrent avec les termes T_i . La représentation abstraite d'un document est similaire à celle du profil comme illustré dans la figure 2.5 (b), à la seule différence que les arcs reliant les termes planètes aux termes satellites ne sont pas pondérés.

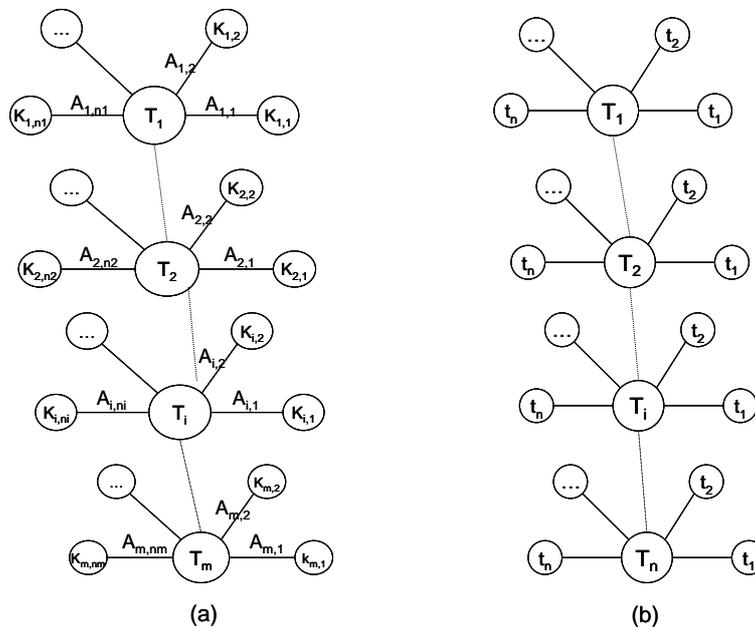


FIG. 2.5 – Représentation du profil utilisateur et du document dans le système Wifs

Infoweb [68] est un système de personnalisation interactif développé pour la recherche dans les librairies digitales. Par analogie au système Wifs, ce système

⁵<http://www.altavista.com>

présente une base de stéréotypes, appelée "*stereotype knowledge base*". Un stéréotype représente une classe d'utilisateurs et contient un vecteur représentatif des documents associés. La classe est définie préalablement par un expert de domaine. Les stéréotypes de la base sont utilisés pour l'initialisation du modèle de l'utilisateur.

La représentation hiérarchique des centres d'intérêts (UIH) développée dans [102] prend en compte des niveaux de généralité et de spécificité liés aux centres d'intérêts. La construction de cette hiérarchie est à la base d'un algorithme de clustérisation des termes des pages web jugés pertinents par l'utilisateur. Les termes reliés sémantiquement et qui se trouvent dans un même document sont groupés dans une même classe de la hiérarchie. Les centres d'intérêts généraux ou à long terme sont représentés par les hauts niveaux de la hiérarchie et les niveaux bas de la hiérarchie représentent des centres d'intérêts spécifiques à court terme. La figure 2.6 présente un exemple d'un profil utilisateur selon le modèle de représentation proposé dans [102].

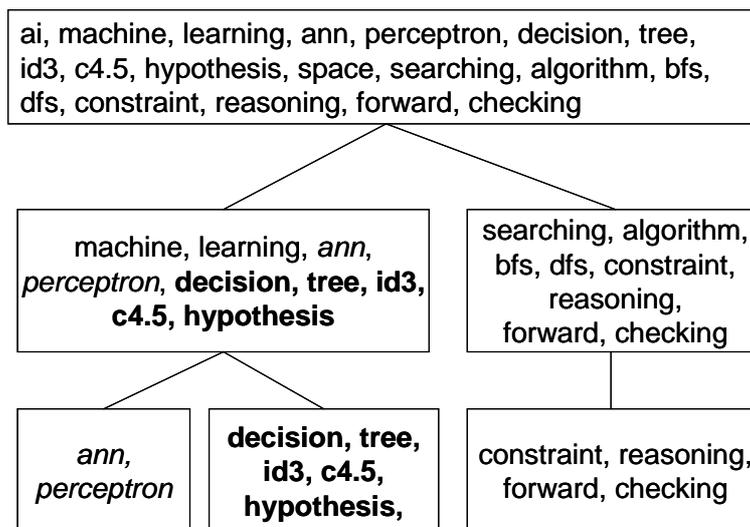


FIG. 2.6 – Exemple d'un profil utilisateur représenté par un hiérarchie de concepts

Dans [109], le profil utilisateur est représenté par un ou plusieurs réseaux de termes connectés par des opérateurs logiques : conjonction, disjonction, substitution et négation qui traduisent des réécritures possibles entre termes. La figure 2.7 illustre un exemple de ce type de profil. Les opérateurs entre les termes servent à transformer une requête utilisateur en une version personnalisée traduisant plus finement le besoin en information de l'utilisateur. A chaque itération de recherche, un sous-graphe est créé à partir des termes de la requête et les termes issus des documents jugés explicitement pertinents par l'utilisateur en respectant des règles de création des arcs. Ces règles se basent principalement sur des relations des co-occurrences entre termes dans le but de

créer des arcs de conjonction, de disjonction, de substitution ou de négation.

La pondération d'un nœud terme dans le profil est donné dans la formule suivante :

$$w_t = f_t * (N_t/N) \tag{2.2}$$

N_t est le nombre de documents précédents contenant le terme t ; N est le nombre total de documents contenant le terme t ; f_t est la fréquence du terme dans les documents précédents sur le nombre total des termes des documents contenant ce terme. Le poids d'un arc est calculé comme suit :

$$w_e = (N_e/N') \tag{2.3}$$

N_e est le nombre de documents précédents contenant les termes reliés par l'arc e , N' est le nombre total de documents indexés par les termes de l'arc e .

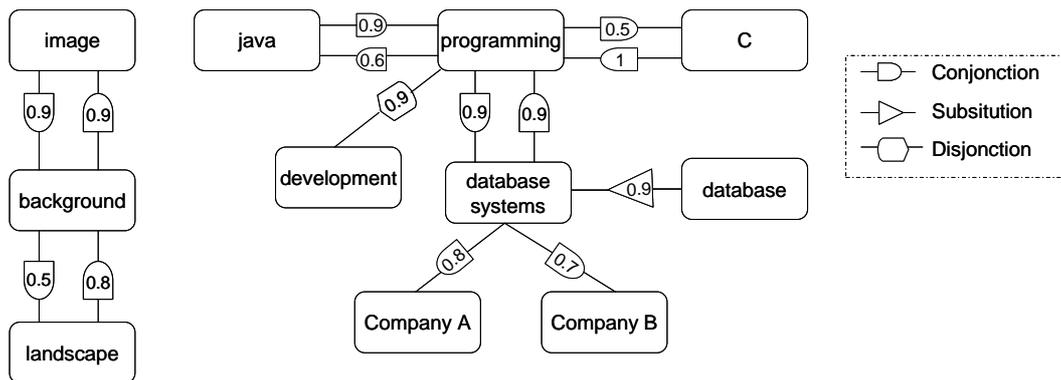


FIG. 2.7 – Exemple d'un profil utilisateur représenté par un graphe de termes reliés par des opérateurs logiques

La représentation connexionniste du profil apporte de la sémantique au modèle de l'utilisateur. Même si ce modèle de représentation permet de résoudre les problématiques liées à la représentation ensembliste en enrichissant le profil par des relations sémantiques entre les centres d'intérêts le constituant, elle présente certaines limitations. En effet, les termes du réseau sémantique représentant le profil utilisateur sont issus à partir de l'historique de recherche de l'utilisateur qui est souvent limité. Le système dispose d'un profil composé des termes avec lesquels l'utilisateur est familier. Cette situation rend la détection d'un nouveau besoin en information pour une nouvelle requête utilisateur, une tâche difficile à accomplir.

2.4.4.2.4 Représentation conceptuelle Dans le but de pallier les limitations de la représentation connexionniste, plusieurs études ont recours à une modélisation du profil utilisateur selon une représentation conceptuelle qui se

base sur l'exploitation des ontologies de domaines ou des hiérarchies de concepts préalablement définies [123, 67, 164]. La représentation conceptuelle consiste à représenter les centres d'intérêts de l'utilisateur par un réseau de nœuds conceptuels et reliés entre eux en respectant la topologie des liens définis dans les hiérarchies et les ontologies de domaines. Chaque concept décrivant un centre d'intérêt est représenté par un vecteur de termes pondérés où le poids traduit le degré d'intérêt de l'utilisateur dans ce concept du profil.

Il existe plusieurs hiérarchies de concepts ou des ontologies de domaines conçues dans le but de répertorier le contenu des pages web pour une navigation facile par les utilisateurs. On cite les portails en ligne tels que *Yahoo*, *Magellan*, *Lycos*, et l'*Open Directory Project (ODP)*. Ces ressources sont exploitées dans plusieurs systèmes de RI personnalisée [123, 67, 164] comme une source de connaissance sémantique dans le processus de construction du profil utilisateur.

La construction d'un profil utilisateur conceptuel repose principalement sur l'utilisation de deux sources d'évidence, les données utilisateurs collectées au cours de ses sessions de recherche et des ressources sémantiques prédéfinies. L'approche de construction du profil conceptuel consiste tout d'abord à spécifier les niveaux des concepts de l'ontologie à considérer, et ensuite appliquer le procédé de déploiement des données dans des techniques de pondération de ces concepts.

Le profil sémantique dans [15, 124] est représenté par un ensemble de concepts issus de l'ontologie. Liu et Yu [123] modélisent deux types de profils utilisateurs, le premier consiste en des concepts représentés par un ensemble de termes pondérés et est construit sur la base de l'historique de recherche de l'utilisateur et le deuxième est un profil général applicable à tous les utilisateurs et représenté par les concepts des trois principaux niveaux de la hiérarchie de l'ODP. Le profil personnel est inféré à partir de l'historique de recherche de l'utilisateur représenté par l'ensemble des requêtes, les catégories associées à chaque requête et l'ensemble de résultats sélectionnés par l'utilisateur et classifiés dans les catégories associées à la requête. Les catégories associées à la requête sont issus d'une hiérarchie de concepts exploités dans le but de présenter à l'utilisateur des résultats groupés thématiquement dans ces concepts. A partir de cet historique, une matrice documents-catégorie et une matrice document-terme permettent d'inférer le profil utilisateur selon une matrice catégorie-terme pondérée.

Sieg et al. [165] exploitent simultanément des centres d'intérêts de l'utilisateur préalablement appris et représentés selon des vecteurs de termes pondérés et l'hiérarchie de concepts "Yahoo" dans le but de construire un profil utilisateur sémantique. Celui-ci contient ainsi des contextes associés aux requêtes soumises par l'utilisateur, formés chacun d'une paire de concepts de la hiérarchie : le premier représente le concept adéquat à la requête, et l'autre représente le

concept à exclure dans la recherche. L'identification du contexte associé à la requête est effectuée selon deux phases : la première phase consiste à projeter la requête sur l'hierarchie et extraire l'ensemble des concepts les plus similaires. La deuxième phase consiste à sélectionner à partir de l'ensemble des concepts obtenus, le concept le plus similaire à un centre d'intérêt déjà appris (celui qui excède un seuil de similarité prédéterminé) et le concept le moins similaire (qui n'excède le seuil de similarité avec aucun centre d'intérêt déjà appris).

Une modélisation conceptuelle hiérarchique du profil utilisateur a été faite dans [164, 67] qui consiste à représenter le profil de l'utilisateur par une instance de l'ontologie de l'ODP. Chaque concept de l'ontologie est associé à un poids résultant de la classification de l'information jugée pertinente à l'utilisateur dans ce concept. Cette classification est basée sur une mesure de similarité vectorielle entre le document et chaque concept de l'ODP. Le poids d'un concept c_j du profil est calculé par accumulation de ses scores de similarités avec les pages p_k classifiés sous ce concept. La similarité d'un concept c_j avec une page p_k es calculée comme suit :

$$\text{similarity}(c_j, p_k) = \sum_{i=1, \dots, n} w_{ij} * p_{ik} \quad (2.4)$$

où n est le nombre total de termes uniques du vocabulaire des documents, w_{ij} est le poids du terme t_i dans le concept c_j et p_{ik} est le poids du terme t_i dans la page p_k .

Contrairement à la construction du profil utilisateur sur plusieurs sessions de recherche dans OBIWAN [67], le profil dans [164] est construit en exploitant les documents qui correspondent à une même requête vu que le profil traduit le centre d'intérêt associé au contexte de la requête elle même. Les poids des concepts du profil sont accumulés par application d'un algorithme de propagation des scores des concepts via les liens hiérarchiques de l'ontologie aboutissant ainsi à une instance de l'ontologie pondérée représentant le profil utilisateur.

Généralement, les hauts niveaux de la hiérarchie permettent de représenter le profil à long terme et les concepts de bas niveau permettent de représenter un niveau de spécificité élevé du profil utilisateur à court terme. Dans un contexte de recherche à grande échelle, telle que le Web, la représentation éparsée des centres d'intérêts de l'utilisateur selon une diversité des concepts de l'ontologie présente des limitations liées à l'identification des concepts pertinents à une recherche donnée parmi une masse importante des concepts de l'ontologie représentés dans le profil utilisateur. Ceci peut augmenter considérablement le temps d'exécution des requêtes personnalisées d'une part et la gestion d'évolution du profil utilisateur d'autre part.

2.4.5 Évolution du profil utilisateur

La gestion de l'évolution du profil utilisateur est un processus complémentaire à la construction d'un profil utilisateur et désigne leur adaptation à la variation des centres d'intérêt des utilisateurs au cours du temps. L'évolution du profil utilisateur se fait souvent selon un processus incrémental basé sur l'addition de nouvelles informations dans la représentation du profil.

Sous l'angle de la dimension temporelle, les approches gèrent l'évolution du profil à court terme, à long terme ou les deux à la fois. La gestion de l'évolution du profil utilisateur consiste principalement à capturer les changements des centres d'intérêts de l'utilisateur dans une première phase et propager ces changements au niveau de la représentation du profil.

Les techniques de collecte des informations utilisées dans la gestion de l'évolution du profil utilisateur sont relativement dépendantes de la portée temporelle du profil. On distingue le profil à court terme et le profil à long terme. Le premier représente les centres d'intérêts liés aux activités de recherche courantes de l'utilisateur. Le second représente les centres d'intérêts persistants de l'utilisateur et issus de son historique de recherche tout entier. S'il s'agit d'un profil utilisateur à court terme, la collecte de l'information est conjointement liée à des mécanismes de groupage des activités de recherche liées à un même besoin en information dans des sessions de recherche ou alors des algorithmes de délimitation des activités de recherche récentes dans un intervalle de temps précis.

L'évolution de la représentation du profil utilisateur implique un changement des degrés d'intérêts dans certains domaines qui se traduit par une mise à jour de la structure/contenu des centres d'intérêts préalablement appris ou alors l'apparition d'un nouveau besoin en information qui se traduit par un ajout d'un nouveau centre d'intérêt au profil utilisateur.

Nous présentons dans cette section la gestion de l'évolution du profil utilisateur à court terme et/ou à long terme dans quelques systèmes de RI personnalisés. Plus précisément, nous abordons l'initialisation, le changement des centres d'intérêts au cours de l'interaction de l'utilisateur avec le système suivi par l'évolution de la représentation du profil à court et/ou à long terme.

2.4.5.1 Évolution du profil utilisateur à court terme

Le profil utilisateur à court terme décrit des centres d'intérêts et des besoins utilisateurs liés aux activités et la tâche de recherche courante. Souvent ces besoins en information sont partiellement représentés par le sujet de la requête. On admet que le profil à court terme sert à mieux cibler la recherche

vu qu'il contient des données considérées spécifiques et pertinentes au besoin en information courant de l'utilisateur [163]. Le but fondamental de l'évolution du profil à court terme est d'améliorer la précision de recherche en utilisant le profil le plus utile et approprié à la requête et non bruité par des centres d'intérêts hors contexte de recherche. Par conséquent, ce profil permet d'adapter efficacement le processus de RI aux besoins en information spécifiques de l'utilisateur.

Dans certaines approches [57, 67], le profil à court terme ne représente pas nécessairement un même besoin en information mais peut traduire des multiples centres d'intérêts. Dans ces approches, l'évolution du profil utilisateur est liée à la délimitation des activités récentes de l'utilisateur par un intervalle de temps qui peut englober plusieurs sujets d'intérêts recherchés. D'autres travaux [163, 203, 164] définissent le profil utilisateur à court terme dans une session de recherche par un besoin en information unique. L'évolution du profil dans ce cas requiert des mécanismes de délimitations des sessions de recherche, où une session est définie par un ensemble de requêtes liées à un même besoin en information.

Parmi les SRI personnalisés qui s'adaptent aux besoins courants de l'utilisateur, on cite les portails Web (tels que MyYahoo, InfoQuest, etc.), *Yahoo Contextual Search* (Y!Q), Letizia [121], JITIR [143], WebPersonae [69] OBIWAN [67] et d'autres systèmes [165, 174, 180, 164, 110].

Yahoo Contextual Search (Y!Q) est un outil de recherche personnalisée à grande échelle et considéré le plus concurrent parmi les outils de RI personnalisée sur Internet. Cet outil est conçu pour aider les utilisateurs à trouver plus des informations pertinents liées à leur recherche courante. L'évolution du profil utilisateur à court terme n'est pas géré par le système vu qu'il est saisi par l'utilisateur en sélectionnant un texte du document en cours de lecture.

Dans [69], le profil à court terme (représenté par un vecteur de termes pondérés) évolue par agrégation de l'information issue à partir des activités de recherche courantes de l'utilisateur. En effet, dans un intervalle de temps précis, un vecteur de termes pondérés V est créé à partir des documents issus des activités de recherche de l'utilisateur, et est comparé à l'ensemble des centres d'intérêts préalablement appris selon une mesure de similarité vectorielle. Le centre d'intérêt qui excède un seuil de similarité est combiné avec V et permet de définir le profil à court terme.

Le profil à court terme dans [165] est identifié pour chaque requête soumise au système comme étant une paire de concepts associé au sujet de la requête. Ce profil varie à chaque soumission d'une requête.

Le profil utilisateur à court terme dans [179] évolue dans une même session de recherche selon une structure appelée, contexte d'usage. Cette évolution repose sur la mise à jour incrémentale de l'historique d'interaction (représenté par une matrice documents-termes) de l'utilisateur après chaque itération feed-

back et l'extraction du contexte d'usage selon un vecteur de termes pondérés. Le résultat de cette évolution est traduite par la repondération des termes du contexte d'usage ou l'apparition de nouveaux termes qui devront contribuer à la définition d'une nouvelle session de recherche.

La gestion de l'évolution du profil conceptuel basé sur une instance d'une ontologie dans [164, 67] est assimilé au processus de sa construction. En effet, l'évolution est prise en compte par classification de l'information récemment visualisée ou créée par l'utilisateur dans les concepts de l'ontologie. Ceci contribue à l'accumulation des poids des concepts initiaux du profil ou l'activation de nouveaux concepts à considérer dans la représentation du profil.

2.4.5.2 Évolution du profil utilisateur à long terme

Le profil utilisateur à long terme modélise des centres d'intérêts généraux, persistants, ou récurrents. Ce profil peut être exploitable dans le but d'améliorer la recherche pour toute requête soumise par l'utilisateur.

Les premiers systèmes permettant de s'adapter aux centres d'intérêts à long terme sont les systèmes de filtrage d'information tels que Grouplens[107]. Plusieurs systèmes développés en RI personnalisée modélisent un profil utilisateur à long terme propre à chaque utilisateur. Parmi ces systèmes, WebPersonae [69, 165, 123, 180] ainsi que les moteurs de recherche sur Internet *Google's Alerts*, et *Google's personalized search 1.1* et *Yahoo My Web*.

Dans les systèmes où le profil utilisateur est spécifié selon un mode d'acquisition explicite des métadonnées par l'utilisateur, l'évolution de ce profil consiste à ajouter explicitement des domaines d'intérêt ou de les supprimer. *Google's Alerts* et la première version de *Google's personalized search 1.1* gèrent des profils utilisateurs génériques incluant les domaines de recherche stables gérés par les utilisateurs.

L'acquisition implicite des données utilisateurs pour l'évolution du profil utilisateur à long terme se fait selon des approches proactives ou réactives. Dans les approches proactives [183, 69], les données sont collectées et traitées dans une étape préliminaire, souvent basée sur une analyse des relations entre requêtes et pages web visités par l'utilisateur. Dans certaines approches réactives [67, 164, 165, 180], la collecte des données est conjointement liée à des mécanismes de délimitation des sessions de recherche dans le cas où le profil à long terme est construit par agrégation des profils appris à court terme [69, 165, 180]. Sinon la collecte des données est indépendante de tout intervalle de temps et est faite durant les interactions de l'utilisateur avec le système [123].

A la différence de prototypes stables de *Google et Yahoo*, la version récente de *Google personnalisé*⁶, appelé "*My Search History*" et son successeur *My Web* introduit par *Yahoo* ainsi que l'approche de personnalisation proposée dans [181], définissent un profil utilisateur à long terme et évolutif à la fois

⁶<http://www.google.com/psearch>

représenté par l'historique de recherche de l'utilisateur. L'évolution de ce profil se fait par accumulation des pages web visitées (*Google personalized search* et *My Web*) ou les requêtes et les pages web cliqués [181] au cours des interactions de l'utilisateur avec le système.

Dans le cadre des approches de construction du profil par accumulation des profils appris à court terme [179, 165], l'évolution du profil utilisateur à long terme consiste à ajouter ou modifier un profil préalablement appris selon des changements éventuels des centres d'intérêts de l'utilisateur au cours des sessions de recherche.

L'approche dans [179] détecte les changements des sujets de recherche en comparant deux vecteurs de termes, appelés contextes d'usage, issus de deux itérations de recherche successifs. La non-corrélation de termes entre ces deux vecteurs indique un éventuel changement des centres d'intérêts. Deux cas peuvent être envisagés : si le nouveau contexte d'usage excède un seuil de similarité avec un centre préalablement appris dans le profil à long terme, le contexte traite un besoin en information récurrent et est combiné avec le vecteur de termes représentant ce centre d'intérêt. Si aucun centre d'intérêt n'est similaire au contexte d'usage courant, celui-ci se considère comme un nouveau centre d'intérêt qui s'ajoute au profil à long terme.

De même, l'évolution du profil conceptuel à long terme dans [165] consiste à ajouter ou modifier un contexte formé d'une paire de concepts associé à une requête de l'utilisateur. L'identification d'un contexte à long terme similaire au contexte de la requête, implique sa combinaison avec celui-ci et la mise à jour éventuelle de sa représentation finale dans le profil à long terme. Un nouveau contexte est ajouté au profil si aucun contexte préalablement appris n'est similaire au contexte de la requête en cours d'évaluation.

Dans une représentation connexionniste du profil utilisateur [22, 129], la gestion de l'évolution du profil s'apparente au processus de construction du profil et implique une mise à jour du réseau de termes. Le système fait évoluer le profil en associant les termes extraits des nouveaux documents collectés à des nœuds existants ou à des nouveaux nœuds dans le réseau en mettant à jour les relations entre ces nœuds.

2.4.5.3 Approches de délimitation des sessions de recherche

L'évolution du profil utilisateur à court terme nécessite des techniques d'identification et de collecte des informations utiles et fortement liées aux activités de recherche courantes de l'utilisateur. Ces techniques se basent souvent sur des mécanismes de délimitation des sessions de recherche définies par un intervalle de temps [173] ou une séquence de requêtes liées à un même besoin en information [180, 163]. D'après [80], une session de recherche est définie

par un groupe de requêtes soumises par un même utilisateur pour une même tâche de recherche. Sur un intervalle de temps, un utilisateur peut faire une ou plusieurs sessions de recherche. Dans le but de clustériser les sessions de recherche, plusieurs approches ont été introduites dans la littérature. Ces approches peuvent être classifiées en trois catégories : les approches basé-temps, les approches basé-contenu et les approches sémantiques.

Les approches basé temps

Les premières approches de clustérisation des sessions de recherche sont basées sur la spécification d'un intervalle de temps moyen pour une session, appelé *Timeout* [77]. Dans ce type d'approches, la session est définie par une séquence de requêtes telle que l'intervalle de temps séparant deux requêtes successives ne dépasse pas un certain seuil. L'analyse est faite sur deux fichiers logs et montre qu'un intervalle de temps entre 10 et 15 minutes est identifié comme le seuil optimal d'identification des sessions de recherche basé temps. Cette méthode souffre du problème de la spécification du meilleur intervalle de temps pour identifier une session. En effet, des utilisateurs différents peuvent avoir des comportements de navigation différents et l'intervalle de temps représentant le seuil d'identification des sessions peut être significativement différent. De même, cet intervalle peut varier entre les sessions de recherche pour un même utilisateur. D'autres approches dédiées à l'analyse des fichiers log des moteurs de recherche identifient les sessions par regroupement des données des utilisateurs sur la base de l'adresse IP, les *cookies* et un intervalle de temps optimal [89].

Une méthode d'identification transactionnelle appelé "reference length", est proposé dans [49]. Cette méthode assume que le temps de lecture d'une page est corrélé au fait que la page est une page de contenu ("content") qui intéresse l'utilisateur ou une page auxiliaire. Une nouvelle session est détectée à chaque détection d'une page "contenu". La limitation de cette méthode réside par le fait qu'un utilisateur peut s'intéresser à plus qu'une page "contenu" pour un même but de recherche.

Une autre méthode d'identification des sessions, appelée *maximal forward reference*, a été proposée dans [42]. Dans cette méthode, une session est définie sur un intervalle de temps par un ensemble de pages agrégées à partir de la première page visitée par l'utilisateur pour une séquence de requêtes jusqu'à ce qu'une page soit revisitée dans la session. La limite de cette méthode est qu'elle traite une session par un ensemble de pages durant un intervalle de temps sans considérer la séquence des *clics* sur les pages visitées. L'approche de clustérisation des sessions proposée dans [193] se base sur le principe de l'alignement séquentiel et prend en compte l'ordre des pages visitées dans une session dans le calcul de similarité des sessions.

Les approches basé-contenu

Ces approches sont basées sur des mesures de similarité textuelle qui se catégorisent

en des mesures basé mots clés ou phrases ou alors des mesures basées sur la distance d'édition des chaînes de caractères entre deux requêtes successives [196]. Les mesures basées mots clés consistent à calculer le nombre de termes présents en commun entre deux requêtes successives p et q . Cette similarité est définie par la formule suivante :

$$similarity_{keyword}(p, q) = KN(p, q) / Max(kn(p), kn(q)) \quad (2.5)$$

où $kn(.)$ est le nombre de termes présents dans une requête, $kn(p, q)$ est le nombre de termes présents simultanément dans les deux requêtes p et q . Des dérivations de cette formule consistent à calculer une similarité où les termes de la requête sont pondérés et peut être étendue pour calculer une similarité plus précise entre les requêtes basée sur les phrases plutôt que des termes uniques. Dans cette approche étendue, l'unité élémentaire représentant partiellement la requête n'est plus un terme mais un ensemble de termes groupés selon des règles syntaxiques. Ceci augmente la similarité entre deux requêtes ayant une phrase en commun en réduisant le nombre des unités élémentaires (phrase) différentes. D'autres mesures consistent à calculer la distance d'édition des chaînes de caractères entre deux requêtes successives [73]. Cette mesure est inversement proportionnelle au nombre d'édérations nécessaires (insertion, suppression, etc.) à unifier deux chaînes de caractères (requêtes). La similarité entre deux requêtes p et q , est calculée selon la formule suivante :

$$similarity_{edit}(p, q) = 1 - EditDistance(p, q) \quad (2.6)$$

Les approches sémantiques

Ces approches sont basées sur des mesures de similarité sémantiques qui se catégorisent en des mesures basées sur le *feedback* utilisateur [196] et des mesures basées sur l'information mutuelle [174].

Les mesures basées sur le *feedback* utilisateur [196] consistent à calculer le nombre de pages visitées en commun pour deux requêtes successives. L'intuition derrière cette mesure est que deux requêtes ayant des documents en commun visitées par l'utilisateur partagent le même sujet. Cette mesure permet de grouper des requêtes sémantiquement liées dans une même session. La mesure de similarité entre deux requêtes p et q , est calculée selon la formule suivante :

$$similarity_{feedback}(p, q) = RD(p, q) / Max(rd(p), rd(q)) \quad (2.7)$$

où RD est le nombre de documents *cliqués* communs entre les deux requêtes, $rd(p)$ est le nombre de documents *cliqués* pour une requête p .

Une mesure de similarité plus élaborée dérivée de la mesure précédente est proposée dans [196] et a pour but d'intégrer en plus du *feedback* utilisateur, une distance conceptuelle entre les documents cliqués communs entre deux requêtes. La distance conceptuelle entre deux documents est calculée sur la base d'une hiérarchie de concepts (Encarta) dans laquelle chaque document de

la collection est classifié dans le concept correspondant. Le système proposé dans [174] intègre une mesure de similarité sémantique qui consiste à calculer le nombre de documents indexés par les termes provenant des deux requêtes successives. Le but dans cette étude est de développer un SRI basé-session où le contexte est représenté par l'ensemble de requêtes et ses résultats associés dans une même session de recherche.

2.4.5.4 Synthèse des approches de modélisation du profil utilisateur

Nous présentons dans cette section une synthèse des approches de construction et d'évolution du profil utilisateur abordées précédemment. Ces approches sont groupées dans le tableau 2.2 selon le modèle de représentation du profil, le mode de construction du profil, son mode d'évolution ainsi que des exemples de systèmes correspondants.

Représentation	Mode de construction	Mode d'évolution	Exemple
Historique de recherche	Accumulation des pages Web sélectionnées au cours des sessions de recherche	évolution à long terme assimilable au mode de construction	Google psearch ⁷ , Yahoo Y!Q [110]
Ensembliste	Classification non supervisée des pages Web visitées ou pertinents dans des classes/centres d'intérêts	Le profil est construit de plusieurs centres d'intérêts en mode proactif sans subir une évolution	Gowan [69]
	Construction d'une matrice document-terme et extraction du centre d'intérêt à court terme selon un vecteur de termes pondérés	agrégation des centres d'intérêts appris au cours des sessions de recherche en mode réactif	[179]
Connexioniste	construction d'un réseau de termes reliés par des arcs [129, 22] ou des connecteurs logiques [109] selon des relations de co-occurrence	augmentation du profil par ajout de noeuds et des arcs issus de nouveaux documents sélectionnés par l'utilisateur	Wifs [129], Ifweb [22],[109]

⁷<http://www.google.com/psearch>

	Extraction d'un graphe de requêtes-documents à l'intermédiaire des sessions de recherche	Le processus d'évolution est similaire au processus de construction par ajout des relations de corrélations pondérées entre les requêtes-documents	[149]
Conceptuelle	Classification supervisée des documents dans des ontologies de domaines prédéfinies	évolution assimilable au mode de construction avec accumulation des poids des concepts initiaux	[164], Obiwan [67]
	Classification hiérarchique non supervisée des documents dans des concepts reliés par des relations de généralité /spécificité	L'évolution du profil est similaire à sa construction avec un ensemble de nouveaux documents collectés en entrée	[102]
	construire le contexte de la requête en l'associant à une paire de concepts de l'ontologie de l'ODP	agrégation et mise à jour des contextes des requêtes au cours des sessions de recherche	[165]

TAB. 2.2: Synthèse des différents modèles de représentation du profil utilisateur

2.4.6 Les modèles d'accès personnalisé à l'information

La personnalisation du processus d'accès à l'information consiste à intégrer ou exploiter le profil utilisateur dans la chaîne d'accès à l'information. Le but fondamental des modèles d'accès personnalisé à l'information est de restituer, en haut de la liste des résultats, des documents qui intéressent l'utilisateur dans sa recherche, en d'autres termes qui semblent les plus similaires à son profil. Nous présentons dans cette section les différentes approches qui consistent à intégrer le profil dans au moins l'une des phases du processus d'accès à l'information telles que la reformulation de requête, l'appariement requête-document et l'ordonnancement des résultats de recherche.

combinaison interpolée des trois modèles de langages comme suit :

$$p(w/\theta_k) = \alpha * p(w/Q_k) + (1 - \alpha) * p(w/H_Q) + \beta * p(w/H_c) \quad (2.8)$$

où $\alpha \in [0 \ 1]$ et $\beta \in [0 \ 1]$.

D'autres approches en RI personnalisée tels que le système de recherche ARCH [165] identifie le profil adéquat à la requête comme étant une paire de concepts issus d'une ontologie et l'exploite dans la reformulation de la requête dans le but d'obtenir une requête plus riche ciblant plus de documents pertinents d'une part et permettant de lever l'ambiguïté au niveau du sens des mots de requête d'autre part. La reformulation de requête est faite en appliquant l'algorithme de Rocchio comme suit :

$$Q2 = \alpha.Q1 + \beta. \sum T_{sel} - \gamma T_{desel} \quad (2.9)$$

Dans cette formule, T_{rel} est un vecteur de termes représentant le concept pertinent et adéquat à la requête et T_{Nrel} est un vecteur de termes représentant le concept non pertinent sélectionnés par l'utilisateur. Les facteurs α, β, γ sont des poids associés respectivement à la requête originale, au concept pertinent et au concept non pertinent.

L'exploitation d'un profil connexionniste dans la reformulation de requêtes est adopté dans [109]. La reformulation de la requête consiste à appliquer toutes les écritures possibles via les arcs du profil (conjonction, substitution et autres). Il est à noter que les arcs sont pondérés de façon à traduire un degré de corrélation entre les termes. Étant donnée une requête, le profil utilisateur, et un seuil de corrélation des arcs reliant les termes, un processus de reformulation de requête génère une nouvelle requête en appliquant toutes les écritures possibles définies par les arcs du profil et satisfaisants le critère du seuil de corrélation considéré.

Une autre approche de reformulation de la requête [51] exploite le profil utilisateur représenté par un réseau de termes dans le but d'identifier des termes fortement corrélés aux termes des requêtes par l'intermédiaire des sessions de recherche. En effet, l'ensemble des termes de la requête est relié à l'ensemble des termes extraits des documents associés par l'intermédiaire des sessions de recherche où une session est définie par une requête et l'ensemble des documents que l'utilisateur a sélectionnés pour cette requête. Chaque session est identifiée comme suit :

$$\text{session} := (\text{requête}, [\text{document sélectionné}]^*)$$

La reformulation d'une requête consiste à sélectionner des termes candidats à partir de tous les termes des documents appartenant aux sessions de la requête.

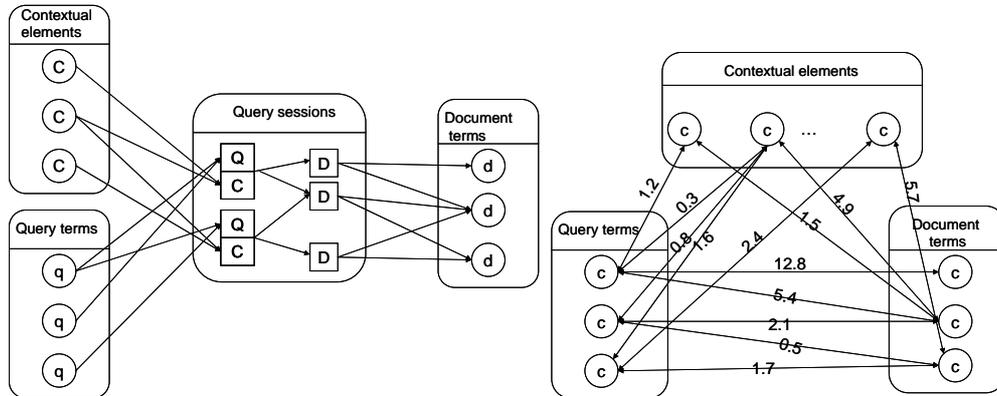


FIG. 2.9 – Les sessions de requêtes sont à la base des corrélations entre les termes des requêtes, les éléments contextuels et les termes des documents

La sélection de n termes candidats à l’expansion de la requête est basée sur des relations de co-occurrence entre chaque terme du document et tous les termes de la requête comme suit :

$$CoWeight_Q(w_j^{(d)}) = \ln\left(\prod_{w_t^{(q)} \in Q} (P(w_j^{(d)}|w_t^{(q)}) + 1)\right) \quad (2.10)$$

Où $P(w_j^{(d)}|w_t^{(q)})$ est la probabilité conditionnelle mesurant le degré de corrélation entre chaque terme du document $w_j^{(d)}$ et chaque terme requête $w_t^{(q)}$. Ceci permet d’identifier les termes des documents qui répondent fréquemment pour une même requête et sont considérés comme candidats au processus de reformulation de la requête.

En plus des relations de corrélations entre les termes des requêtes et ceux des documents, l’approche dans [149] propose d’ajouter des relations de corrélations entre les termes du profil ou éléments contextuels avec les termes des requêtes dans une même session où la session est définie par le triplet requête, contexte et documents associés. La figure 2.9 montre un exemple illustratif du modèle du profil utilisé.

Sur la base de ces corrélations, la reformulation de requête consiste à étendre la requêtes avec des termes issus des termes des documents selon quatre modèles proposés : un modèle orienté-contexte, un modèle indépendant requête-contexte, un modèle orienté requête-contexte et un modèle filtrage orienté-contexte. Le modèle orienté requête-contexte est introduit pour prendre en compte les relations de dépendance entre la requête et le contexte dans la sélection de termes d’expansion. Ce modèle est défini par :

$$M_3(d \triangleleft Q, C \triangleright) =$$

$$= I(d, \langle Q, C \rangle) = \sum_i I(d, c_i) + \sum_j I(d, q_j) + \sum_{ij} I(d, \langle q_j, c_i \rangle)$$

où $\sum_{ij} I(d, \langle q_j, c_i \rangle)$ est l'information mutuelle entre un terme du document et une paire requête-contexte. C'est ce facteur qui introduit une dépendance requête-contexte. Lorsqu'une nouvelle requête est soumise avec un contexte donné, une liste des termes issus des documents corrélés est sélectionnée, puis ordonnée selon les probabilités conditionnelles calculées à travers le modèle. Enfin, les termes les mieux pondérés sont alors utilisés pour reformuler la requête de l'utilisateur.

2.4.6.2 Modèles d'appariement personnalisé requête-document

Les modèles d'appariement personnalisés consistent à exploiter le profil utilisateur dans la fonction de calcul du score du document vis-à-vis une requête. La fonction classique de calcul du score du document se base sur la requête comme la seule ressource d'information qui représente l'utilisateur. Dans le cadre de la RI personnalisée, le calcul du score du document est une fonction qui assigne au document un score de pertinence en fonction non seulement de la requête mais aussi du profil utilisateur. Deux types d'approches d'appariement requête-document existent en RI. On distingue les approches de calcul de pertinence exploitant la structure des hyperliens des pages Web dans un contexte de recherche sur le Web, en se basant sur l'algorithme de PageRank [104] et les approches exploitant le contenu informationnel des documents en se basant sur un modèle de RI par appariement basé mots clés.

2.4.6.2.1 Modèles d'ordonnement personnalisé basé-contenu Il existe peu de systèmes qui fournissent réellement un modèle d'ordonnement personnalisé basé-contenu. Parmi les travaux intéressants qui s'inscrivent dans ce cadre, nous citons les modèles probabilistes d'analyse sémantique latente (PLSA) et le modèle bayésien personnalisé [203].

PLSA est une extension probabiliste à l'analyse sémantique latente (LSA) [78]. Les modèles PLSA se basent sur le principe de corrélation sémantique entre des ensembles d'objets. En RI, l'intégration du profil dans ces modèles revient à résoudre un problème d'ordonnement décisionnel basé sur l'apprentissage d'une fonction de calcul de pertinence d'un ensemble de pages web, étant donné une requête et un modèle d'utilisateur. Le modèle d'un utilisateur u consiste en l'ensemble des documents p^* sélectionnés pour l'ensemble des requêtes sou- mises, et les requêtes soumises q^* , et où l'utilisateur $u \in U\{u_1, u_2, \dots, u_n\}$, la requête $q \in Q\{q_1, q_2, \dots, q_m\}$ et les pages associés $p \in P\{p_1, p_2, \dots, p_l\}$. Pour une requête déjà soumise, le système doit retrouver les pages fréquemment sélectionnées dans le profil de l'utilisateur. Si c'est une nouvelle requête, le

problème revient à calculer un degré de pertinence d'une page p' non existante dans le profil de l'utilisateur et qui a une forte probabilité d'être sélectionnée. Le système se base sur un modèle de probabilité qui calcule la pertinence d'une page p sachant l'utilisateur u et la requête q comme suit :

$$P(p|u, q) = \frac{\sum_{z \in Z} n(u, q, p) P(z|u, q, p)}{\sum_{p'} \sum_{z \in Z} n(u, q, p') P(z|u, q, p')} \quad (2.11)$$

$n(u, q, p)$ est le nombre de fois où l'utilisateur u sélectionne la page p pour la requête q . Puis, le système ordonne les résultats sur la base des valeurs de ces probabilités et retourne les pages ayant les valeurs les plus élevées.

Dans [203], le profil utilisateur représenté par des vecteurs de mots clés représentant les centres d'intérêts de l'utilisateur est intégré dans la fonction d'appariement du modèle bayésien de RI. Plus précisément, L'approche est basée sur l'utilisation des diagrammes d'influence qui permettent de formaliser l'utilité des décisions associées à la pertinence des documents compte tenu de la requête et du profil de l'utilisateur. Plus précisément, la probabilité de pertinence du document d , considérant la requête q et l'utilisateur u , noté $P(d/q, u)$, est calculé comme suit :

$$RSV_u(q, d) = P(q/d, u) P(d/u) \quad (2.12)$$

Ainsi le score de pertinence d'un document D est calculé à travers le diagramme d'influence, noté $ID(D, C, \mu)$, où C est l'ensemble des centres d'intérêts c_1, c_2, \dots, c_n modélisant l'utilisateur u et $\mu = \{\mu_1, \mu_2, \dots, \mu_n\}$, où μ_k exprime l'utilité du document instancié, D pour le centre d'intérêt C_k de l'utilisateur. D'où le calcul du score de pertinence est exprimé par :

$$RSV_U(q, d) = \Psi_{k=1..n}(\mu_k(d, c_k) * p(q/d, c_1, c_2, \dots, c_n)) \quad (2.13)$$

2.4.6.2.2 Modèles d'ordonnement personnalisé basé-hyperliens

Les algorithmes d'ordonnement personnalisés sur le Web consistent à intégrer le profil utilisateur dans le calcul du score de pertinence des pages Web sur la base de la structure des liens entre les pages. Les systèmes adoptant ce type d'approches sont à la base de l'extension des algorithmes les plus connus, tels que les mesures Hub et Authority proposées dans [104] (également nommé : Hyperlinked Induced Topic Search (HITS)), et aussi le PageRank qui constitue une composante importante dans l'algorithme de recherche utilisé par Google [31]. Brièvement, le PageRank (PR) est un vote assigné à une page A collectée de toutes les pages web T_1, \dots, T_n qui se dirigent vers la page A. Il représente

l'importance de la page dirigée. Le PageRank de la page A est calculé par la formule suivante :

$$PR(A) = (1 - d) + d \left[\frac{PR(T_1)}{C(T_1)} + \dots + \frac{PR(T_n)}{C(T_n)} \right] \quad (2.14)$$

où d est un facteur d'atténuation qui prend des valeurs entre 0 et 1 et $C(T_n)$ est défini comme le nombre de liens dans la page T_n . Les scores PR sont des degrés d'importance estimés *a priori* pour toutes les pages Web indépendamment de la requête utilisateur. Parmi les systèmes en RI personnalisée basé sur ce type d'algorithmes, nous citons PROS [44], [61] et [75].

Chirita et al. [44] propose une plateforme d'ordonnement personnalisé, appelé PROS qui intègre un ordonnancement personnalisé des pages web en fonction des profils utilisateurs, construits automatiquement sur la base des annotations et un ensemble des pages Web fréquemment visitées. Dans cette plateforme, les pages Web jugées les plus intéressantes pour l'utilisateur font l'entrée du module nommé *HubFinder*. Ce module, étant l'extension de l'algorithme *Kleinberg's HITS* [104], consiste à collecter des pages web, appelée pages *hub*, corrélés aux centres d'intérêts de l'utilisateur. Une page *hub* est une page qui contient des liens pointant vers d'autres ressources d'information. Les pages Web collectées par *HubFinder* sont ensuite traitées par un autre algorithme, appelé *HubRank*, qui combine les scores PR et les valeurs *hub* des pages web dans le but d'augmenter la liste des résultats fournis par *hubRank*. La liste finale des résultats est traitée par une version personnalisée de l'algorithme de *PageRank* qui consiste à réordonner les résultats à chaque soumission d'une nouvelle requête par l'utilisateur.

Dans le même sens, l'ordonnement personnalisé des pages Web dans [75] se fait grâce à la mise en place d'une dérivée de l'algorithme de PageRank, appelé *topic-sensitive PageRank* (TPSR) sensitive aux domaines d'intérêts informationnels. Dans cet algorithme, l'importance de la page Web se calcule en adaptant les scores de PageRank à certains domaines. Vu que les pages considérées importantes dans certains domaines, peuvent ne pas l'être dans d'autres, l'algorithme calcule 16 ensembles de scores de *PageRank* associés aux domaines, chacun correspond aux *URLs* des catégories de haut niveau de l'ODP. Chaque nouvelle requête soumise est associée aux catégories prises en compte de l'ODP et le score de chaque page résultat n'est plus fonction du *PageRank* original mais d'une combinaison linéaire des scores de PageRank des domaines pondérés par les poids de similitude avec la requête. Ainsi, le score de TPSR de la page v en considérant t est défini par :

$$TSPR_t(v) = d * \sum_{w \in A_v} TSPR_t(w)/l_w + (1 - d) * E_t(v) \quad (2.15)$$

Où v est la page concernée, A_v l'ensemble des pages liées à v , l_w le nombre de liens sortant de la page w , d la probabilité que l'utilisateur suit le lien sortant et $1 - d$ la probabilité restante du saut de page et $E_t(v) = 1/v_t$ où v_t est le nombre total de pages reliées au domaine t . Cet algorithme offre un temps d'exécution comparable à celui de PageRank vu que les ensembles de valeurs de PageRank des domaines sont calculées avant toute interaction avec le système.

2.4.6.3 Modèle d'ordonnement personnalisé basé-contenu

Les modèles d'ordonnement personnalisé basé-contenu consistent à exploiter le profil utilisateur dans le calcul des rangs finaux des résultats de recherche restitués par le système pour une requête donnée. L'ordre final des documents à retourner se calcule souvent par combinaison des rangs/scores originaux des documents et de leurs rangs/scores personnalisés résultant de leur similitude avec le profil utilisateur.

Plusieurs approches en RI personnalisé adoptent ce type de personnalisation [67, 164, 123]. Dans l'approche présentée dans [69], le réordonnement des résultats de recherche consiste à combiner le score de similarité entre le document et le centre d'intérêt courant, représenté par un vecteur de termes pondérés, avec le score d'appariement original du document.

La personnalisation du processus de recherche dans [190] consiste à réordonner les résultats de recherche en ajoutant au score original du document un score dit "score de pertinence personnalisé". Ce dernier est calculé en appliquant une mesure de similarité vectorielle basé sur le *cosinus* entre le document et le profil représentés dans l'espace des concepts de l'ontologie. Par opposition à cette dernière approche, un nombre limité de concepts du profil sont considérés dans le réordonnement des résultats de recherche dans [67, 164] par combinaison de rangs ou de scores. En effet, pour chaque document résultat, le système identifie les premiers N concepts les plus similaires à ce document et calcule pour chacun un score personnalisé basé sur la mesure de similarité *cosinus* avec le document. Le score personnalisé du document sera calculé comme étant la somme de ses scores de similarités avec les N concepts identifiés et du poids des concepts dans le profil comme suit :

$$\text{ConceptualScore} = \sum_{i=1..N} \text{score}(c_i) * \cos(\vec{d}, \vec{c}_i) \quad (2.16)$$

Le rang final d'un document est calculé par combinaison de son rang original et de son rang personnalisé selon la formule suivante :

$$\text{FinalRank} = \alpha * \text{ConceptualRank} + (1 - \alpha) * \text{KeywordRank} \quad (2.17)$$

La personnalisation du processus de recherche dans [123] consiste à exploiter le profil utilisateur représenté par un ensemble de concepts de l'ontologie de

l'ODP dans le réordonnement des résultats de recherche. Étant donné une requête utilisateur, l'algorithme identifie N catégories associées du profil selon une mesure de similarité entre la requête q et chacune des catégories C_j selon la formule suivante :

$$simC_j = sim(q, C_j) = \sum_{d_i \in kNN} cos(q, d_j) * DC(i, j) \quad (2.18)$$

où kNN est un ensemble de documents les plus similaires à la requête retrouvées selon la méthode *k-nearest neighbors* et $DC(i, j) \in \{0, 1\}$ désigne si le document d_i est relié à la catégorie C_j . Sur la base de l'ensemble de N catégories associées à la requête, N listes de résultats sont produites pour chaque concept du profil et pour le même ensemble de résultats restitués par le système. Le rang d'un document dans une liste dépend de sa similarité avec le concept représentant la liste. Un schéma de fusion des listes des résultats est ainsi nécessaire afin de produire une liste unique. Ce schéma de fusion est à la base de la combinaison de vote associé à chaque résultat de recherche dans plusieurs listes. Le vote pour un document d_i ayant un rang i dans une liste L_j est calculé comme suit :

$$vote(d_i, L_j) = W_j * (MM - i + 1) \quad (2.19)$$

où MM est la longueur maximale des listes des résultats obtenus, W_j est le poids de la liste L_j calculée selon la formule suivante :

$$W_j = rankC_j * \sqrt{simC_j * numC_j} \quad (2.20)$$

et où $rankC_j$ est le rang de la catégorie associée à la liste L_j , $simC_j$ est le score de similarité de la catégorie C_j avec une requête q et $numC_j$ est le nombre de documents retrouvés dans la liste associée à la catégorie C_j . Finalement, si un résultat de recherche apparaît dans plusieurs listes de concepts, son vote sera égal à la somme des votes dans chaque liste. Des variantes des approches de réordonnement des résultats consistent en une catégorisation personnalisée [124] basée sur la classification des résultats de recherche dans des catégories représentant les centres d'intérêts du profil utilisateur.

2.4.6.4 Synthèse des modèles d'accès personnalisé à l'information

Nous présentons dans cette section une synthèse des modèles d'accès personnalisé à l'information abordées précédemment. Ces approches sont groupées dans le tableau 2.3, catégorisées selon la phase du processus d'accès personnalisé à l'information, le modèle de représentation du profil utilisé, la technique de personnalisation et des exemples de systèmes correspondants.

Méthode	Profil associé	Technique	Exemple
Reformulation de requête	Une paire de concepts associé à la requête	Ajout ou repondération des termes selon Rocchio	[165]
	Historique de requêtes H_Q et historique de clics H_C	Repondération des termes de la requête selon des modèles de langage statistiques associés à H_C et H_Q	[181]
	Graphe de termes connectés par des opérateurs logiques	Ajout de termes reliés aux termes de la requête par des opérateurs logiques ayant les poids les plus élevés	[109]
	Graphe de requêtes, documents et contextes liés à travers les sessions	Ajout de termes issus de documents occurrents dans une même session selon un modèle de sélection orienté requête-contexte	[149]
Appariement requête-document	Centres d'intérêts multiples représentés par des vecteurs de termes pondérés	Intégration dans le calcul de pertinence du document du modèle de recherche bayésien	[180]
	Domaines d'intérêts de l'ODP issus de des premiers niveaux	Combinaison des <i>PageRanks</i> (TSPR) d'un document calculés relativement aux domaines d'intérêts d'une ontologie de domaine prédéfinie (l'ODP)	[75]
Présentation des résultats	Liste de concepts d'une ontologie de domaines	Combinaison du score/rang original et personnalisé du document telle que le score personnalisé est calculé selon la mesure cosinus entre le document et les concepts du profil	[164, 67]

	Un ensemble de concepts/catégories associé à la requête	Fusion des listes de résultats associées aux catégories/concepts de la requête	[123]
--	---	--	-------

TAB. 2.3: Techniques d'accès personnalisé à l'information

2.5 Conclusion

Nous avons présenté au cours de ce chapitre l'émergence de la RI contextuelle en abordant tout d'abord les fondements de la RI classique et ses limitations en présence du contexte, l'évolution de la RI classique à la RI adaptative jusqu'à l'orientation des travaux vers la RI contextuelle. En RI contextuelle, nous avons présenté les travaux portant sur l'accès contextuel à l'information guidé par le contexte de la tâche de recherche (en tant que informationnel, navigationnel ou transactionnel) et l'accès contextuel à l'information guidé par le profil utilisateur. Pour chacune de ces deux pistes de recherches, nous avons développé les techniques de construction et d'exploitation du contexte dans la chaîne d'accès à l'information.

Nous pouvons constater que les défis majeurs pour faire asseoir une personnalisation efficace dépendent du modèle de représentation du profil utilisateur, des mécanismes de dérivation et d'évolution du profil utilisateur au cours du temps. Ces éléments sont à la base de la différence de performance des SRI personnalisés. Compte tenu de ces éléments, l'évaluation de l'efficacité de la recherche personnalisée est aussi importante que l'évaluation de la qualité du profil appris par le système. Nous présentons dans le chapitre suivant l'émergence des cadres d'évaluation orientés-contexte dédiés à l'évaluation des SRI personnalisés.

Chapitre 3

Accès personnalisé à l'information : de l'évaluation orientée-laboratoire à l'évaluation orientée-contexte

3.1 Introduction

L'évaluation d'un SRI consiste à mesurer ses performances et estimer sa capacité à répondre aux besoins en information des utilisateurs. La performance ou la qualité d'un SRI est mesurée en comparant les réponses du système renvoyés à l'utilisateur pour une requête donnée, aux réponses idéales que l'utilisateur espère recevoir.

Les premières approches d'évaluation des SRI sont basées sur une approche de type laboratoire où (*laboratory-based model*) initiée par *Cleverdon* [45] dans le cadre du projet *Cranfield project II*. Cette approche fournit des ressources de base pour l'évaluation d'un SRI, notamment une collection de requêtes, une collection de documents et des jugements de pertinence associés à chaque requête. Ce modèle est adopté dans les campagnes d'évaluation des SRI tels que TREC, INEX, etc.

L'orientation des modèles de RI vers une nouvelle génération des moteurs de recherche basés sur la RI contextuelle a remis en cause le modèle d'évaluation de *Cranfield* [177]. En effet, les limites de ce modèle d'évaluation sont principalement liées à la non implication du contexte de recherche, de l'utilisateur et ses interactions dans le protocole d'évaluation des systèmes de RI. Ces limites ont motivé les réflexions autour de méthodologies d'évaluation intégrant l'utilisateur dans le processus d'évaluation. Les pistes d'évaluation développées dans

ce sens concernent trois grandes tendances [177] : les initiatives de TREC dans les tâches interactives et HARD, les approches d'évaluation par simulation des contextes ou par utilisation des contextes réels (*user studies*) proposées pour l'évaluation des SRI personnalisés intégrant le profil utilisateur dans le processus d'accès à l'information.

Nous présentons dans ce chapitre un aperçu des différentes approches d'évaluation des SRI. La section 3.2 décrit l'approche d'évaluation classique orientée-laboratoire basée sur le modèle de *Cranfield*. La section 3.3 décrit l'émergence des approches d'évaluation orientées-contexte où nous citons les objectifs de ces approches, les différentes tendances d'évaluation orientée-contexte ainsi que les mesures d'évaluation orientées-contexte. Les sections suivantes détaillent le principe de l'évaluation pour chacune des tendances d'évaluation citée. La section 3.4 décrit les premières tentatives d'évaluation orientées-contexte, notamment les tâches interactive et HARD de TREC. La section 3.5 traite de l'émergence des approches d'évaluation par simulation de contexte, et la section 3.6 décrit le principe de l'évaluation par utilisation des contextes réels (*user studies*). La dernière section conclut le chapitre.

3.2 Les approches d'évaluation des SRI orientées-laboratoire

La performance d'un SRI traduit sa capacité à retourner de l'information pertinente qui répond au besoin en information de l'utilisateur. L'évaluation de la qualité d'un SRI selon les approches orientées-laboratoire consiste à mesurer ses performances sur la base d'une collection de test contrôlée et des métriques d'évaluation standards définis selon des critères d'efficience ou d'efficacité. Les mesures d'évaluation communément utilisées sont essentiellement basées sur le rappel et la précision. Nous présentons dans cette section le paradigme de base des approches d'évaluation orientée-laboratoire, appelé paradigme de *Cranfield*, les différents éléments de la collection de test utilisée dans ces approches ainsi que les limites de ce modèle d'évaluation en présence du contexte de recherche.

3.2.1 Le paradigme de *Cranfield*

L'évaluation des SRI débutait dans les années 1950 où des petites collections de documents (références bibliographiques) ont été utilisées à petite échelle. Dans les années 1960 et 1970, un nouveau paradigme d'évaluation des systèmes de type laboratoire (*laboratory-based model*), appelé paradigme de *Cranfield*, est initié par *Cleverdon* [45] dans le cadre du projet *Cranfield project II*. Le premier

but des expérimentations dans le projet *Cranfield II* est la comparaison entre langages d'indexation basés sur l'utilisation du même ensemble de documents, des besoins/requêtes formalisés pour chaque langage en utilisant les mêmes mesures basées essentiellement sur le rappel et la précision. Ce modèle est défini comme suit [46] : ”*a laboratory type situation where, freed as far as possible from the contamination of operational variables, the performance of index languages could be considered in isolation*”.

Généralement, ce modèle est fondé sur la construction des collections de test volumineuses (e.g. *cranfield*, CACM) servant de base à l'évaluation des systèmes différents et leurs impacts en pratique et la comparaison des différentes techniques. Le principe de construction des collections de test dans le modèle de *Cranfield* est basé sur trois hypothèses majeures :

- hypothèse 1 : la pertinence est estimée par la similarité basée sur le critère de sujet, d'où la pertinence thématique. Cette hypothèse présente les implications suivantes : tous les documents pertinents ont le même degré de préférence du point de vue de l'utilisateur ; la pertinence d'un document est indépendante de la pertinence des autres documents : le besoin en information de l'utilisateur est statique.
- hypothèse 2 : un ensemble unique de jugements de pertinence pour une requête est représentatif d'une population d'utilisateurs, ce qui traduit une notion de pertinence commune.
- hypothèse 3 : la liste des documents pertinents pour chaque requête est complète (tous les documents pertinents sont connus).

3.2.2 Collection de test

La collection de test servant à l'évaluation orientée-laboratoire des SRI comprend un ensemble de requêtes, une collection de documents et des jugements de pertinence associant un sous-ensemble de documents, dits pertinents pour chaque requête. Nous présentons dans cette section les propriétés de chacun de ces éléments, en mettant l'accent sur la campagne d'évaluation TREC comme exemple illustratif, et terminons avec les avantages et les limites de ce modèle d'évaluation.

3.2.2.1 Collection de documents

La collection de documents doit contenir des échantillons des types de textes qui sont considérés intéressants dans un contexte de recherche opérationnel. D'après le modèle de *Cranfield*, les documents de la collection doivent être

représentatifs d'une tâche de recherche réelle satisfaisant les critères suivants :

- Diversité de la collection de test : Il est important que la collection de documents soit représentative dans une tâche de recherche réelle et dispose d'une diversité des sujets, choix des termes, styles de littérature, formats de document, genre, quantité, texte intégral ou résumé du texte, etc. A titre d'exemple, les collections de TREC ad hoc contiennent des articles des journaux (*newspaper or newswire articles*), et des documents gouvernementaux (the Federal Register) dans le but d'intégrer la diversité dans la collection de test.
- Volume de la collection : pour qu'une collection de test soit significative, il faut qu'elle comprend un nombre de documents assez élevé. Les premières collections de test développées dans les années 1970 renferment quelques milliers de documents. Les corpus de test plus récentes (par exemple, ceux de TREC) contiennent en général plus 100 000 documents (considérés maintenant comme une collection de taille moyenne), voir des millions de documents (collection de grande taille).

La spécification du volume des collections de documents utilisés dans l'évaluation orientée-laboratoire est relativement dépendante de la tâche de recherche impliquée dans le SRI à évaluer. Classiquement dans les sciences cognitives, une tâche est définie comme étant un but à atteindre dans un environnement donné au moyen d'actions ou d'opérations [115]. La notion de tâche a été introduite dans plusieurs campagnes d'évaluation telles que TREC, INEX, etc. Dans le cadre de la recherche textuelle simple, la tâche ad hoc est la tâche principale dans TREC qui vise à évaluer les performances des systèmes de recherche d'information, sur des ensembles statiques de documents où seules les requêtes changent. Des tâches spécifiques ont été introduites progressivement dans TREC (depuis TREC4 en 1996) afin de permettre l'évaluation de problèmes spécifiques en recherche d'information telles que le filtrage, le croisement de langues, la recherche dans de très larges collections (25 giga-octets et plus), les modèles d'interaction, etc.

Généralement, les collections TREC contiennent environ 2 gigabytes de textes (entre 500,000 et 1,000,000 documents). Les collections de documents utilisées dans différentes tâches étaient plus petites ou plus grandes dépendant des besoins des tâches et la disponibilité des données. Une nouvelle tâche ad-hoc dans laquelle la collection de documents est un ensemble représentatif de documents issus du *world wide web* est proposée dans le cadre de la collection TREC. Une première tâche permet d'étudier l'impact de l'utilisation des liens hypertextes sur les performances de la recherche d'informations sur approximativement 2 GO de pages Web. Une base de 100 GO de documents Web est également mise à disposition des utilisateurs pour l'évaluation des tâches de recherche sur le Web.

3.2.2.2 Collection de requêtes

La requête traduit le besoin en information de l'utilisateur et est formulée souvent par un ensemble de mots clés ciblant les documents recherchés. L'évaluation d'un système ne doit pas reposer seulement sur une requête. Pour avoir une évaluation assez objective, un ensemble de quelques dizaines de requêtes, traitant des sujets variés, est nécessaire. Vu que l'efficacité d'un système varie considérablement entre les requêtes, plus le nombre de requêtes utilisées dans les expérimentations est élevé, plus les conclusions tirées des expérimentations seront fiables [34].

Dans le cadre des campagnes d'évaluation (TREC, INEX, etc.), le terme *topic* désigne un besoin en information de l'utilisateur dans lequel sont spécifiées des métadonnées supplémentaires concernant une description courte et longue du besoin en information et un ensemble de critères de pertinence des documents répondant au *topic*. Généralement un *topic* comprend 4 sections : un identifiant, un titre, une description et un champ narratif. La requête désigne quant à elle la structure du besoin en information soumise au SRI et peut correspondre au titre du *topic* ou sa description. Un exemple typique d'un *topic* dans TREC est illustré dans la figure 3.1. Le champ *Title* contient une description courte du *topic* et le champ *Description* contient une description riche et plus longue du *topic* concernant les critères de pertinence des documents associés au *topic*. Finalement, le champ *Narrative* décrit des informations additionnelles qui recouvrent les critères de pertinence des documents pertinents associés au *topic* dans un paragraphe. TREC utilise de 25 à 50 *topics* comme une norme

```
<top>
<num> Number: 705
<title> Iraq foreign debt reduction
<desc> Description:
Identify any efforts, proposed or undertaken, by world
governments to seek reduction of Iraq's foreign debt.
<narr> Narrative:
Documents noting this subject as a topic for
discussion (e.g. at U.N. and G7) are relevant. Money
pledged for
reconstruction is irrelevant.
</top>
```

FIG. 3.1 – Topic 705 de la tâche Terabyte de TREC

du nombre de requêtes de test. Les *topics* dans TREC sont créés par le même groupe de personnes qui associent à chaque requête l'ensemble de ses documents pertinents. Souvent, chaque assesseur dans TREC participe en définissant des *topics* selon ses propres intérêts et recherche dans la collection de documents

via le système de recherche PRISE de NIST pour estimer le nombre de documents pertinents pour chaque *topic*. Finalement, la liste des topics est choisie par l'équipe de TREC en se basant sur le nombre de documents pertinents associés à chacun et en balançant les choix entre les assesseurs. Dans le cadre de la recherche informationnelle sur le Web, la tâche *Terabyte* dans TREC, créent les champs *Title* des *topics* pour définir des requêtes stéréo-typiquement soumises par des utilisateurs du Web.

3.2.2.3 Jugements de pertinence

Dans le but de comparer les documents résultats fournis par le système et les documents que souhaite recevoir l'utilisateur, il faut spécifier pour chaque requête l'ensemble de réponses idéales du point de vue de l'utilisateur. La spécification des jugements de pertinence des documents associés à la requête constituent la tâche la plus difficile dans la construction d'une collection de test. A la différence de l'évaluation à la *Cranfield* où les documents pertinents doivent être connus et complets pour chaque requête, la technique adoptée dans la plupart des campagnes d'évaluation pour la création des jugements de pertinence ne respectent pas cette hypothèse.

La création des jugements de pertinence dans l'évaluation orientée-laboratoire est basée sur une technique appelée, *Pooling*. Vu que la notion de pertinence est variable entre les utilisateurs, la technique de *Pooling* consiste à trouver une pertinence commune sur une partie de documents en se basant uniquement sur le critère de sujet ou la notion de pertinence thématique. Dans cette technique, les jugements de pertinence sont souvent spécifiés par un groupe de personnes (assesseurs) experts dans le domaine des sujets traités par les requêtes. Dans le but d'établir les listes de documents pertinents pour chaque requête, les utilisateurs (ou des assesseurs simulant des utilisateurs) doivent examiner chaque document du *pool* de la collection, et juger s'il est pertinent indépendamment des autres documents pertinents contenant la même information.

La figure 3.2 montre les étapes de la technique de *Pooling* pour la création des jugements de pertinence, détaillée comme suit :

- Création des *topics* : les assesseurs créent les *topics* de test servant à l'évaluation des systèmes,
- Collecte des documents pour un topic : cette étape consiste à collecter les 1000 premiers documents retournés par une variété des systèmes pour un même *topic*. Dans le cadre de *TREC*, le *NIST* (National Institute of Standards and Technology) fournit une collection de documents et un ensemble de requêtes définies. Les participants exploitent leurs propres systèmes de récupération des résultats pour chaque *topic*, et renvoient au

- *NIST* la liste des documents associée,
- Création des ensembles *pools* des documents uniques : Il s’agit de créer un ensemble de documents uniques à partir de toutes les listes des documents retournés par les systèmes utilisés puisqu’un même document peut figurer dans plusieurs listes de résultats. Dans *TREC*, le NIST juge en commun les documents recherchés pour l’exactitude, et évalue les résultats en prenant les *k*-documents les mieux classés des différents SRI participant à la campagne d’évaluation.
- Création des jugements de pertinence : la dernière étape de la technique de *Pooling* consiste à juger chaque document dans les *pools* de documents associés à chaque *topic*. Chaque document est montré à des assesseurs qui décident finalement de sa pertinence.

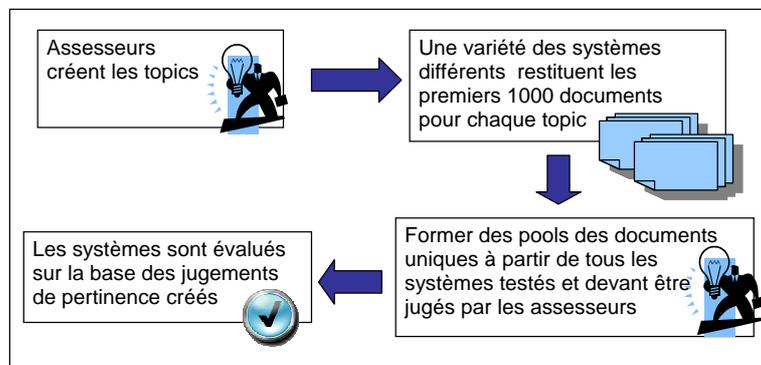


FIG. 3.2 – Technique de *Pooling* pour la création des collections de test larges

Notons que les jugements de pertinence peuvent être binaires (pertinent, non pertinent) ou graduels correspondants à des degrés de pertinence (utilisés dans *Cranfield*) allant de très pertinent (high relevant), moyennement pertinent (soft relevant) ou non pertinent (non relevant).

3.2.3 Mesures d’évaluation

L’évaluation des systèmes peut être abordée sous deux angles : l’efficacité et l’efficacités. L’efficacité regroupe le temps et l’espace [30, 191]. En effet, un système est considéré meilleur lorsque le temps entre la formulation de la requête et la réponse du système est court et l’espace occupé par le système est faible. L’efficacités d’un système peut être mesurée par plusieurs critères :

- L’effort intellectuel ou physique, nécessaire aux utilisateurs pour formuler les requêtes, conduire leur recherche, voir les documents résultats, etc.
- La présentation du résultat peut être mesurée par la capacité de l’utilisateur à utiliser les documents retrouvés,
- La qualité de la collection de documents vis à vis du besoin de l’utilisateur. Ce critère dépend de la disponibilité de l’information pertinente dans la

collection de test,

- La capacité à retrouver tous les documents pertinents pour une requête et à rejeter les documents jugés non pertinents. Cette caractéristique semble être la plus importante.

Les mesures d'évaluation largement adoptées en RI sont des mesures d'évaluation de la capacité du système à retrouver les documents pertinents pour une requête et sont souvent basées sur le rappel et la précision ou des mesures dérivées du rappel et de la précision. Nous citons la précision à X documents noté Précision@n (n peut prendre différentes valeurs : 5 ; 10 ; 15 ; ... ; 1000), la précision moyenne (ou MAP pour Mean Average Precision), mesure F, R-precision, etc.

A. Rappel et précision

- *Rappel* : le rappel mesure la proportion des documents pertinents retrouvés parmi tous les documents pertinents dans la base documentaire. Elle mesure la capacité du système à retrouver tous les documents pertinents répondant à une requête.

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

- *Précision* : la précision mesure la proportion des documents pertinents retrouvés parmi tous les documents retournés par le système. Elle mesure la capacité du système à rejeter tous les documents non pertinents à une requête donnée par le rapport entre l'ensemble des documents pertinents retrouvés et l'ensemble des documents retournés. La précision est calculée selon la formule suivante :

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

Idéalement, on voudrait qu'un système donne de bons taux de précision et de rappel en même temps. En pratique, on ne peut pas avoir un système qui aurait 100% de rappel et de précision (l'algorithme trouve la totalité des documents pertinents (rappel) et ne fait aucune erreur (précision)). Plus souvent, on peut obtenir un taux de précision et de rappel aux alentours de 30%. Notons que les deux métriques ne sont pas indépendantes. Le comportement d'un système peut varier en faveur de la précision ou en faveur du rappel (au détriment de l'autre métrique). Ainsi, pour un système, on a une courbe de précision-rappel qui a en général la forme illustrée dans la figure 3.3.

Les mesures de précision-rappel ne sont pas statiques non plus, (c'est-à-dire qu'un système n'a pas qu'une mesure de précision et de rappel). La performance du système en termes de ces mesures change en fonction de plusieurs facteurs tels que le degré de difficulté/ambiguïté des requêtes de test ou autres.

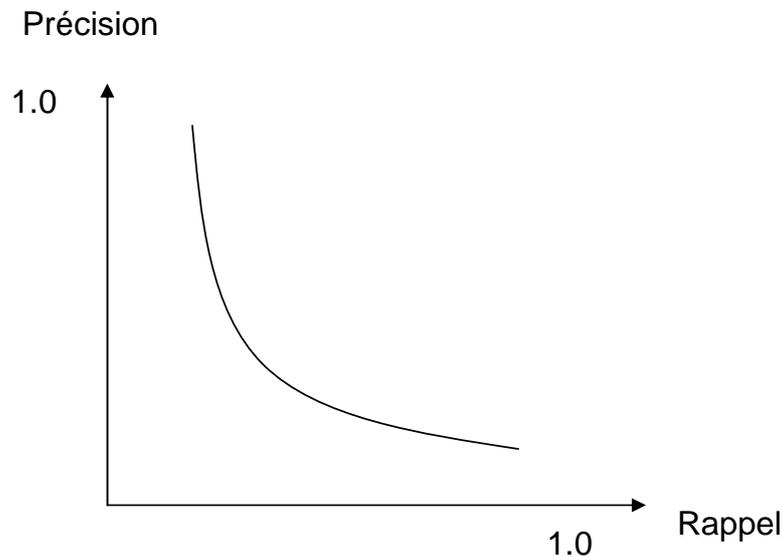


FIG. 3.3 – Forme générale de la courbe de précision-rappel d'un SRI

B. Mesures dérivées de rappel et précision

D'autres mesures d'évaluation sont proposées pour mesurer l'efficacité d'un SRI. Ces mesures peuvent être dérivées des mesures de rappel et précision ou alors orientées-rangs. Nous les résumons comme suit :

- Précision à n (P_n) : c'est la précision à différents niveaux de coupe. Cette précision mesure la proportion des documents pertinents retrouvés parmi les n premiers documents retournés par le système. Elle permet en particulier de s'intéresser à la haute précision, lorsque peu de documents sont restitués (ceci est à rapprocher de l'intérêt des internautes à la seule première page de résultats par exemple). Cette mesure ne prend pas en compte le fait que certaines requêtes comportent peu de documents pertinents.

$$P_n = \frac{\#documents\ pertinents\ retrouvés}{n}$$

- R-precision : Cette précision mesure la proportion des documents pertinents retrouvés après que R documents ont été retrouvés, où R est le nombre de documents pertinents pour la requête considérée. C'est une Variante de la P_n où n est le nombre de documents pertinents pour la requête.
- Précision moyenne non interpolée (MAP) : est la moyenne des précisions moyennes (AP) obtenues sur l'ensemble des requêtes à chaque fois qu'un document pertinent est retrouvé.

$$MAP = \frac{\sum_{q \in Q} AP_q}{|Q|} \tag{3.1}$$

avec AP_q est la précision moyenne d'une requête q , Q est l'ensemble des

requêtes et $|Q|$ 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 P_n .

- précision moyenne interpolée : Cette précision est calculée à différents niveaux de rappel (0%, 10%, 20%, ..., 100%). Pour chaque niveau de rappel, les valeurs calculées sont moyennées sur tout l'ensemble des requêtes.

C. Mesures orientées-rang

- Mean Reciprocal Rank (MRR) : correspond à une mesure alternative de calcul de la précision. Si l'utilisateur analyse la liste des documents, cette mesure permet d'évaluer le nombre de documents qu'il faut considérer avant de retrouver le premier document pertinent. Elle est égale à la moyenne (calculée sur l'ensemble de requêtes) du rang du premier document pertinent.

$$MRR = \frac{1}{|Q|} \sum_{i=1}^Q \frac{1}{rank_i}$$

MRR est nulle pour une requête si aucun document pertinent n'est retourné par le système. Cependant, MRR donne un score élevé pour un système qui retourne des documents pertinents en haut de la liste présentée à l'utilisateur. Cette mesure est souvent utilisée dans les systèmes Questions-réponses où l'utilisateur s'intéresse à recevoir la bonne réponse en premier rang.

Une grande variété des mesures d'évaluation (CG (cumulative gain), DCG (discounted cumulative gain) [91, 90], etc.) ont été développées où certaines sont moins stables que d'autres. Par exemple, les mesures basées sur un peu de données comme la précision au premier document pertinent retrouvé introduisent du bruit, et la moyenne des précisions moyennes (MAP), qui mesure la précision globale du système recouvrant tous les points de rappel, est beaucoup plus stable [191].

3.2.4 Protocole d'évaluation

Le protocole d'évaluation dans le modèle d'évaluation orienté-laboratoire consiste à comparer plusieurs systèmes, stratégies de recherche ou algorithmes entre eux en spécifiant trois composantes non indépendantes qui sont le nombre de *topics* utilisés, les mesures d'évaluation utilisés et la différence de performance requise pour considérer qu'une stratégie de recherche est meilleure qu'une autre [34].

L'évaluation de l'efficacité de chaque stratégie de recherche consiste à évaluer la liste de résultats obtenus pour chaque requête de test. Cette évaluation est à la base de la correspondance entre la pertinence algorithmique calculée par le système et la pertinence donnée par les assesseurs. L'efficacité globale d'une

stratégie de recherche est calculée comme étant la moyenne des précisions calculées selon une mesure donnée sur l'ensemble des requêtes dans la collection de test.

3.2.5 Limites de l'évaluation orientée-laboratoire en présence du contexte

L'évaluation orientée laboratoire d'un SRI est principalement fondée sur l'utilisation d'une collection de test où les requêtes sont les seules ressources clés qui traduisent le besoin en information de l'utilisateur. Ce type d'évaluation révèle des limites pour l'évaluation des SRI contextuels [113, 99]. Ces limites concernent le non réalisme des collections de test et la non-exhaustivité des mesures d'évaluation, détaillées dans une étude [177] effectuée sur les protocoles d'évaluation des SRI en présence du contexte.

3.2.5.1 Limites des collections de test

Les limites liées aux collections de test concerne l'inadéquation de ces collections pour l'évaluation de la recherche d'information en présence du contexte [177, 85]. Plus particulièrement, l'aspect du non-réalisme touche la collection des requêtes et de documents ainsi que des facteurs réels non impliqués dans les jugements de pertinence.

– L'absence du contexte de recherche :

1. Absence des vrais utilisateurs et du contexte de recherche : le modèle d'évaluation classique n'intègre pas des vrais utilisateurs dans des contextes de recherche ou situations précises et les remplace par des assesseurs experts dans le domaine chargés de la création des jugements de pertinence pour chaque *Topic*.
2. Absence de la dynamique et d'interactions : le modèle d'évaluation orienté laboratoire est fait en mode *batch* où les requêtes sont supposées ainsi représenter à elles seules l'utilisateur. Les utilisateurs directs ayant émis ces requêtes, leurs centres d'intérêt et interactions avec le SRI ne figurent pas dans l'évaluation du système. D'où l'absence d'interactions et de dynamique requis pour l'apprentissage du modèle de l'utilisateur dans le cadre de la RI personnalisée.

– La faible représentativité des besoins :

la problématique dans l'évaluation orientée-laboratoire concerne également les requêtes quant à leur représentativité du besoin en information d'une part et la non-incertitude des requêtes d'autre part.

1. Représentativité du besoin en information : le besoin en information

de l'utilisateur est représenté par une seule ressource clé, notamment une requête mots-clés exprimée souvent en langage naturel. Or, c'est tout le problème de la représentation du besoin d'information de l'utilisateur. Les écarts entre ce que nous pensons et ce qui est interprété sont en effet importants.

2. Incertitude des requêtes : les requêtes d'une collection dans l'évaluation orientée-laboratoire sont bien définies et ne traduisent pas l'incertitude lors des tâches ou situations de recherche réelles [27].
- Les limites de la pertinence thématique :
- concernant les jugements de pertinence, les limites sont issues du fait qu'ils sont créés sur la base de l'hypothèse de la pertinence thématique et statique ainsi que de l'indépendance entre les documents pertinents.

1. Pertinence thématique : les jugements de pertinence sont thématiques et indépendants des situations et du contexte de recherche. En outre la pertinence considérée est statique qui ne révèle pas la nature dynamique de la perception de pertinence de l'utilisateur qui change avec la dynamique de son besoin pour un même *topic* d'une situation à une autre [27, 189].

Or, il a été bien montré dans [140, 131] que la notion de pertinence est plus complexe, couvrant des niveaux divers liés à la situation de recherche en cours : pertinence cognitive, pertinence affective, pertinence situationnelle, etc.

2. Indépendance de la pertinence entre documents :

l'hypothèse d'indépendance entre les documents pertinents semblent non réelle vu que certains documents contenant de l'information similaires dans d'autres documents pertinents peuvent être également pertinents [145]. Nous citons l'exemple des hyperliens reliant sémantiquement des pages web dans le cadre de la recherche sur le web ou des citations des références bibliographiques dans le cadre de la recherche dans des bibliothèques digitales.

3.2.5.2 La non-exhaustivité des mesures d'évaluation

Les mesures d'évaluation classiques ne sont pas exhaustives dans le sens où le document est jugé pertinent s'il recouvre le *topic* de la requête indépendamment du contexte et de la tâche de recherche. Nous citons quelques limites de ces mesures comme suit :

- Insuffisance des mesures quantitatives : les mesures de rappel et précision sont des mesures quantitatives considérées insuffisantes pour l'évaluation des SRI en présence du contexte. En effet, l'évaluation par le biais de ces mesures se fait par rapport au nombre de documents retrouvés par le

système. Ces mesures ne permettent pas d'évaluer la qualité du système où figurent les interactions de recherche de l'utilisateur.

- Utilisation de la pertinence binaire : la pertinence binaire est largement adoptée dans l'évaluation de performance des systèmes selon les mesures de rappel et précision. Dans le cadre de la RI contextuelle, la pertinence graduelle est la plus adaptée pour l'évaluation de la performance des systèmes en fonction des facteurs liés à l'environnement de recherche et encore de la disposition d'esprit de la personne ayant réellement besoin de ces documents [28]. La pertinence graduelle n'est pas largement utilisée en RI contextuelle. Pourtant, elle est utilisée dans la plupart des campagnes d'évaluation telles que TREC, CLEF, INEX, etc.

Malgré les critiques associées aux expérimentations effectuées dans *Cranfield II* [171], ce modèle d'évaluation est largement adopté dans les campagnes d'évaluation internationales telles que TREC, INEX¹ et CLEF² [155, 93, 172] vu qu'il présente une base d'évaluation comparative des systèmes tout en exploitant des ressources communes et des mesures d'évaluation standards.

3.3 Vers l'évaluation orientée-contexte pour une tâche d'accès personnalisé à l'information

Nous présentons dans cette section l'émergence de l'évaluation orientée-contexte des SRI personnalisés. Nous nous intéressons à explorer les modalités d'évaluation du contexte cognitif de l'utilisateur ou son profil ainsi que l'évaluation des modèles de RI en présence du contexte. Nous présentons tout d'abord les objectifs de l'évaluation orientée-contexte ainsi que les différentes approches et mesures d'évaluation permettant d'intégrer le contexte de recherche dans le processus d'évaluation.

3.3.1 Objectifs de l'évaluation orientée-contexte

Le but fondamental des approches d'évaluation orientées-contexte est de mettre en place un scénario d'évaluation permettant d'intégrer le contexte de recherche et les interactions de l'utilisateur dans l'évaluation de l'efficacité du système.

L'intégration du contexte de recherche dans les approches d'évaluation du système porte des objectifs répondant aux limites de l'évaluation orienté la-

¹Initiative for the Evaluation of XML retrieval

²Cross Language Evaluation Forum

boratoire. Ces objectifs concernent des aspects liés à la prise en compte du contexte de recherche dans l'approche de l'évaluation, la représentativité des requêtes de test, la contextualisation des jugements de pertinence et celles des mesures d'évaluation.

- Intégration du contexte de recherche dans la collection de test : l'intégration du contexte de recherche comme étant un composante principale de la collection de test en plus des requêtes, documents et des jugements de pertinence permet d'évaluer les SRI dans des situations de recherche réelles et où les interactions de l'utilisateur avec le système et la nature dynamique du besoin en information sont pris en compte dans le scénario d'évaluation. Toutes les approches en RI personnalisée et la tâche HARD de TREC évaluent le SRI en intégrant respectivement le profil utilisateur ou des éléments contextuels associés à la requête dans la collection de test.
- Réalisme des besoins en information : l'utilisation des requêtes formulées par des utilisateurs usagers permettent d'évaluer la réponse du système à des besoins en information naturels et qui se qualifient par l'incertitude lors des tâches de recherche réelles. La tâche interactive de TREC ainsi que quelques approches en RI personnalisée investissent des utilisateurs qui formulent des requêtes traduisant des besoins en information naturels.
- Contextualisation des jugements de pertinence : la création des jugements de pertinence par l'utilisateur qui l'a émis en fonction du contexte associé permet d'évaluer le SRI sur la base de la pertinence perceptionnelle ou alors situationnelle telle qu'elle est vue de l'utilisateur qui a émis la requête. Ceci permet de prendre en compte la dynamique du besoin en information ainsi que celle des jugements de pertinence au cours de la recherche. Cet objectif est atteint dans le cadre de la tâche interactive de TREC ainsi que dans certaines approches en RI personnalisée qui permettent à l'utilisateur qui a émis la requête de juger les documents renvoyés par le système.
- Contextualisation des mesures d'évaluation : l'évaluation des SRI selon des mesures d'évaluation orientées-contextes permet d'atteindre un double objectif. Le premier consiste à estimer la performance du système en fonction du comportement d'interaction de l'utilisateur et de la dynamique besoin en information. Le deuxième consiste à évaluer la capacité de l'utilisateur à accomplir une tâche de recherche selon des mesures appropriées à l'évaluation interactive. La tâche interactive de TREC a principalement introduit des mesures d'évaluation interactive dans ce sens.

3.3.2 Approches d'évaluation orientées-contexte

A nos jours, il n'existe pas un cadre standard pour l'évaluation des modèles d'accès personnalisé à l'information. L'élaboration des approches d'évaluation orientées-contexte sont conçues dans le cadre des approches développées en RI personnalisée.

Vu que les SRI personnalisés consistent en la modélisation du contexte/profil dans une première phase et l'exploitation du profil dans le processus d'accès à l'information dans une deuxième phase, les approches d'évaluation orientées-contexte doivent porter sur l'évaluation de deux composantes complémentaires :

- évaluation de la qualité du profil utilisateur servant comme contexte de recherche,
- évaluation de l'efficacité de la recherche personnalisée intégrant le profil utilisateur dans le processus d'accès à l'information.

Nous présentons dans la suite de cette section le principe de l'évaluation de la qualité du profil abordée dans quelques SRI personnalisés et l'évaluation de l'efficacité de recherche personnalisée selon des approches d'évaluation orientées-contexte.

3.3.2.1 Evaluation de la qualité du profil utilisateur

L'évaluation de la qualité du profil est prise en compte dans l'évaluation des SRI personnalisés vu que la qualité du profil a un impact direct sur la performance de recherche. Cette évaluation n'est pas nécessaire généralement dans les approches qui définissent le contexte par des métadonnées saisies explicitement par l'utilisateur ou dans des approches où l'évaluation est effectuée en utilisant une collection de requêtes associées à des métadonnées servant comme éléments contextuels (la tâche HARD de TREC). Pourtant, l'évaluation de la qualité du profil est requise dans les approches de modélisation implicite du contexte dans le but de valider la précision du contexte construit automatiquement.

La question fondamentale qui se pose au niveau de l'évaluation de la qualité du profil utilisateur est si le modèle de l'utilisateur ou son profil construit automatiquement par le système est une représentation adéquate de ses centres d'intérêts effectifs. Pour cette raison, une étape fondamentale avant l'évaluation de l'efficacité du modèle de RI personnalisée consiste à séparer le modèle de l'utilisateur du processus de RI personnalisée et mesurer sa précision par rapport à des utilisateurs réels.

Dans le but d'évaluer la qualité du profil indépendamment de l'efficacité du modèle de RI, et pour ne pas biaiser les résultats sur une même collection de documents, un découpage de la collection est souvent effectué en deux sous-

collections : une sous-collection pour l'apprentissage du profil utilisateur et une sous-collection pour tester l'efficacité du système. L'évaluation de la qualité du profil peut inclure des tests pour évaluer l'efficacité de l'algorithme d'apprentissage du profil. Ces tests peuvent intégrer des vrais utilisateurs pour juger le profil appris, des mesures comparatives entre plusieurs algorithmes [139] ou des mesures de convergence de l'algorithme [123]. Compte tenu des travaux dans le domaine, nous supposons que :

1. Selon un modèle de représentation conceptuel du profil, la mesure de la qualité du profil consiste à quantifier le degré de correspondance entre les annotations manuelles (faites par des utilisateurs réels) des éléments constitutifs du profil avec la construction automatique de ce profil par le système [123, 173, 56]. Il s'agit dans certains travaux d'investir des utilisateurs qui jugent de la pertinence des concepts du profil [123, 173] ou alors de calculer automatiquement la précision d'un profil (représenté par un ensemble de concepts associés à une requête donnée) comme étant la robustesse de classification des documents annotés préalablement dans les concepts associés.
2. Il n'existe pas des mesures d'évaluation standards. Les mesures d'évaluation proposées sont relativement dépendantes du modèle de représentation du profil.

3.3.2.2 Évaluation de l'efficacité du modèle de RI personnalisée

L'objectif de cette étape est d'évaluer l'impact de l'utilisation du contexte de recherche dans le processus d'accès à l'information. Généralement, l'évaluation se fait en comparant la performance de deux scénarios de recherche :

- la recherche classique effectuée en utilisant seulement la requête,
- la recherche personnalisée effectuée en intégrant le contexte de recherche ou le profil utilisateur dans le modèle d'accès à l'information.

Nous pouvons dégager 3 grandes tendances des approches d'évaluation de l'efficacité des modèles de RI personnalisée orientée-contexte :

- les approches d'évaluation orientées-contexte initiées par TREC (approches numériques comme dans les campagnes traditionnelles),
- les approches d'évaluation par simulation des contextes,
- les approches d'évaluation par utilisation des contextes réels (*user studies*).

3.3.2.2.1 Approches d'évaluation orientées-contexte initiées par TREC

Les premières tentatives effectuées dans le cadre de l'évaluation des SRI en présence du contexte ont été proposées dans TREC à travers les tâches *Interactive* et *HARD*. La tâche interactive de TREC consiste à étudier les interactions de l'utilisateur avec le système en utilisant des mesures d'évaluation spécifiques

dédiées à mesurer qualitativement l'efficacité du système dans une tâche de recherche précise. La tâche HARD de TREC consiste à intégrer le contexte de recherche comme un ensemble de critères de préférences de l'utilisateur aussi bien dans le processus d'accès à l'information que dans l'évaluation du système. Le principe de l'évaluation de ces approches est traité plus tard dans la section 3.4 p.98.

- La tâche Interactive : Cette tâche a été menée dans TREC depuis 1995 (TREC-4) jusqu'en 2002 (TREC-10). L'objectif de cette tâche est d'étudier les modèles d'interaction en RI. Plus précisément, elle a eu comme double objectif le développement de méthodologies appropriées à l'évaluation du processus de RI interactive et la mesure de l'impact des différentes caractéristiques des utilisateurs sur l'évaluation de la pertinence des résultats.
- La tâche HARD (*High Accuracy Retrieval from Document*) : cette tâche a été menée dans TREC depuis 2003 (TREC-12) jusqu'en 2005 (TREC-14). L'objectif de la tâche HARD est de permettre aux systèmes d'atteindre une grande précision de recherche en mettant en place une recherche orientée-contexte. Celle-ci considère le contexte de recherche ou de l'utilisateur comme étant une partie intégrante du processus de recherche et de l'évaluation. L'évaluation du système de recherche dans cette tâche consiste à mesurer sa capacité à fournir les résultats pertinents en fonction du contexte de recherche associé à la requête. Les mesures d'évaluation adoptées considèrent des niveaux de pertinence dans le but de prendre en compte la différence de perception de pertinence de l'information entre les utilisateurs.

3.3.2.2.2 Approches d'évaluation par simulation des contextes

L'évaluation d'un SRI par simulation de contextes intègre un scénario d'évaluation défini par des contextes de recherche et des interactions hypothétiques *a priori* permettant de simuler des utilisateurs [133, 152]. Un scénario représente une situation de recherche qui met en œuvre un utilisateur simulé à travers ses centres d'intérêt [164, 179], ou des interactions hypothétiques avec le SRI [152] à travers des interfaces comme celles proposées dans le cadre de la tâche interactive de TREC. Les pistes d'évaluation par simulation des contextes consistent à simuler un scénario de recherche par une ou plusieurs requêtes liées à un centre d'intérêt de l'utilisateur et exploiter des jugements de pertinence préalablement donnés [179] ou considérés pertinents s'ils sont classifiés dans le domaine d'intérêt simulé [164].

L'évaluation de l'efficacité du SRI dans ce type d'approches consiste à évaluer l'impact de l'addition du contexte de recherche dans le processus de recherche d'information. Cette évaluation est basée sur la comparaison des deux listes de résultats obtenues selon le scénario de recherche classique et personnalisé aux jugements de pertinence hypothétiques. Les jugements de pertinence

sont soit réutilisés à partir d'une collection de type laboratoire [180], ou basés sur des hypothèses de pertinence dépendantes du contexte de recherche simulé [164]. Les mesures d'évaluation dans ce type d'évaluation sont souvent basées sur des mesures classiques dérivées du rappel et de la précision. Le principe de l'évaluation de ces approches est traité plus tard dans la section 3.5 p.104.

3.3.2.2.3 Approches d'évaluation par utilisation des contextes réels

À la différence des approches d'évaluation des SRI personnalisés par simulation de contexte, l'évaluation basée sur l'utilisation des contextes réels (*User studies*) est adoptée dans le but d'intégrer les activités et les interactions des utilisateurs dans le processus d'évaluation des systèmes. L'exploitation des contextes de recherche réels dans ce type d'approches permet de prendre en compte de la nature dynamique du besoin en information et considérer des jugements de pertinence selon la perception de pertinence de l'utilisateur qui a émis la requête dans des situations de recherche réelles et bien spécifiées (*work task situation*) [39].

Dans ces approches, les interactions des utilisateurs (tels que les *clics*, temps passé sur une page, etc.) sont enregistrés dans un fichier log exploité dans en partie pour l'apprentissage du profil utilisateur et dans quelques approches dans l'évaluation de performance du système [173]. Les mesures d'évaluation classiques sont utilisées dans ce type d'approches. En plus, l'approche proposée dans [173] se base sur les données de *clics* sur un document comme un indice de pertinence du document.

Le principe de l'évaluation de ces approches est traité plus tard dans la section 3.6 p.108.

3.3.3 Mesures d'évaluation orientées-contexte

Les mesures d'évaluation classiques des SRI se basent souvent sur le rappel et la précision et leurs dérivées. Ces mesures ne sont pas suffisantes pour l'évaluation de la RI contextuelle et où le contexte de recherche et les interactions utilisateur-système ont un impact sur la perception de pertinence de l'information chez l'utilisateur. L'introduction des mesures d'évaluation orientées contexte concernent l'évaluation de la qualité du contexte d'une part et l'évaluation de l'efficacité de recherche dans le sens où ces mesures soient des mesures qualitatives et non seulement quantitatives tenant en compte des aspects d'interactions et contextuels.

Nous présentons deux types de mesures d'évaluation orientées contextes :

- les mesures d'évaluation de la qualité du profil exploitées dans le cadre

de la RI personnalisée,

- les mesures d'évaluation de l'efficacité du modèle de recherche. Plus particulièrement, nous présentons des mesures d'évaluation orientés-système et des mesures orientées-tâches utilisées dans les SRI interactifs.

3.3.3.1 Mesures d'évaluation de la qualité du profil

La plupart des mesures d'évaluation de la qualité du profil proposées sont adaptées au modèle de représentation conceptuel du profil utilisateur. Peu de travaux en RI personnalisée ont évalué la qualité du profil appris. Nous résumons dans ce qui suit les mesures d'évaluation utilisées dans l'estimation de la qualité du profil utilisateur dans quelques systèmes de RI personnalisée [187, 123, 55].

- Dans l'approche présentée dans [187], le profil utilisateur est représenté par un ensemble de concepts issus d'ontologie Web prédéfinie. La précision du profil est estimée par la moyenne des rangs des concepts non pertinents inclus dans le profil utilisateur. Dans le but d'évaluer la qualité du profil, les mesures de rappel, précision et la mesure F sont calculées sur les premiers 5%, 10%, .., 100% concepts pondérés du profil utilisateur. Finalement, la qualité du profil est mesurée par la moyenne des valeurs de la mesure F obtenues pour tous les utilisateurs. La mesure F est la moyenne harmonique des mesures de rappel et de précision calculée comme suit :

$$F - measure = \frac{2 * precision * recall}{precision + recall} \quad (3.2)$$

- Dans l'approche présentée dans [123], le profil utilisateur est également représenté par un ensemble de concepts issus d'une ontologie Web prédéfinie et est utilisé dans la désambiguïsation d'une requête donnée. Dans le but d'évaluer la précision du profil, les requêtes sont associés à n concepts du profil les plus similaires. Ces listes sont ensuite comparées avec les listes de concepts attribués manuellement par des utilisateurs pour chaque requête. La mesure de précision du profil est donnée par la formule suivante :

$$Accuracy = \left(\sum \frac{1}{1 + rank_i - ideal_rank_i} \right) / n \quad (3.3)$$

où n est le nombre de concepts associés à la requête, $rank_i$ est le rang du concept c_i donné par le système, $ideal_rank_i$ est le rang idéal attribué manuellement au concept c_i par les utilisateurs.

- Dans l'approche présentée dans [55], le profil utilisateur est représenté également par un ensemble de concepts issus d'une ontologie Web prédéfinie et utilisé dans un algorithme de classification des requêtes basé sur les réseaux de neurones. Les données d'apprentissage du classifieur consistent

en des documents pré-annotés (classifiés) par des concepts de l'ontologie de l'ODP. Dans le but d'évaluer la précision de la qualité du profil, des requêtes de test annotées des concepts corrects sont utilisées. L'évaluation se fait en classifiant des documents de test dans les catégories les plus similaires. La liste des concepts obtenus est comparée avec les annotations données manuellement par des utilisateurs réels. La mesure de précision utilisée est calculée comme suit :

$$Accuracy = \frac{N_{dc}}{N_{dt}} \quad (3.4)$$

où N_{dc} est le nombre de documents correctement classifiés comparative-ment à la classification manuelle, N_{dt} est le nombre total des documents utilisés dans l'évaluation.

3.3.3.2 Mesures d'évaluation de l'efficacité de RI

L'évaluation de l'efficacité du système de RI en présence du contexte repose sur deux classes de mesures : les mesures orientées-système et les mesures orientées-tâche.

3.3.3.2.1 Mesures d'évaluation orientées-système Les mesures d'évaluation orientées-système permettent d'évaluer l'efficacité du processus de RI personnalisée selon des mesures de rappel et précision. Toutefois ces mesures sont basées sur la pertinence binaire ("*pertinent*" or "*non pertinent*") qui ne tiennent pas en compte la subjectivité de l'utilisateur amenant à juger dans des contextes situationnels. L'étude effectuée dans [170] consiste à juger les documents pertinents de TREC selon une pertinence graduelle. Cette étude montre que 50% des documents jugés pertinents sont re-jugés comme faiblement pertinents, 34% comme moyennement pertinents et 16% comme fortement pertinents. Selon cette vue, des mesures d'évaluation supportant une pertinence graduelle ont été proposés utilisées pour l'évaluation des modèles de RI.

1. *la mesure RR (Relative Relevance).* Cette mesure permet de calculer un degré de correspondance entre la pertinence système (algorithmique) et la pertinence situationnelle de l'utilisateur qui intègre des niveaux de pertinence (1 : pertinent, 0.5 : partiellement pertinent, 0 : non pertinent). En pratique, deux valeurs de jugements de pertinence R_1 et R_2 sont calculées pour chaque niveau de pertinence et comparés selon les mesures cosinus ou Jaccard comme suit :

$$RR(R_1, R_2) = \frac{\sum(R_1 * R_2)}{(\sum R_1^2)^{\frac{1}{2}} * (\sum R_2^2)^{\frac{1}{2}}} \quad (3.5)$$

$$RR(R_1, R_2) = \frac{\sum(R_1 * R_2)}{\sum R_1 + \sum R_2 - \sum(R_1 * R_2)} \quad (3.6)$$

Plus la valeur de RR est petite, moins est la correspondance entre la pertinence système et la perception de pertinence de l'utilisateur dans un contexte de recherche. Cette mesure permet de comprendre les caractéristiques de la performance d'un SRI vis à vis la subjectivité de l'utilisateur. Néanmoins, elle pose un problème lors de l'évaluation comparative entre différents algorithmes de recherche voire entre différents SRI [140]. En effet les scores de pertinence algorithmique ne sont pas étalonnés à la même échelle entre différents SRI, ce qui rend la comparaison de mesures RR non significative.

2. *La mesure bpref (Binary PReFerence measure)*. La *bpref* mesure le nombre de fois que des documents jugés non pertinents sont retrouvés avant un document pertinent [35]. Contrairement aux autres mesures, cette mesure se focalise sur les documents réellement jugés. Par conséquent cette mesure n'est pas sensible à des jugements de pertinence incomplets.
3. *La mesure RHL (Ranked Half-Life measure)*. Contrairement à la mesure RR, la mesure RHL [28] est une mesure orientée-rang qui permet de comparer des systèmes entre eux. L'objectif de cette mesure est alors de déterminer, pour un SRI donné, le rang à partir duquel la moitié du nombre total de documents pertinents (ou de la quantité d'information pertinente dans le cas d'une pertinence graduelle) est présentée à l'utilisateur. Plus la mesure RHL est petite, plus les documents les plus pertinents sont en début de liste et donc, plus le SRI est efficace, en ce sens que l'ordre des documents retournés s'accorde avec l'ordre de pertinence contextuelle de l'utilisateur. La formule de calcul de la mesure RHL est la suivante :

$$RHL = R_m + \left(\frac{R_{tot}/2 - \sum_{R < R_m} Rel(R)}{Rel(R_m)} \right) \quad (3.7)$$

où R_m est le plus petit rang qui délimite la classe des documents dont la valeur de pertinence constitue la moitié de la pertinence totale, R_{tot} est la valeur de pertinence totale cumulée sur l'ensemble des documents présentés à l'utilisateur, $Rel(R)$ est la valeur de pertinence au rang R .

4. *The Cumulative Gain (CG) and Discounted Cumulative Gain (DCG)* [90, 91] Les mesures CG et DCG sont des mesures orientés-rang proposées dans un contexte de pertinence graduelle. L'objectif de ces mesures est d'évaluer la performance globale en estimant le gain de l'utilisateur en terme de pertinence cumulée en observant les documents situés jusqu'à un rang donné. Chaque niveau de pertinence i (e.g. pertinent, partiellement pertinent, non pertinent) est assigné une valeur $G[i]$ où le niveau de pertinence le plus élevé prend la valeur la plus élevée. Ces mesures sont données comme suit :

$$CG[i] = \begin{cases} G[1], & \text{if } i = 1 \\ CG[i - 1] + G[i], & \text{otherwise} \end{cases} \quad (3.8)$$

vu qu'il est moins probable qu'un document de rang élevé soit examiné par le utilisateur, la mesure DCG introduit un facteur d'atténuation, permet d'atténuer le gain de pertinence apporté par un document en fonction de son rang. La mesure DCG (Discount Cumulative Gain) est définie comme suit avec $\log i$ est utilisée comme un facteur d'atténuation :

$$DCG[i] = \begin{cases} CG[i], & \text{if } i < b \\ DCG[i - 1] + \frac{G[i]}{\log_b i}, & \text{otherwise} \end{cases} \quad (3.9)$$

5. *La mesure GRP (Generalised Recall and Precision)*. Cette mesure [91] est également une mesure orientée-rang qui étend les mesures rappel-précision classiques en considérant une pertinence graduelle. Le rappel généralisé (GR) et la précision généralisée (GP) sont calculés comme suit :

$$gP = \sum_{d \in R} r(d) / |R| \quad (3.10)$$

$$gR = \sum_{d \in R} r(d) / \sum_{d \in D} r(d) \quad (3.11)$$

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

3.3.3.2 Mesures d'évaluation orientées-tâche Ces mesures permettent d'évaluer l'interaction, la satisfaction, et l'efficacité du processus de recherche dans un contexte où l'utilisateur est intégré dans une tâche de recherche spécifique. Les premières mesures d'évaluation orientées-tâche ont été proposées dans le cadre de l'évaluation des systèmes interactifs dans TREC.

1. *User effort* : cette mesure permet d'évaluer la capacité de l'utilisateur pour accomplir une tâche de recherche en estimant la durée de la recherche (*elapsed time*), le nombre d'itérations (nombre de requêtes soumises), nombre de documents visualisés durant la navigation et le nombre de documents sauvegardés durant l'interaction de l'utilisateur avec le système. Des enquêtes peuvent être faites dans le but de mesurer des mesures subjectives de la performance du processus de recherche comme la satisfaction et la perception du succès du système. Ces mesures peuvent être reliées au niveau d'expertise ou de familiarité de l'utilisateur avec la tâche de recherche en question.
2. *Précision/rappel aspectuel* : ces mesures sont utilisées dans le cadre d'une tâche de recherche dite aspectuelle dans la tâche interactive de TREC, où la tâche consiste à retrouver des documents pertinents contenant les différents aspects recouverts par un *topic*. Ces mesures sont proposées dans ce cadre pour évaluer la différence des jugements de pertinence fournis par les utilisateurs et ceux fournis par les assesseurs de TREC. La précision ou rappel aspectuel est défini (par NIST) par la fraction de

tous les aspects qui s'occurrent dans au moins un document sauvegardé durant la recherche.

3.4 Les approches d'évaluation orientée-contexte initiées par TREC

Les tâches interactives et HARD initiées par TREC sont considérées les premières tentatives de l'évaluation orientée-contexte. Ces tâches sont des extensions de l'évaluation orientée-laboratoire dans le sens où ils ajoutent :

- la modélisation d'une interaction minimale entre l'utilisateur et le système (tâche interactive),
- l'introduction des facteurs contextuels associés à chaque requête et ayant un impact sur le comportement de l'interaction et par suite sur la différence de la perception de pertinence de l'information entre les utilisateurs (tâche HARD).

L'évaluation des SRI interactifs dans TREC consistent à comparer la performance du système en réponse à des requêtes construites automatiquement à la performance obtenue en réponse à des requêtes construites interactivement via l'intervention humaine. Les mesures d'évaluation introduites dans la tâche interactives sont dérivées des mesures classiques permettant d'évaluer le processus de recherche en lui même, et des mesures d'évaluation interactives utilisées dans l'évaluation de l'interaction utilisateur avec le système et l'impact de ses facteurs.

3.4.1 La tâche interactive de TREC

Le but fondamental de cette tâche consiste à examiner le rendement du processus de recherche dans un contexte de recherche interactif. Nous présentons dans la suite les objectifs de chacune des tâches interactives de TREC depuis TREC 1 jusqu'à TREC 8 ainsi qu'un récapitulatif des composantes expérimentales utilisées dans chacune des deux tâches, notamment les requêtes, les documents, les jugements de pertinence, le contexte de recherche ainsi que le protocole et les mesures d'évaluation. .

3.4.1.1 Aperçu des tâches interactives de TREC

- TREC 1, 2, 3 : les tâches principales dans TREC 1,2,3 consistent à examiner le mode interactif de construction des requêtes (*routing*) dans le

- but de comparer les systèmes interactifs et les systèmes purement automatiques.
- TREC 4 : la tâche consiste à retrouver le maximum des documents pertinents sur un intervalle de temps déterminé (15 à 20 minutes) et construire la requête associée optimale. Les mesures classiques de rappel-précision sont généralement utilisées pour mesurer l'effet de certaines variables isolées (liées aux systèmes) sur les performances de recherche.
 - TREC 5, 6 : dans cette tâche, le *topic* possède des aspects qui consistent en une description de l'information recherchée associée à différents aspects thématiques et pour lesquels plusieurs instances de réponses sont exactes. La tâche est similaire à celle dans TREC 4, à la seule différence que les documents pertinents retrouvés doivent recouvrir tous les aspects du *topic*. En plus de la mesure de la précision, les mesures de rappel aspectuel (qui mesure le nombre de documents pertinents retrouvés et recouvrant quelques aspects du *topic* dans l'ensemble total des documents pertinents recouvrant tous les aspects du *topic*) et le temps de recherche nécessaire pour accomplir une tâche (*elapsed search time*) sont utilisés dans TREC 5 et 6.
 - TREC 7, 8 : dans TREC 7 et 8, la tâche est similaire à celle dans TREC 6, le terme *aspect* est remplacé par le terme *instance* et où le temps de recherche est limité à 15 minutes dans TREC 7 et 20 minutes dans TREC 8. Dans ces tâches la comparaison entre systèmes n'est pas prise en compte. Dans TREC-7, la mesure de rappel aspectuel est remplacée par le rappel au niveau instance, qui mesure pour chaque utilisateur le nombre d'instances des réponses correctes trouvées pour une requête donnée.

3.4.1.2 Composantes expérimentales de la tâche interactive de TREC

Nous présentons dans le tableau 3.4.1.2 un récapitulatif des principales composantes des cadres d'évaluations de la tâche *inetractive* de TREC. Plus précisément, nous spécifions les caractéristiques de la collection de test, les mesures d'évaluation et le protocole d'évaluation utilisés.

- Collection de documents :
La tâche interactive de TREC utilise la collection .GOV créé dans le cadre de la tâche Web dans TREC 2002. Cette collection est indexée et utilisée via le moteur de recherche *Panoptic*.
- Collection de requêtes :
Les requêtes sont soit formulées par les utilisateurs dans le cas où la tâche de recherche consiste à créer la requête optimale (*routing query*) dans TREC 3, soit ils sont définis et évalués par les utilisateurs.
- Jugements de pertinence :
Les documents sont annotés selon une pertinence binaire par : perti-

Collection de test	Tâches interactives de TREC
Documents	collection Web .GOV (TREC10), Associated Press (disks 1-3) Wall Street Journal (disks 1-2), etc. (TREC 9)
Requêtes	Requêtes typiques de TREC ad hoc
Jugements de pertinence	Pertinent, Non-pertinent ou donnés explicitement par les participants (TREC 9,10)
Utilisateurs / contextes	Utilisateurs réels avec des tâches prédéfinies / contexte interactif : interactions utilisateurs, comportement de recherche
Evaluation	
Mesures	précision, rappel, rappel aspectual, précision aspectual, effort, temps pour accomplir une tâche (elapsed search time)
Protocole	Comparaison entre les systèmes interactifs (utilisateur + système) Vs. systèmes purement automatiques (TREC 3) Comparaison entre les systèmes interactifs (TREC 4,...,8), Systèmes Questions-Réponses (QA systems) (TREC 9,10)

TAB. 3.1 – Les composantes expérimentales des tâches interactives de TREC

ment ou non-pertinent. Cependant, dans TREC 9 et 10, les participants appliquent leur propre pertinence graduelle dans la détermination des documents pertinents plus précisément dans la tâche QA.

– *Utilisateurs*

Des utilisateurs réels sont sollicités pour effectuer les tâches de recherche dans la tâche interactive de TREC.

– Mesures d'évaluation :

Les mesures d'évaluation dans la tâche interactive de TREC permettent d'évaluer l'efficacité, l'efficacité (complexité de temps et d'espace) et la satisfaction de l'utilisateur. Plus précisément, les mesures standards de précision et de rappel sont utilisées dans TREC 3 et 4. En plus de ces mesures, une nouvelle mesure orientée-tâche a été introduite dans TREC 5, 6, 7 and 8, notamment la précision et le rappel aspectuel, le temps de recherche et l'effort. Dans la dernière tâche interactive (TREC 9, 10), la tâche principale se focalise sur les systèmes Question-Réponses où les mesures d'évaluation sont celles qui sont utilisées dans les tâches précédentes.

– Protocole d'évaluation

Le protocole d'évaluation dans le cadre de la tâche interactive de TREC

est mis en place dans le but de comparer les systèmes interactifs (utilisateur + système) avec les systèmes automatiques quant à la construction des requêtes et à la recherche. Dans ce but, des utilisateurs réels formulent des requêtes optimales pour un certain *topic* en utilisant des jugements de pertinence associés (TREC 3).

Le but dans le reste des tâches interactives de TREC (TREC 4, 5, 6, 7, 8) est de comparer les systèmes interactifs entre eux tout en éloignant l'effet du système de celui de l'utilisateur et du besoin en information. Pour cela, le protocole consistait à sélectionner un *topic* et c'est à l'utilisateur de trouver un nombre de documents pertinents dans un intervalle de temps bien précis. Dans TREC 9 et 10, la tâche principale se focalise sur les systèmes Questions-Réponses où les utilisateurs doivent rechercher la réponse exacte à une question et identifier un ou plusieurs documents supportant la réponse pour des *topics* liés au domaine de la santé ou du voyage.

3.4.2 La tâche HARD de TREC

Nous présentons dans cette section un aperçu de la tâche HARD de TREC ainsi que les composantes expérimentales impliquées dans le cadre de cette tâche.

3.4.2.1 Aperçu de la tâche HARD de TREC

Le but fondamental de la tâche HARD de TREC est d'intégrer le contexte de recherche dans le processus d'accès à l'information ainsi que dans le processus d'évaluation de l'efficacité de recherche.

Le contexte est défini par des métadonnées introduites aussi bien dans les documents que dans les requêtes de la collection de test. Ces métadonnées concernent les préférences de l'utilisateur et peuvent être rangées par des classes de données bibliographiques de l'utilisateur, des données concernant le type de résultats retournés la familiarité, le genre du document et la granularité (paragraphe ou document intégral), ou toute autre information qui permet de désambiguïser la recherche. D'autres informations additionnelles sont éventuellement demandées aux utilisateurs participant à la campagne d'évaluation à l'aide de formulaires de clarification et pouvant être considérés comme données du contexte.

3.4.2.2 Composantes expérimentales de la tâche HARD de TREC

Nous présentons dans le tableau 3.4.2.2 un récapitulatif des principales composantes expérimentales impliquées dans le cadre d'évaluation de la tâche *HARD* de TREC. Plus précisément, nous spécifions les caractéristiques de la collection de test utilisée, les mesures d'évaluation et le protocole d'évaluation.

Collection de test	Tâche <i>HARD</i> de TREC
Documents	corpus AQUAINT et documents du gouvernement des états unis
Requêtes	Requêtes typiques de TREC ad hoc
Jugements de pertinence	Non pertinent, moyennement pertinent, fortement pertinent
Utilisateurs /contextes	assesseurs TREC/ métadonnées de la requête
Evaluation	
Mesures	évaluations de recherche par passage : R-précision, passage R-recall, F score, bpref (tâche <i>HARD</i> de TREC 2003,2004) évaluations de recherche par document : MAP (tâche <i>HARD</i> de TREC 2003,2004), R-précision (tâche <i>HARD</i> de TREC 2005)
Protocole d'évaluation	Comparaison entre recherche de base (requête seulement) et recherche contextuelle (requête + contexte des métadonnées de la requête)

TAB. 3.2 – Composantes expérimentales de la tâche *HARD* de TREC

- Collection de documents :
Avant 2005, les collections de documents sont constituées de textes d'actualité issus d'une collection de documents, appelé AQUAINT, et de documents gouvernementaux des U.S. En 2005, la collection de documents utilisée consistait en toute la collection AQUAINT entier. Les fichiers résultats ne contenaient plus d'informations concernant les passages sélectionnés dans les documents.
- Collection de requêtes :
Les *topics* dans la tâche *HARD* 2005 sont sélectionnés à partir d'un ensemble de *topics* existants utilisés dans la tâche robuste de TREC et ayant le même format classique des *topics* de TREC. Les *topics* de la tâche *HARD* 2003 et 2004 ont principalement le format que les *topics* dans TREC mais sont enrichies avec des métadonnées qui décrivent l'utilisateur et le contexte de la requête. Ces métadonnées sont définies dans

des champs supplémentaires tels que *Purpose, Genre, Familiarity, Granularity, Geography, etc.*

- Jugements de pertinence : Dans la tâche HARD de TREC 2003, 2004, trois granularités de jugements de pertinence ont été fixés :
 1. *Non-relevant* : document non pertinent pour la requête ;
 2. *Soft-relevant* : le document est pertinent pour la requête mais il ne satisfait pas les métadonnées appropriées ;
 3. *Hard-relevant* : le document est pertinent et satisfait les métadonnées appropriées.

où *Soft-relevant* et *Hard-relevant* sont équivalents dans la tâche HARD 2005.

- *Utilisateurs*

Les utilisateurs dans la tâche HARD de TREC sont les assessseurs qui fournissent les jugements de pertinence en fonction de la similarité du sujet entre la requête et le document ainsi qu'en fonction des métadonnées associées à un topic donné.

- Mesures d'évaluation :

Dans la tâche HARD de TREC, les mesures d'évaluation sont définies au niveau document intégral ou au niveau passage. Les mesures d'évaluation au niveau document correspond à la moyenne des précisions moyennes standard (MAP). Pour l'évaluation au niveau passage, les mesures d'évaluation utilisées sont le rappel et la précision au niveau passage, la mesure F et la mesure $bpref$.

- Protocole d'évaluation

Le protocole d'évaluation dans le cadre de la tâche HARD de TREC consiste à comparer la performance du système en utilisant seulement la requête (recherche classique) à la performance obtenue en utilisant les requêtes et les métadonnées associées comme étant le contexte de recherche (recherche contextuelle). Dans les deux scénarios de recherche, les requêtes peuvent être tirées du champ Title ou le champ Description du *topic* définis dans le format standard d'un *topic* de TREC. Les participants utilisent leur propre moteur de recherche et retournent les listes des résultats pour les requêtes de la collection pour la recherche typique. Pour la recherche contextuelle, les assessseurs remplissent des métadonnées associées aux *topics* comme ils les perçoivent et les renvoient aux participants qui fournissent finalement les listes de résultats obtenues par la recherche contextuelle. Des interfaces de clarification typiques doivent être développées dans le but de permettre la saisie du contexte de la requête par les assessseurs.

3.4.3 Discussion

Les tâches interactives et HARD de TREC ont le mérite d'avoir introduit la dimension contexte dans le paradigme de l'évaluation orienté-laboratoire. Les interactions de l'utilisateur avec le système (tâche interactive) ou son contexte représenté par des métadonnées de la requête (tâche *HARD*) sont pris en compte dans le processus de recherche ainsi que dans le processus d'évaluation des systèmes. Ces tâches ont permis d'aboutir à des conclusions consistantes concernant l'évaluation du processus de recherche en contexte.

Cependant, ces tâches sont très spécifiques, ne permettant pas d'évaluer, sans peine d'extensions, des modèles d'accès intégrant un contexte multidimensionnel incluant un profil utilisateur à centres d'intérêts multiples, un profil utilisateur mobile, etc.

Ceci a permis l'émergence des approches d'évaluation fondées sur l'utilisation des contextes de recherche simulés ou des contextes réels. Ces approches d'évaluation sont conçues dans le cadre de l'évaluation des modèles de RI personnalisés. Le but fondamental de ces approches est d'évaluer les SRI en présence de la nature dynamique du besoin en information de l'utilisateur d'une part ainsi que des jugements de pertinence d'autre part.

3.5 Évaluation des SRI personnalisés par simulation de contexte

Nous présentons dans la suite de cette section le principe de la simulation du contexte et les principales composantes de l'évaluation par simulation de contexte dans quelques SRI personnalisés.

3.5.1 La simulation du contexte

Les approches d'évaluation par simulation du contexte ne requièrent pas des utilisateurs réels ni des interactions avec le système de recherche. Ces approches consistent à créer des situations ou des contextes de recherche hypothétiques représentant les profils utilisateurs simulés. Nous distinguons entre les approches d'évaluation qui réutilisent des collections de test de type laboratoire [180] et les approches définissent des requêtes et des jugements de pertinence hypothétiques par utilisation d'une collection de documents accessible au large public [164].

La simulation du contexte dans [180] consiste à représenter un utilisateur ayant un centre d'intérêt par un ensemble de requêtes associées à un même domaine d'intérêt prédéfini dans la collection de TREC, et exploiter les jugements de pertinence donnés pour chaque requête dans la collection comme étant des interactions hypothétiques résultant de la soumission de ces requêtes. L'évaluation de l'efficacité de la recherche personnalisée est basée sur l'utilisation des jugements de pertinence prédéfinies dans la collection [180].

La simulation de contexte dans [164] consiste à représenter les centres d'intérêts d'un utilisateur simulé par des concepts d'une ontologie de domaines prédéfinies (l'ODP) et à partir de chaque concept, une requête est générée automatiquement traduisant ainsi un contexte de recherche défini par une requête et un centre d'intérêt associé. Les jugements de pertinence sont hypothétiquement défini par l'ensemble de documents classifiés sous un concept simulant un centre d'intérêt.

3.5.2 Composantes expérimentales de l'évaluation par simulation de contexte

Nous présentons dans le tableau 3.5.2 un récapitulatif des principales composantes de l'évaluation basée sur la simulation des contextes dans quelques SRI personnalisés. Plus précisément, nous spécifions, pour chaque approche, la collection de tests utilisée, les mesures d'évaluation et la méthodologie d'évaluation.

- Documents :
Les collections de documents utilisés dans l'évaluation par simulation des contextes sont souvent réutilisées à partir d'une collection de test contrôlé pour l'évaluation des systèmes tels que TREC [180]. La collection de documents de Disk 1&2 de TREC est utilisé dans [180]. D'autres approches [164] utilisent une collection de pages Web disponibles via des ressources ouvertes tels les pages Web classifiées dans l'ontologie de domaines l'ODP³.
- *Requêtes*
Les requêtes sont soit fournies par une collection de test réutilisée [180], soit générées automatiquement [164] relativement aux contextes de recherche simulés. Dans [180], les requêtes correspondent aux mots clés du champ *Title* des *topics* de la collection Disk 1&2. Dans [164], les requêtes comportent les mots clés les mieux pondérés représentatifs du concept de l'ontologie considérés comme contexte de recherche à simuler.
- *Jugements de pertinence*
Les jugements de pertinence peuvent être réutilisés à partir de la col-

³<http://www.dmoz.org>

Collection de test	
Documents	Collection de documents TREC [180], collection de pages web [164], etc.
Requêtes	Requêtes de la collection TREC 1 (Disk 1&2) [180], requêtes générées des concepts d'une ontologie prédéfinie [164]
Jugements de pertinence	Pertinent, Non-pertinent
Utilisateurs /contextes	Utilisateurs simulés par leur profil/contexte cognitif : ou profil décrivant les centres d'intérêts
Evaluation	
Mesures	Précision@n, Rappel@n, MAP évaluations de recherche par document : MAP (tâche HARD de TREC 2003,2004), R-précision (tâche HARD de TREC 2005)
Protocole d'évaluation	Les interactions des utilisateurs sont non requises Apprentissage du profil par exploitation des jugements de pertinence donnés pour chaque requête [180] dans TREC 1, ou les page web classifiés sous un concept simulé de l'ontologie de l'ODP [164] Comparaison entre la recherche classique (requête seulement) et la recherche personnalisée (requête + contexte)

TAB. 3.3 – Composantes expérimentales de l'évaluation par simulation des contextes

lection de test prédéfinie ou alors spécifiés selon des critères dépendants des contextes simulés. Dans le cas de l'utilisation d'une collection de test prédéfinie [180], les documents sont assignés à des valeurs de pertinence binaire : pertinent, non-pertinent. Dans l'approche d'évaluation présentée dans [164], les documents sont assignés à des valeurs *signal* ou *noisy* qui signifient respectivement que le document est pertinent ou non pertinent. Un document est dit *signal* pour une requête générée à partir d'un concept simulé de l'ontologie de l'ODP si ce document est classifié sous ce concept, sinon il est considéré *noisy*.

– *Utilisateurs/Contextes*

Les utilisateurs sont remplacés par des contextes simulés et sont représentés dans la collection par leur profil. Un contexte de recherche est simulé dans [180] par une séquence de requêtes liées à un même besoin en information permettant de construire le profil de l'utilisateur. Celui-ci contient un centre d'intérêt représenté par un domaine prédéfini dans la collection de test utilisée de TREC 1. Il est construit sur l'ensemble des requêtes d'un domaine à partir des jugements de pertinence fournis par les assesseurs de TREC pour chaque requête.

L'utilisateur est simulé dans [164] via un concept de l'ontologie représentant ainsi un profil utilisateur à un instant donné. A chaque concept/contexte simulé correspond une requête générée automatiquement et un ensemble de documents classifiés sous ce concept choisi aléatoirement, appelé documents "‘profil’", servant à la construction du profil utilisateur pour le concept associé.

– *Protocole d'évaluation*

Le protocole d'évaluation impliqué dans le cadre de l'évaluation par simulation des contextes consiste à comparer la performance de la recherche classique (ignorant le profil utilisateur) à celui de la recherche personnalisée intégrant le profil utilisateur dans la chaîne d'accès à l'information. Nous distinguons entre deux types de protocoles d'évaluation. Le premier type des protocoles est basé sur la construction d'un profil pour chaque requête de test à partir des sources d'évidence liées à la requête elle-même [164]. Le deuxième type des protocoles d'évaluation consiste à apprendre le profil à partir d'un ensemble de requêtes d'apprentissage [180, 123] (Cross-validation). Le profil ainsi construit est utilisé pour évaluer les requêtes de test.

Plus précisément, le protocole d'évaluation dans [164] suit le premier type de protocoles d'évaluation et consiste à construire un profil pour chaque requête de test à partir d'un ensemble de documents classifiés sous le concept générant la requête, appelé *documents profil*. Ce profil est exploité dans le but de réordonner les résultats de recherche de la requête de test associée. Dans [180], le protocole d'évaluation consiste à subdiviser l'ensemble des requêtes annotées d'un même domaine en un sous-ensemble d'apprentissage qui permettent de construire le profil en utilisant les documents pertinents associés à la requête (fournis par les assesseurs de TREC) et un sous-ensemble de test permettant d'évaluer la performance du système tout en intégrant le profil de test dans la fonction d'appariement du modèle de recherche d'information.

– *Mesures d'évaluation*

Les mesures d'évaluation consistent en des mesures standards de rappel et précision à n documents et la moyenne des précisions moyennes sur l'ensemble des requêtes de test (MAP) dans le cas d'une collection de test prédéfinie (e.g. TREC). Le référentiel d'évaluation d'une requête issue d'un concept / centre d'intérêt de l'ontologie dans [164] est construit sur la base des documents Web classifiés sous ce concept.

3.5.3 Discussion

La mise en œuvre d'un cadre d'évaluation générique pour l'évaluation des systèmes de RI en présence du contexte permet d'asseoir une base d'évaluation comparative des systèmes, mais présente toutefois des limitations concernant le réalisme au niveau de l'évaluation. Généralement, l'évaluation par simulation de contextes présente deux avantages majeurs :

- n'est pas coûteuse en temps puisqu'elle n'implique pas des utilisateurs réels,
- permet d'effectuer une évaluation comparative et reproductible tel est le cas de l'évaluation orientée-laboratoire.

En revanche, les principales limites se recourent avec des limites de l'évaluation orientée-laboratoire.

- Les requêtes de la collection peuvent ne pas intéresser beaucoup d'utilisateurs. Dans quelques approches, elles sont réutilisées à partir d'une collection de type laboratoire [180] ou générées selon un processus automatique [164].
- les jugements de pertinence sont fournis par des assesseurs professionnels et peuvent être en contraste avec l'incertitude et nature dynamique du besoin en information ainsi que la perception de l'utilisateur au cours de la recherche.

Compte tenu de ces limites, beaucoup d'études adopte le paradigme d'évaluation de l'usager basé sur l'utilisation des utilisateurs réels qui interagissent avec le système dans des contextes de recherche réels.

3.6 Évaluation des SRI personnalisés par utilisation des contextes réels

Nous nous intéressons dans cette section aux approches d'évaluation des SRI personnalisés qui se focalisent sur la construction des profils utilisateurs à partir des contextes de recherche réels. Nous présentons tout d'abord le principe de l'évaluation par utilisation des contextes réels (*user studies*) et les composantes expérimentales de ce type d'évaluation adoptés dans quelques SRI personnalisés.

3.6.1 Principe de l'évaluation

Le but fondamental de l'évaluation par utilisation des contextes réels est d'évaluer la performance qualitative et quantitative d'un système par l'inves-

tissement des vrais utilisateurs en interaction naturelle avec le système [140]. L'objectif de ce type d'évaluation est de prendre en compte des situations de recherche réelles révélant ainsi la dynamique du besoin en information de l'utilisateur ainsi que la dynamique des jugements de pertinence qui traduisent non seulement une pertinence thématique mais aussi une pertinence perceptionnelle ou situationnelle.

Dans ce but, ce type d'évaluation consiste à solliciter des utilisateurs pour accomplir des tâches de recherche spécifiques en interrogeant le SRI [39]. Les utilisateurs peuvent être des enfants [168], des membres d'une équipe de travail dans une variété d'organisations [169], des utilisateurs du web [123], etc. La méthodologie d'évaluation consiste à tester et évaluer l'efficacité du système à travers les utilisateurs, leurs interactions et leur utilisation naturelle du système [140]. Cette évaluation est accomplie selon trois composantes [140] :

1. l'implication des utilisateurs participants dans des tâches de recherche dédiées à l'évaluation du système,
2. l'application des interprétations des besoins en information dynamique,
3. la création des jugements de pertinence multidimensionnels et dynamiques.

La problématique majeure dans l'évaluation basée sur l'utilisation des contextes réels consiste à balancer entre le réalisme et le contrôle. En effet, si des utilisateurs différents sont utilisés dans des situations de recherche différentes, il sera difficile de répéter les expérimentations à cause des différences entre les utilisateurs (intelligence, connaissance, familiarité, location, etc.) et par conséquent il sera difficile de déterminer les facteurs (variables expérimentales) qui influencent sur l'efficacité de recherche. Dans le but de résoudre ces problèmes, plusieurs recommandations sont données

[43, 140] et se résument comme suit :

- attribuer arbitrairement un nombre suffisant d'utilisateurs dans des groupes ; chaque groupe est identifié par une même tâche de recherche (appelé *search tasks* [140]),
- garantir un minimum d'interactions entre les utilisateurs,
- effectuer les expérimentations dans des situations aveugles dans le sens où les participants ignorent le scénario d'évaluation expérimenté (e.g. évaluation d'un modèle de RI personnalisé intégrant un profil de l'utilisateur, ou l'évaluation d'un modèle de RI classique qui ne considère pas un profil de l'utilisateur,
- permuter l'ordre des tâches entre les groupes des utilisateurs,
- effectuer une étude pilote avant l'étude principale de test.

3.6.2 Composantes expérimentales de l'évaluation par utilisation des contextes réels

L'évaluation par utilisation des contextes réels est appliquée pour l'évaluation des systèmes de RI personnalisés [123, 173, 114]. Généralement, l'évaluation se base sur l'utilisation d'un moteur de recherche commercial ou des interfaces de recherche dans de larges collections. Deux types d'évaluation par utilisation des contextes réels existent :

- le premier type est basé sur une approche hybride [162] qui intègre une collection de test prédéfinie (e.g. TREC) dans un protocole d'évaluation impliquant des vrais utilisateurs.
- Le deuxième type d'évaluation [173, 123] implique des vrais utilisateurs qui formulent leur requêtes et fournissent les jugements de pertinence sur une collection de documents ouverte au public pour chacune des requêtes formulées.

Le type d'évaluation hybride exploite seulement la collection de requêtes et de documents d'une campagne d'évaluation tandis que les jugements de pertinence sont fournis par des vrais utilisateurs pour chaque requête et où le contexte de recherche est représenté par le profil utilisateur. Un exemple de cette approche d'évaluation est adopté dans [162] qui étend les collections TREC par des contextes représentés par l'historique de recherche de l'utilisateur englobant les requêtes et les documents *cliqués*. Chaque requête de la collection TREC est reformulée par les utilisateurs dans le but de définir des requêtes liées à un même besoin en information et par la suite définir une session de recherche.

L'évaluation dans [173, 123] adoptent le deuxième type d'évaluation où des vrais utilisateurs formulent leur requête selon un besoin spécifique en exploitant une collection web via une interface Web (Google API, Yahoo API, etc.) et où les jugements de pertinence sont fournies pour chaque requête par l'utilisateur ayant soumis la requête.

Nous présentons dans le tableau 3.6.2 un récapitulatif des principales composantes de l'évaluation basée sur l'utilisation des contextes réels. Plus précisément, nous spécifions, pour chaque approche, la collection de tests utilisée, les mesures d'évaluation et la méthodologie d'évaluation adoptées.

- *Collection de documents*

La collection de documents utilisée dans ce type d'approches peut être fournie par une collection de type laboratoire, tel est le cas de TREC [163] dans le cas d'une approche hybride, ou alors une collection Web accessible via une interface publique (Google API, ODP API) [173, 123, 55, 162].

- *Topics*

Selon la collection de documents utilisée, les requêtes sont soit for-

Collection de test	
Documents	Collection de documents fournie par TREC (Disk1&2, TERABYTE, tâche Web de TREC 2003, etc.) [162], collection des pages Web accessible via une interface web d'un moteur de recherche en ligne (Google API) [173]
Requêtes	Requêtes fournies par TREC (Disk1&2, TERABYTE, etc.) [162], requêtes formulées par l'utilisateur lui-même [173, 123]
Jugements de pertinence	Jugements donnés explicitement par l'utilisateur (Pertinent, Non-pertinent) [162, 123], mesures de pertinence implicite (données de clics, temps de lecture etc.) [173]
Utilisateurs / contextes	Utilisateurs réels/ profil utilisateur décrivant des centres d'intérêts de l'utilisateur
Evaluation	
Mesures	Précision@n, Rappel@n, rang moyen des documents cliqués
Protocole d'évaluation	Les interactions des utilisateurs sont non requises Apprentissage du profil par exploitation de l'historique d'interaction de l'utilisateur avec le système Comparaison entre la recherche classique (requête seulement) et la recherche personnalisée (requête + profil utilisateur)

TAB. 3.4 – Composantes expérimentales de l'évaluation par utilisation des contextes réels

mulées par des utilisateurs réels en donnant comme indication la tâche de recherche simulée [173, 123, 55, 181], soient formulés par affinement d'un topic prédéfini dans un contexte de tâche précise dans la cadre d'une approche intégrant une collection de test de TREC et des utilisateurs réels [163].

– *Jugements de pertinence*

Les jugements de pertinence sont basés dans quelques approches sur des indicateurs issus des interactions de l'utilisateur avec le système [173]. Nous citons les données de *clics*, le temps de lecture, etc. Ainsi, un docu-

ment est considéré pertinent et contribue à la construction du profil utilisateur s'il a été *cliqué* par l'utilisateur [173]. Dans le cas d'une évaluation où les utilisateurs jugent la liste de résultats retournés par le système, le document prend l'une des valeurs pertinent, non-pertinent.

– *Utilisateurs/Contextes*

Des vrais utilisateurs sont impliqués dans le processus de RI afin d'accomplir les tâches de recherche servant à l'évaluation du système. Ils peuvent être des experts dans les tâches de recherche simulés, des membres des équipes dans une organisation [169] ou des utilisateurs qui accomplissent une recherche web usuelle [123]. A partir des interactions réelles d'un utilisateur avec le système dans des tâches de recherche précises, le système sauvegarde un historique de recherche de l'utilisateur, déployé ensuite pour l'apprentissage de son profil décrivant ses centres d'intérêts.

– *Protocole d'évaluation*

Le protocole d'évaluation consiste à faire interagir les utilisateurs avec le systèmes selon des tâches de recherche précises. Les interactions de l'utilisateur avec le système sont sauvegardés souvent dans un fichier *log* d'interaction utilisateur-système comprenant souvent l'identifiant de l'utilisateur, l'identifiant de la session, la requête basée mots clés, le temps de soumission la requête, les documents *cliqués* et leur rang pour la requête donnée.

Dans certaines approches, le fichier *log* contient des informations additionnelles dépendantes de l'approche de personnalisation en question, tels que les résumés de documents [162, 173]. Dans le cas ou le contexte dépend des catégories associées à la requête [56], les utilisateurs sont sollicités pour fournir des jugements sur les catégories pertinentes au sujet de la requête pour les sauvegarder dans le fichier *log*.

L'ensemble des requêtes de test servant à l'évaluation du système sont soient sélectionnées à partir du fichier *log* [162, 173], ou formulées par l'utilisateur pour lesquelles ils jugent les top- N documents retournés par le système.

– *Mesures d'évaluation*

Les mesures d'évaluation de performance du système peuvent être des mesures classiques de rappel et précision et leur dérivée. L'utilisation de la mesure de la précision moyenne MAP est possible grâce à la construction d'un référentiel de pertinence sur la base de l'ensemble des documents jugés pertinents par les différents utilisateurs pour une même requête [123]. D'autres mesures orientées-rang telles que le rang moyen [173] et DCG (Discounted Cumulative Gain) [91] et NDCG [14] sont également utilisées et se basent sur la position des documents pertinents dans la liste de résultats, la précision à n documents [55].

3.6.3 Discussion

Le défi majeur introduit par les approches d'évaluation par utilisation des contextes réels consiste à balancer entre le réalisme et le contrôle dans les expérimentations. L'introduction du réalisme permet de pallier les limites de l'évaluation basée sur la simulation des contextes. Les avantages sont principalement liés à l'évolution du besoin et de la notion de pertinence dans un contexte de recherche interactif, plus précisément :

1. l'application des tâches de recherche précises et individuelles impliquent un besoin en information dynamique et des interprétations dynamiques des résultats recherchés,
2. l'assignation des jugements de pertinence multidimensionnel et dynamique au cours de la recherche.

Toutefois, le réalisme introduit dans les cadres d'évaluation, étant en faveur de la mise en œuvre des expérimentations contrôlées, impose des limites liées principalement au coût en temps et à la reproductibilité des résultats et par conséquent à l'inconsistance de l'évaluation comparative. L'étude des facteurs ayant un impact sur le processus de recherche tels que la variation des paramètres liés à la construction du contexte ou le modèle d'appariement est une tâche difficile à accomplir vu que les expérimentations ne peuvent pas être contrôlables dans les mêmes conditions de recherche.

3.7 Conclusion

Nous avons présenté dans ce chapitre l'évolution des approches d'évaluation classiques des SRI basées sur le paradigme d'évaluation de *Cranfiled* à des approches d'évaluation fondées sur un paradigme d'évaluation orienté-contexte. Nous avons présenté le principe de l'évaluation des systèmes selon les approches d'évaluation orienté-laboratoire basé sur le paradigme de Cranfiled. Ce paradigme est à la base de l'utilisation d'un corpus de test comprenant une collection des documents, une collection des requêtes et l'ensemble des documents pertinents associés à la requête par des juges experts.

Nous avons ensuite évoqué les problématiques majeures de ce type d'évaluation en présence du contexte. Ces limites sont principalement liées à la faible représentativité des requêtes, l'absence de la notion de pertinence dynamique et du besoin en information dynamique, ainsi que l'évaluation des SRI en utilisant des mesures quantitatives ignorant les interactions de l'utilisateur avec le système.

Ceci a conduit à l'émergence des approches d'évaluation des SRI orientées-

contexte développées en RI personnalisée. Dans ce cadre, nous avons abordé l'évaluation de la qualité du profil, un aperçu des approches d'évaluation orientées-contexte dédiées à l'évaluation de l'efficacité du modèle de RI personnalisée ainsi que les mesures d'évaluation orientées-contexte de la qualité du profil et de l'efficacité de recherche. Ensuite, nous avons exploré les trois grandes tendances pour l'évaluation des SRI en présence du contexte, les expérimentations en laboratoire initiées par TREC, l'évaluation par simulation de contexte et l'évaluation par utilisation des contextes réels. Pour chacune de ces tendances, nous avons mis le focus sur le principe de l'évaluation et les composantes expérimentales impliquées.

Nous pouvons constater que l'évaluation par utilisation des contextes réels semble efficace en termes d'utilité et d'utilisabilité réelles, mais cette méthode d'évaluation n'est pas toujours réalisable et demande énormément de temps comparativement à l'évaluation par simulation de contexte. La combinaison de toutes les techniques possibles semble nécessaire si l'on veut obtenir une évaluation qui soit la plus juste possible.

Dans le but de balancer entre le réalisme et les expérimentations contrôlables dans les cadres d'évaluation, plusieurs projets (QUAERO⁴, Lemur Query Log Project⁵ introduit par Croft, LogCLEF⁶ 2009) ont été lancés récemment dans le but de collecter des données des utilisateurs réels, comprenant des requêtes et les comportements de *clicks* des utilisateurs au cours de leurs activités de recherche. Ce type de projets offre l'opportunité d'exploiter ces données dans des cadres d'évaluation orientés-laboratoire reproductibles dans les mêmes conditions de recherche.

⁴<http://www.quaero.org>

⁵<http://lemurstudy.cs.umass.edu/>

⁶<http://www.uni-hildesheim.de/logclef/>

Deuxième partie

Conception et évaluation d'un
modèle d'accès personnalisé à
l'information basé-session dans
des cadres d'évaluation
orientés-contexte

Chapitre 4

Prédiction contextuelle du type de besoin derrière la requête : proposition et évaluation

4.1 Introduction

L'objectif fondamental de l'accès contextuel à l'information est de répondre au mieux aux besoins en information de l'utilisateur en intégrant le contexte de recherche défini par des facteurs dépendants directement de l'utilisateur, de sa requête ou de l'environnement de recherche dans le processus d'accès à l'information [16]. Le chapitre 2 de l'état de l'art présenté dans la première partie de cette thèse a permis de cerner le domaine de l'accès contextuel à l'information guidé par la tâche de recherche informationnelle, navigationnelle ou transactionnelle. Les défis majeurs dans ce cadre, consistent à identifier les caractéristiques des requêtes induisant chacune des tâches de recherche à prédire, servant ainsi à leur estimation précise, et ensuite à l'identification des sources d'évidence appropriées permettant d'améliorer la performance de la recherche. Notre contribution présentée dans ce chapitre porte sur la prédiction du type de besoin derrière la requête, traduisant la nature de la tâche de recherche, en exploitant les caractéristiques morphologiques de la requêtes et le type de besoin des requêtes précédentes dans la session. Nous évaluons notre approche selon un protocole d'évaluation adapté à la prédiction du type de besoin en intégrant le profil de la session dans la classification des requêtes. Ce protocole est à la base de la construction d'une séquence des sessions de recherche où une session est simulée par une séquence de requêtes ayant le même type de besoin.

Ce chapitre est organisé comme suit. La section 5.2 présente les problématiques et les motivations de notre contribution. La section 5.3 présente l'architecture

générale du système qui supporte notre modèle. La section 4.3 décrit notre approche de prédiction du type de besoin derrière la requête. La section 4.4 présente l'évaluation de notre approche. La dernière section conclut le chapitre.

4.2 Problématique et motivations

Le but fondamental de la prédiction du type de besoin derrière la requête est de mettre en place un modèle de RI orientée-tâche. Un besoin en information de l'utilisateur peut être informationnel lié à la recherche du contenu, navigationnel lié à la recherche des sites d'accueil et transactionnel lié à la recherche des services en ligne. Pour chaque type de besoin, des stratégies de recherche sont dédiées à exploiter des sources d'information appropriées permettant ainsi d'améliorer la performance de la recherche selon le type de besoin.

Nos motivations dans ce but consiste à prédire le type de besoin informationnel derrière la requête dans un système d'accès personnalisé à l'information [2, 13]. Nous nous intéressons à la prédiction du type de besoin des requêtes sur le *Web* en fonction du contexte de la session, plus particulièrement les requêtes ambiguës, dites multi-facettes, qui satisfont des caractéristiques associées à plusieurs types de besoins. Notre approche de prédiction du type de besoin consiste à exploiter simultanément des caractéristiques morphologiques de la requête d'une part et le profil de la session défini par le type des requêtes de la session d'autre part [1].

Notre intuition concernant l'exploitation du profil de la session se base sur le fait que l'utilisation seule des caractéristiques de la requête est insuffisante pour la prédiction de son type du fait de son aspect multi-facettes ou son ambiguïté. Par exemple, pour une requête navigationnelle contenant un verbe comme "*Preserve and Prevent*", les approches classiques basées sur l'utilisation seule des caractéristiques de la requête ne sont pas capables de prédire correctement le type navigationnel de la requête. En effet, l'existence d'un verbe dans les termes de la requête est une caractéristique des requêtes informationnelles et transactionnelle [87] à la fois. En outre, notre méthode permet de renforcer la probabilité de la classification d'une requête selon le type du profil de la session même si elle satisfait moyennement les autres types de besoins.

La plupart des approches dans le domaine exploitent seulement des caractéristiques morphologiques de la requête [96, 95] ou des indicateurs du type de besoin dépendants du comportement et des données de *clics* de l'utilisateur sur l'ensemble des résultats retournés en réponse à une requête donnée [114]. Contrairement à ces approches nous exploitons le type de besoin porté par les requêtes de la session dans la prédiction du type de besoin d'une nou-

velle requête. Comparativement à ces approches classiques, notre approche se distingue par :

1. l'intégration du profil de la session dans la prédiction du type de besoin des nouvelles requêtes.
2. l'estimation du type de besoin de la requête en calculant des probabilités que la requête satisfait des propriétés morphologiques d'un type de besoin particulier d'une part et selon le profil de la session d'autre part.

4.3 Prédiction du type du besoin derrière la requête

Notre approche de prédiction du type de besoin derrière la requête est basée sur la combinaison des sources d'évidence issues des caractéristiques de la requête et du profil de la session courante défini par le type unique de ses requêtes [1]. Cette approche est illustrée dans la figure 4.1. L'estimation du type de besoin selon les propriétés morphologiques de la requête est basée sur le degré de satisfaction de la requête aux caractéristiques propres à ce type. Ces caractéristiques dépendent principalement du nombre de termes de la requêtes, des règles d'existence des verbes, des termes transactionnels, la distribution des termes de la requête dans les titres et les hyperliens des pages web. L'estimation du type de besoin selon le profil de la session est basée sur le calcul d'une probabilité de basculement en se basant sur une longueur moyenne des sessions définie à chaque type de besoin.

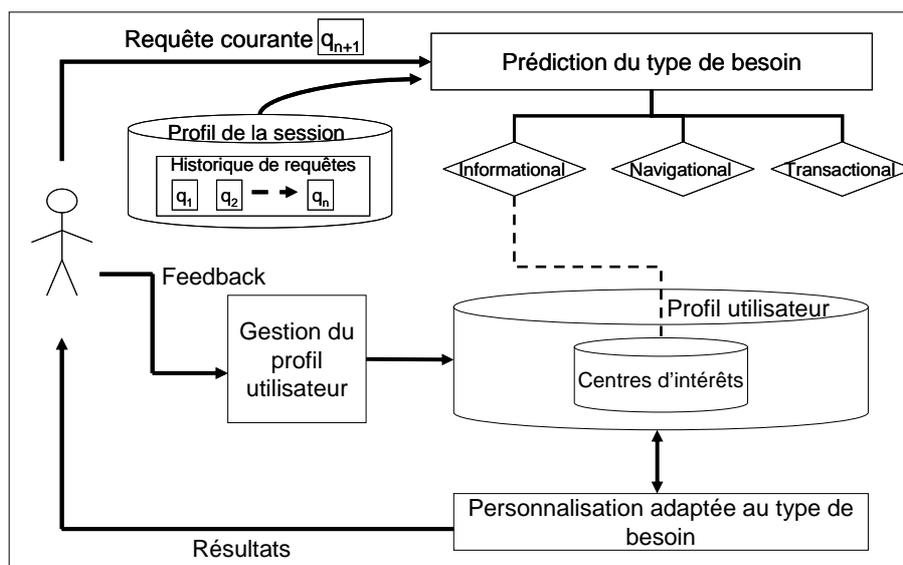


FIG. 4.1 – Architecture générale de notre modèle

Nous nous orientons vers le calcul des probabilités de classification des requêtes selon le type informationnel, navigationnel ou transactionnel dans un mécanisme de prédiction contextuelle du type de besoin. Nous présentons dans cette section :

1. l'estimation du type de besoin selon la morphologie de la requête,
2. l'estimation du type de besoin selon le profil de la session,
3. le principe de classification en combinant les estimations issues de la morphologie de la requête et du profil de la session.

4.3.1 Estimation du type de besoin selon la morphologie de la requête

Nous présentons dans cette section l'estimation du type de besoin de la requête en se basant seulement sur la définition du descripteur morphologique de la requête.

4.3.1.1 Le descripteur morphologique de la requête

Selon différentes études analysant les requêtes des utilisateurs, les caractéristiques les plus importantes permettant de prédire le type de besoin derrière la requête sont classifiées en deux types : les caractéristiques linguistiques et morphologiques [96, 95] et les caractéristiques liées au comportement de *clicks* des utilisateurs observées sur l'ensemble des résultats retournés en réponse à la requête [114].

Les caractéristiques linguistiques et morphologiques de la requête se basent souvent sur la longueur de la requête [87], l'existence des noms propres ou des noms des organisations [96, 87], des verbes (*buy*, *download*, etc.) ou des extensions des fichiers répondant à des besoins transactionnels (*exe*, *mp3*, *jpeg*, *mvi*, etc.) [95, 87], la distribution des termes de la requête dans une collection de pages annotées par le type navigationnel (sites d'accueil personnel ou d'organisations, etc.) ou transactionnel (existence des hyperliens permettant d'accomplir des transactions) [96] ou le taux d'utilisation des termes de la requête dans les titres des pages et les textes des hyperliens [96].

Nous sélectionnons les caractéristiques des requêtes les plus importantes exploitées dans les travaux du domaine, résumées et argumentées comme suit :

1. **La longueur de la requête** : cette caractéristique désigne le nombre de termes de la requête. Elle permet d'identifier les requêtes navigationnelles

ayant généralement une longueur plus petite que les requêtes transactionnelles ou informationnelles [87]. Nous fixons la longueur critique d'une requête informationnelle à une valeur plus grande que trois (3) comme étant une longueur moyenne calculée selon une analyse d'un fichier d'un moteur de recherche sur le Web [87].

2. **L'existence des verbes Vb** : cette caractéristique est une propriété générale des requêtes informationnelles ou transactionnelles, qui consiste souvent en un verbe représentant respectivement le besoin en information ou le service transactionnel demandé. Elle a été définie spécialement pour les requêtes transactionnelles contenant des verbes particuliers (tels que *Buy, Download, etc.*) [87].
3. **L'existence des termes transactionnels Tr** : cette caractéristique est une propriété des requêtes transactionnelles. Elle est souvent identifiée à l'aide d'une liste de termes préalablement définie [87] ou par apprentissage supervisé [120].
4. **L'existence des termes informationnels Ti** : nous avons introduit cette nouvelle caractéristique comme étant une propriété des requêtes informationnelles. En effet, la plupart des approches dans le domaine considèrent que le type par défaut des requêtes est informationnel [96, 114].
5. **Le taux d'utilisation des termes de la requête dans les titres des pages web $\tau_t(Q)$** : nous définissons ce taux comme étant une nouvelle caractéristique des requêtes qui permet de quantifier l'occurrence des termes de la requête dans les titres des pages Web. Ce taux permet d'identifier principalement les requêtes navigationnelles et transactionnelles des requêtes informationnelles. Cette propriété a été exploitée implicitement dans l'approche de Savoy et al. [86] dans une indexation des documents basée seulement sur les termes issus des balises du titre des pages *Web*. En effet, les termes des requêtes navigationnelles contiennent souvent des noms propres ou des noms des organisations qui apparaissent fréquemment dans les titres des pages et les textes des hyperliens [96]. Les requêtes transactionnelles contiennent également des termes ou des verbes transactionnels qui ocurrent souvent dans les titres des pages contenant la transaction à effectuer [95, 120].
6. **Le taux d'utilisation des termes de la requête dans les textes des hyperliens $\tau_a(Q)$** : cette caractéristique est déjà proposée dans [96] et permet également d'identifier les requêtes navigationnelles et transactionnelles [86, 95] des requêtes informationnelles. En effet, les termes de ces deux types de requêtes ocurrent dans les textes d'ancrage des hyperliens des pages pointant vers les sites d'accueil recherchés dans le cas des requêtes navigationnelles ou vers les pages qui permettent d'accomplir la tâche transactionnelle [86].

Sur la base de ces caractéristiques sélectionnées, nous définissons un descri-

peur morphologique d'une requête Q , noté $F(Q)$ comme suit :

$$F(Q) = (\|Q\|, Vb, Tr, Ti, \tau_t, \tau_a)$$

où $\|Q\|$ est la longueur de la requête estimée par le nombre de termes de la requête, Vb , Tr et Ti sont des valeurs booléennes qui indiquent si la requête contient ou non respectivement un verbe, des termes transactionnels (*download*, *buy*, etc.) et des termes interrogatifs. τ_a (resp. τ_t) est le taux d'utilisation des termes de la requête dans les textes d'ancrage des hyperliens des pages Web (resp. des titres des pages Web).

Étant donnée une collection de pages Web indexés, τ_a et τ_t sont calculées comme suit :

$$\tau_a(Q) = \frac{\sum_{t_i \in Q} \frac{nA_i}{n_i}}{\|Q\|}, \tau_t(Q) = \frac{\sum_{t_i \in Q} \frac{nT_i}{n_i}}{\|Q\|}$$

nA_i (resp. nT_i) est le nombre de pages Web contenant le terme t_i dans les textes d'ancrage de ses hyperliens (resp. dans les titres des pages), n_i est le nombre total de pages *Web* collectés contenant le terme t_i .

Pour chaque requête, nous représentons un descripteur de caractéristiques navigationnelles $QF^N(Q)$ et un descripteur de caractéristiques transactionnelles $QF^T(Q)$ en considérant que l'estimation du type de besoin par défaut est informationnel. Chaque descripteur est représenté par un vecteur de valeurs booléennes comme suit :

- $QF^N(Q) = (CF_1^{(N)}, CF_2^{(N)}, \dots, CF_6^{(N)})$, où
 - $CF_1^{(N)} = 1$ si $\|Q\| \leq 3$, 0 sinon
 - $CF_2^{(N)} = 1$ si $Vb=F$, 0 sinon
 - $CF_3^{(N)} = 1$ si $Tr=F$, 0 sinon
 - $CF_4^{(N)} = 1$ si $Ti=F$, 0 sinon
 - $CF_5^{(N)} = 1$ si $\sigma_t(Q) \geq \tau_t^*$, 0 sinon
 - $CF_6^{(N)} = 1$ si $\sigma_a(Q) \geq \tau_a^*$, 0 sinon
- $QF^T(Q) = (CF_1^{(T)}, CF_2^{(T)}, \dots, CF_6^{(T)})$, où
 - $CF_1^{(T)} = 1$ si $\|Q\| > 3$, 0 sinon
 - $CF_2^{(T)} = 1$ si $Vb=V$, 0 sinon
 - $CF_3^{(T)} = 1$ si $Tr=V$, 0 sinon
 - $CF_4^{(T)} = 1$ si $Ti=V$, 0 sinon
 - $CF_5^{(T)} = 1$ si $\sigma_t(Q) < \tau_t^*$, 0 sinon
 - $CF_6^{(T)} = 1$ si $\sigma_a(Q) < \tau_a^*$, 0 sinon

Nous spécifions que les seuils d'utilisation des termes de la requête dans les titres des pages (τ_t^*) ainsi que dans les textes des hyperliens des pages Web (τ_a^*) permettent de différencier entre les requêtes navigationnelles et transaction-

nelles. Au delà de chacun de ces seuils, la requête est considéré navigationnelle sur la base de deux suppositions :

1. Les termes de la requête navigationnelle ont des taux d'occurrence dans les textes d'ancrage et les titres des pages plus grandes que ceux de la requête transactionnelle. En effet, les mots clés des requêtes navigationnelles concernent souvent des noms des compagnies/organisation/personnes qui sont beaucoup plus distribués dans les titres des pages et les textes d'ancrage des hyperliens des pages *Web* que les termes des requêtes transactionnelles.
2. Les requêtes transactionnelles contiennent des termes spécifiques qui ne se trouvent que dans les pages cibles permettant d'effectuer la tâche transactionnelle, considérées moins nombreux que les pages navigationnelles.

4.3.1.2 Principe de l'estimation du type de besoin selon la morphologie

En se basant sur les vecteurs descriptifs de la requête (navigationnel et transactionnel), nous estimons le type de besoin derrière la requête en tant qu'informationnel, navigationnel ou transactionnel sur la base de l'hypothèse suivante : *"plus la requête satisfait les caractéristiques d'un type de besoin donné, plus la probabilité que la requête soit associée à ce type de besoin est élevée"*. Nous considérons que le type par défaut de la requête est informationnel.

Formellement, soit c_i (resp. c_j) une variable aléatoire correspondant à la catégorie C_i (resp. C_j) prenant des valeurs dans I, N, T, où I dénote une requête informationnelle, N dénote une requête navigationnelle, T dénote une requête transactionnelle. La probabilité que la requête q soit classifiée dans la classe C_i ($C_i \in \{N, T\}$) selon les descripteurs $F^N(Q)$ et $F^T(Q)$ est calculée comme suit :

$$p(q = c_i/ qf) = \begin{cases} \alpha_i * p(qf = c_i) & \text{si } c_i \in \{N, T\} \\ 1 - \sum_{c_j \in \{N, T\}} \alpha_j * p(qf = c_j) & \text{sinon} \end{cases} \quad (4.1)$$

où

$$p(qf = c_i) = \frac{\sum_{k=1..6} CF_k^{(c_i)}}{6}$$

$p(qf = c_i)$ est la probabilité que les caractéristiques morphologiques de la requête qf satisfont les caractéristiques du type c_i . α_i est un facteur dépendant de la performance de la classification de la classe C_i . Ce facteur est introduit dans le but de prendre en compte la fiabilité de la classification de chacun des types de besoins. Ce paramètre est adaptatif pouvant être calculé comme

étant la robustesse de la méthode de classification d'un type de besoin donné, par la fraction du nombre de requêtes correctement classifiées dans une classe représentant un type de besoin sur le nombre total de requêtes.

4.3.1.3 Exemple illustratif

Nous présentons un exemple illustratif de la classification d'une requête transactionnelle selon les caractéristiques linguistiques et morphologiques. Soit la requête de l'utilisateur q «*download music and lyrics*».

L'analyse lexicale de la requête par un outil d'analyse lexicale (ex. TreeTagger¹) donne la sortie suivante : *download* :verbe, *music* :nom, *and* :préposition, *lyrics* :nom. Nous extrayons ensuite les caractéristiques de la requête présentée dans le tableau 4.1 sachant que les termes «download», «music», «lyrics» sont prédéfinis dans une liste de termes transactionnels :

Type	Valeur
$\ Q\ $	4
Vb	TRUE
Tr	TRUE
Ti	FALSE
τ_t	0.0680
τ_a	0.3846

TAB. 4.1 – Exemple des caractéristiques d'une requête transactionnelle

Le vecteur descriptif de la requête $F(Q)$ est représenté comme suit :

$$F(Q) = (4, TRUE, TRUE, FALSE, 0.3846, 0.0680)$$

Supposons $\tau_a^* = 0.4419$ et $\tau_t^* = 0.1021$, nous calculons les vecteurs $QF^T(Q)$ et $QF^N(Q)$ comme suit :

$$\begin{aligned} QF^T(Q) &= (1, 1, 1, 0, 1, 1) \\ QF^N(Q) &= (0, 0, 0, 0, 0, 0) \end{aligned}$$

Nous estimons ensuite le type de besoin navigationnel, transactionnel ou informationnel selon la formule 4.1 comme suit :

$$\begin{aligned} p(Q = N/qf) &= 0 \\ p(Q = T/qf) &= 5/6 \\ p(Q = I/qf) &= 1 - (0 + 5/6) = 1/6 \end{aligned}$$

¹<http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/>

En maximisant ces valeurs de probabilités, on en déduit que la requête admet le type transactionnel.

4.3.2 Estimation du type de besoin de la requête selon le profil de la session

A ce niveau, nous estimons le type de besoin d'une requête en se basant seulement sur le profil de la session défini par le type de besoin des requêtes associées. Notre intuition à ce niveau est basée sur le fait que le type de besoin est globalement véhiculé par une séquence de requêtes inscrites dans une même tâche de recherche. Cette hypothèse est validée par différentes études des *logs* des moteurs de recherche [135], qui montrent que les sessions de recherche peuvent être catégorisées selon le type de besoin [135], comme étant informationnel, navigationnel ou transactionnel.

4.3.2.1 Principe de l'estimation du type de besoin selon le profil de la session

Sur la base de l'hypothèse de catégorisation des sessions de recherche selon le type de besoin, l'estimation du type de besoin selon le profil de la session consiste à définir une longueur moyenne des sessions pour chaque type de besoin et calculer la probabilité d'appartenance de la requête à chaque type de besoin en se basant sur le fait que : *plus le nombre de requêtes de la session se rapproche de la longueur moyenne d'une session de ce type, nous supposons que la fin de la session est plus probable dans le sens où l'utilisateur tend à basculer vers un nouveau type de besoin.*

Formellement, soit Q la requête à classifier, nous définissons le profil de la session QP^{c_j} par le type C_j des requêtes associées, noté $QP^{(c_j)} = \{Q_1, \dots, Q_{m-1}, Q_m\}$. L'estimation du type de besoin de la requête en se basant sur le profil de la session consiste principalement à calculer la probabilité de basculement vers un nouveau type et déduire ensuite la probabilité que la requête ait le même type de la session courante. Nous détaillons le calcul des probabilités associées à chacun des types de besoin en distinguant le cas où le profil de la session est plus petit ou égal à la longueur moyenne des sessions du même type :

- **Probabilité de basculement élémentaire du type c_j à un type c_i :** cette probabilité est traduite par le fait que plus la longueur du profil des requêtes de type c_j se rapproche de la longueur moyenne prédéfinie

des sessions de type c_j , plus la probabilité que l'utilisateur change le type de besoin à un nouveau type est élevée. Elle est donnée par la formule suivante :

$$p(q = c_i / qp = c_j) = \frac{\|QP^{(c_j)}\|}{w_j} \text{ si } [(c_i \neq c_j) \wedge (QP^{(c_j)} < w_j)]$$

La probabilité de basculement à un autre type, $\frac{\|QP^{(c_j)}\|}{w_j}$ est uniformément distribuée sur les deux possibilités restantes d'où la formule suivante :

$$p(q = c_i / qp = c_j) = \frac{\|QP^{(c_j)}\|}{2 * w_j} \text{ si } [(c_i \neq c_j) \wedge (QP^{(c_j)} < w_j)]$$

- **Probabilité de non-basculement** : la classification de la requête selon le type porté par le profil des requêtes de la session courante est déduite de la probabilité de basculement définie comme suit :

$$p(q = c_i / qp = c_j) = 1 - \frac{\|QP^{(c_j)}\|}{w_j} \text{ if } [(c_i = c_j) \wedge (QP^{(c_j)} < w_j)]$$

- **Équi-probabilité des types de besoins** : dans le cas où la longueur du profil de la session $\|QP^{(c_j)}\|$ devient égale à la longueur moyenne prédéfinie w_j d'une session de type c_j , l'utilisateur bascule à l'un des types de besoins de façon équitable en réinitialisant ainsi une nouvelle session. Ceci implique que la probabilité que la requête soit associée à un type de besoin c_i est égale à celles des autres types, d'où la formule suivante :

$$p(q = c_i / qp = c_j) = 1/3 \forall c_i, c_j \wedge (QP^{(c_j)} = w_j)$$

Nous donnons la définition générale du calcul de la probabilité qu'une requête q appartienne au type de besoin C_i sachant que le profil de la session qp est de type C_j comme suit :

$$p(q = c_i / qp = c_j) = \begin{cases} 1 - \frac{\|QP^{(c_j)}\|}{w_j} \text{ if } [(c_i = c_j) \wedge (QP^{(c_j)} < w_j)], \\ \frac{\|QP^{(c_j)}\|}{2 * w_j} \text{ if } [(c_i \neq c_j) \wedge (QP^{(c_j)} < w_j)], \\ 1/3 \text{ sinon} \end{cases} \quad (4.2)$$

où $\|QP^{(c_j)}\|$ est la longueur moyenne du profil de la session contenant des requêtes de type c_j , w_j est la longueur moyenne d'une session de type c_j qui peut être calculée statistiquement à partir de l'historique de recherche de l'utilisateur.

Dans le cas où la longueur moyenne du profil de la session de type c_j ne dépasse pas la longueur moyenne prédéfinie pour une session de ce type w_j , nous garantissons que la somme des probabilités de classification de la requête selon le type navigationnel, informationnel et transactionnel est égale à 1. Dans le cas contraire, la somme des probabilités de classification est garantie à 1 dans la dernière ligne de la formule.

4.3.2.2 Exemple illustratif

Dans cet exemple, nous Supposons que l'utilisateur réalise une tâche de recherche transactionnelle, et il avait émis deux requêtes lors de la session. Le profil de la session est donc défini comme suit $QP^T = \{Q1, Q2\}$. Soit une nouvelle requête $Q3$ transactionnelle émise par l'utilisateur. Nous estimons la probabilité que $Q3$ appartienne à chaque type de besoin selon la formule 4.2 étant donné la longueur du profil $|QP^T| = 3$, et la longueur moyenne des sessions transactionnelles $w_T = 5$:

$$\begin{aligned} p(Q = T/qp = T) &= 1 - 3/5 = 4/10 \\ p(Q = N/qp = T) &= 3/(2 * 5) = 3/10 \\ p(Q = I/qp = T) &= 3/(2 * 5) = 3/10 \end{aligned}$$

La valeur de probabilité maximale obtenue est $p(Q = T/qp = T)$. Par conséquent, nous admettons que le type de besoin induit par la requête $Q3$ est transactionnel. Selon le profil de la session, la requête a plus tendance de rester dans la session tant que la longueur du profil de la session n'atteint pas la longueur moyenne prédéfinie pour un certain type.

4.3.3 Classification de la requête par combinaison de la morphologie et du profil de la session

La classification des requêtes par combinaison de la morphologie et du profil de la session est basée sur l'hypothèse que les requêtes ne sont pas soumises indépendamment les unes des autres, mais elles se trouvent dans un contexte précis lié à une tâche de recherche informationnelle, navigationnelle ou transactionnelle. La classification de la requête selon le type besoin consiste donc à combiner l'estimation du type de besoin selon la morphologie (Sect. 4.3.2) et selon le profil de la session (Sect. 4.3.2).

4.3.3.1 Description de l'approche de combinaison

La prédiction du type de besoin d'une requête q en exploitant la morphologie, noté qf et le profil de la session, noté qp revient à évaluer la la probabilité que la requête soit d'un certain type de besoin selon la règle de *Bayes* : $p(q/qf, qp)$ où q , qf et qp sont des variables aléatoires prenant des valeurs dans I, N, T.

Nous constatons que la prédiction du type de la requête selon sa morphologie est indépendante de la prédiction de son type selon le profil de la session. Par conséquent les variables qf et qp sont indépendantes d'où la formule suivante :

$$p(q/qf, qp) = p(q/qf) * p(q/qp) \tag{4.3}$$

où $p(q/qf)$ est l'estimation du type de besoin de la requête selon sa morphologie et $p(q/qp)$ est l'estimation du type de besoin de la requête selon le profil de la session.

Plus précisément, nous déduisons une valeur de probabilité de classification d'une requête Q pour chaque type de besoin, C_i , sachant que le profil de la session est de type C_j comme suit :

$$p(q = c_i/qp = c_j) = \begin{cases} p(q = c_i/qf) * (1 - \frac{\|QP^{(c_j)}\|}{w_j}) & \text{si } [(c_i = c_j) \wedge (QP^{(c_j)} < w_j)] , \\ p(q = c_i/qf) * (\frac{\|QP^{(c_j)}\|}{2*w_j}) & \text{si } [(c_i \neq c_j) \wedge (QP^{(c_j)} < w_j)] , \\ p(q = c_i/qf) * 1/3 & \text{sinon} \end{cases} \quad (4.4)$$

où $p(q = c_i/qf)$ est calculée selon la formule 4.1.

La maximisation des probabilités calculées permet d'identifier précisément le type de besoin $c^*(q)$ à attribuer à la requête q comme suit : $c^*(q) = argmax_{c_i}(p(q = c_i/qf, qp))$.

4.3.3.2 Exemple illustratif

Dans cet exemple, nous supposons que le profil de la session est de type transactionnel contenant deux requêtes, et il est défini par $QP^T = \{Q1, Q2\}$. Soit une nouvelle requête à classifier définie par : $Q = \text{«download music and lyrics»}$. Étant donné $|QP^{(T)}| = 3$, la longueur moyenne des sessions transactionnelles est de $w_T = 5$ et en utilisant les valeurs de probabilités calculées dans les exemples précédents, les probabilités de classification de la requête Q dans chacun des types de besoin sont calculés selon la formule 4.4 comme suit :

$$\begin{aligned} p(Q = N) &= p(Q = N/qf) * p(Q = N/qp = T) = 0 * 3/10 = 0 \\ p(Q = T) &= p(Q = T/qf) * p(Q = T/qp = T) = 5/6 * 4/10 = 1/3 \\ p(Q = I) &= p(Q = I/qf) * p(Q = I/qp = T) = 1/6 * 3/10 = 1/20 \end{aligned}$$

La probabilité maximale calculée est $p(Q = T)$, par conséquent le type de besoin de la requête Q est transactionnel.

4.4 Evaluation expérimentale de la prédiction du type de besoin

L'évaluation de la prédiction du type de besoin consiste à mesurer la robustesse de la méthode de classification des requêtes. En effet, il s'agit de comparer

la précision de la classification obtenue automatiquement aux annotations manuelles qui assignent à chaque requête un type de besoin unique.

Notre objectif dans cette évaluation consiste à étudier l'impact de l'intégration du profil de la session en plus de l'utilisation des caractéristiques linguistiques sur la précision de la classification des requêtes. Plus précisément, nous comparons trois méthodes de classification des requêtes :

- la classification basée sur l'estimation du type de besoin selon la morphologie de la requête (QF) (cf. sect. 4.3.1 p.119),
 - notre méthode de classification basée sur l'utilisation de la morphologie de la requête et du profil de la session (QF+QP)(cf. sect. 4.3.3 p.126),
 - la classification par apprentissage supervisé obtenu par un classifieur classique, notamment *TiMBL* [52], basée sur l'exploitation seule du descripteur morphologique de la requête $F(Q)$ (cf. sect. 4.3.1.1 p.119). Cette classification a été adoptée par [95] pour la classification des requêtes selon les trois types de besoins. Le but de la comparaison de la performance de classification obtenue par *TiMBL* à nos deux méthodes de classification présente un avantage double :
1. garantir une précision de classification selon la méthode *QF* proche d'une classification de base (selon *TiMBL*, noté *baseline*) et combiner ensuite les probabilités calculées par cette méthode avec l'estimation du type de besoin selon le profil de la session (QF+QP),
 2. les résultats obtenus selon *TiMBL* servent comme *baseline* en plus de la méthode *QF* qui n'intègre aucun facteur contextuel dans la classification.

4.4.1 Description des ressources expérimentales

Nous décrivons dans cette section, la collection de test utilisée pour l'évaluation de la classification des requêtes, les outils d'extraction des caractéristiques des requêtes ainsi que le classifieur *TiMBL* adopté pour la classification des requêtes par apprentissage supervisé.

4.4.1.1 Collection de test

Dans le but d'évaluer la classification des requêtes selon le type informationnel, navigationnel ou transactionnel, nous avons utilisé trois ensembles de requêtes issus des collections de TREC, présentés dans la table 4.4.1.1. Cette collection comporte comme indiqué dans le tableau des requêtes annotées par le type de besoin informationnel issues de la collection TREC-2001 et TREC-9

et des requêtes de type navigationnel ou transactionnel issues de la collection TREC 2001.

Nous avons subdivisé l'ensemble des requêtes annotées chacune par un type de besoin en un sous-ensemble d'apprentissage et un sous-ensemble de test.

- Les requêtes d'apprentissage servent à calculer les taux d'utilisation optimaux des termes de la requête dans les titres (τ_t^*) et les textes d'ancrage des hyperliens (τ_a^*) dans les pages Web de la collection WT10G. Ces taux sont définis dans le descripteur morphologique de la requête comme étant des propriétés des requêtes navigationnelles et transactionnelles (cf. chap.4.3.1.1 sect.4.3.1.1 p.119). Elles servent également à l'apprentissage de la classification basée sur *TiMBL*.
- Les requêtes de test sont utilisées dans l'évaluation des méthodes de classification.

Type de la requête	Apprentissage	Test
<i>Informationnel</i>	TREC-9 (thème 451-500)	TREC-2001 (thème 501-550)
<i>Navigationnel</i>	TREC-2001 (thème 1-100)	TREC-2001 (thème 101-145)
<i>Transactionnel</i>	TREC-2001 (thème 20001-20100)	TREC-2001 (thème 20101-20150)

TAB. 4.2 – Choix des collections de requêtes

4.4.1.2 Outils et ressources d'extraction des caractéristiques de la requête

Nous spécifions dans cette section les outils utilisés pour extraire le descripteur morphologique de la requête (cf. chap.4.3.1.1 sect.4.3.1.1 p.119). Pour chaque requête de la collection de test Q , exprimé en langage naturel, nous calculons le descripteur morphologique $F(Q)$ comme suit :

$$F(Q) = (\|Q\|, Vb, Tr, Ti, \tau_a, \tau_t)$$

- $\|Q\|$ est calculé par le nombre des termes de la requête Q ,
- Vb désigne l'apparition d'un verbe dans la requête. Nous avons utilisé *Treetagger*, un outil linguistique, écrit en *Python*, permettant d'analyser la syntaxe grammaticale d'une phrase ou d'une expression.
- les deux caractéristiques Tr et Ti sont identifiées respectivement sur la base d'une liste prédéfinie de termes transactionnels et de termes interrogatifs. La présence ou l'absence d'un terme transactionnel (resp.interrogatif) dépend de la définition de ces listes.

- Les seuils des taux d'utilisation des termes de la requête dans les balises (τ_a et τ_t) sont calculés sur la collection WT10G, fournie dans le cadre d'évaluation TREC selon la formule 4.3.1.1 (cf. chap.4.3.1.1 sect.4.3.1.1 p.119).

Le tableau ci dessous décrit les caractéristiques de la collection WT10G indexée en utilisant le moteur de recherche *Mercur*e [29].

Nombre de documents	1,557,918
Nombre de termes	3,199,772
Longueur moyenne du document	338,6880
Nombre moyen de mots simples par document	142,3035
Taille de la collection	10 Giga-octets

TAB. 4.3 – Propriétés de la collection WT10G

4.4.1.3 Outil de classification des requêtes par apprentissage supervisé basé sur TiMBL

TiMBL [52] est un outil de classification par apprentissage supervisé permettant de classifier des objets dans des classes prédéfinies selon la méthode des K-plus proche voisins. Il implémente plusieurs algorithmes d'apprentissage basé sur le stockage de tous les exemples des classes des requêtes prédéfinies sous forme de vecteurs descriptifs. Sur la base de ces exemples, il classifie selon diverses mesures de similarité chaque nouvelle requête représentée également par un vecteur descriptif selon le type de besoin représenté par les exemples les plus proches.

4.4.2 Protocole d'évaluation

Généralement, l'évaluation de la précision de classification consiste à comparer le type de la requête obtenu automatiquement au type annoté manuellement par des utilisateurs.

Notons que la classification par apprentissage supervisé basé sur TiMBL ainsi que notre méthode de classification utilisant seulement la morphologie de la requête (QF) exploitent le descripteur morphologique de la requête. L'évaluation de ces deux méthodes de classification ne nécessite que la collection des requêtes d'apprentissage et de test. Cependant, l'évaluation de la précision de notre méthode de classification combinant la morphologie de la

requête et le profil de la session nécessite de simuler une séquence des sessions de recherche de différents types de besoins.

Le principe de l'évaluation est basé sur la méthode de la cross-validation qui consiste à subdiviser l'ensemble des requêtes en un sous-ensemble d'apprentissage et un sous-ensemble de test. Le protocole d'évaluation est basé sur deux phases : une phase d'apprentissage et une phase de test. Pour chacune des méthodes de classification, nous spécifions ces deux phases comme suit :

- Évaluation de la classification de la *baseline* selon TiMBL : étant donné l'ensemble des requêtes d'apprentissage et l'ensemble de requêtes de test. Le fonctionnement de TiMBL selon ces deux phases est décrit comme suit :

1. La phase d'apprentissage selon cette méthode consiste à sauvegarder des exemples de requêtes dans trois classes représentant les types de besoin informationnel, navigationnel et transactionnel. Chaque exemple d'une requête Q correspond à son descripteur morphologique contenant les caractéristiques de la requête ($F(Q)$).
2. La phase de test consiste à créer le descripteur morphologique de chaque requête de test et la classifier selon son degré de similarité avec les exemples de requêtes d'apprentissage sauvegardés pour chaque type de besoin.

- Évaluation de la classification basée sur la morphologie de la requête (QF) :

1. La phase d'apprentissage consiste à identifier les seuils des taux d'utilisation des termes de la requête dans les titres des pages ($\tau_t^* = 0.1021$) et dans les textes d'ancrages des hyperliens des pages ($\tau_a^* = 0.4419$). Ces taux sont calculés sur la collection WT10G comme suit :
 - Pour chaque requête d'apprentissage Q , calculer les taux τ_t et τ_a selon la formule 4.3.1.1 (cf. sect.4.3.1.1 p.119),
 - Pour chaque valeur τ_t (resp. τ_a), calculer la précision de classification $P(\tau_t)$ (resp. $P(\tau_a)$) des requêtes d'apprentissage (de type navigationnel et transactionnel) en utilisant seulement la propriété définie par le taux d'utilisation des termes dans les titres des pages (resp. le taux d'utilisation des termes dans les hyperliens des pages) comme suit :

$$P(\tau_t) = \frac{\|\{Q \in Nav/\tau_t(Q) \geq \tau_t\} \cup \{Q \in Trans/\tau_t(Q) < \tau_t\}\|}{\|\{Q/Q \in Nav \cup Trans\}\|}$$

$$P(\tau_a) = \frac{\|\{Q \in Nav/\tau_a(Q) \geq \tau_a\} \cup \{Q \in Trans/\tau_a(Q) < \tau_a\}\|}{\|\{Q/Q \in Nav \cup Trans\}\|}$$

- Identifier les taux optimaux τ_a^* et τ_t^* comme suit :

$$\tau_a^* = \operatorname{argmax}_{\tau_a} P(\tau_a), \tau_t^* = \operatorname{argmax}_{\tau_t} P(\tau_t),$$

2. La phase de test consiste à construire pour chaque requête de test deux vecteurs, navigationnel $F^N(Q)$ et transactionnel $F^T(Q)$, et calculer les valeurs de probabilités de classification de la requête dans chaque classe/type de besoin selon la morphologie selon la formule 4.1 (cf. sect.4.3.1.2 p.122). Les types attribués aux requêtes par le classifieur seront ensuite comparés aux types des requêtes prédéfinies dans la collection.
- Classification basée sur la combinaison de la morphologie de la requête et du profil de la session (QF+QP) : l'évaluation de cette méthode est basée sur une phase de test qui réutilise les valeurs de probabilités calculées dans la classification basée seulement sur la morphologie de la requête. La phase de test est basée sur les étapes suivantes :
1. Créer une séquence des sessions de recherche des trois types de besoins, issue de la collection des requêtes de test. Nous avons alterné entre les trois types de requêtes (I, N, T) où chaque session est créée en alignant aléatoirement des requêtes de même type. Pour des buts expérimentaux, nous fixons tout au long de la séquence la longueur moyenne des sessions de type informationnel (w_I), navigationnel (w_N) et transactionnel (w_T) respectivement à 5, 3 et 5. Nous avons pris $w_N = 3$ comme étant la longueur moyenne des sessions de recherche navigationnelles. Selon différentes études [89, 77], analysant le fichier *log* des moteurs de recherche, plus que 79% des sessions sont en effet composées de trois ou moins des requêtes. Puisque les sessions transactionnelles et informationnelles demandent plus d'effort et du temps pour être accomplies [135], nous avons augmenté la longueur moyenne de ces types de sessions par rapport à celle des sessions navigationnelles en $w_I = w_T = 5$. La figure 4.2 montre un exemple d'une séquence des sessions de trois types informationnel, navigationnel et transactionnel.
 2. Pour chaque requête de test sur la séquence, nous calculons sa probabilité de classification pour chaque type de besoin selon le profil de la session et selon la morphologie de la requête (formule 4.4, sect.4.3.3 p.126). Le type de besoin ayant la valeur de probabilité maximale est considéré la classe/type de besoin attribué à la requête.

4.4.3 Mesures de performance

Dans le but de mesurer la performance des méthodes de classification des requêtes, nous définissons pour chaque type de besoin, la précision et le rappel comme suit :

$$P(c) = \frac{\#classification\ correcte(c)}{\#total\ des\ requêtes}$$

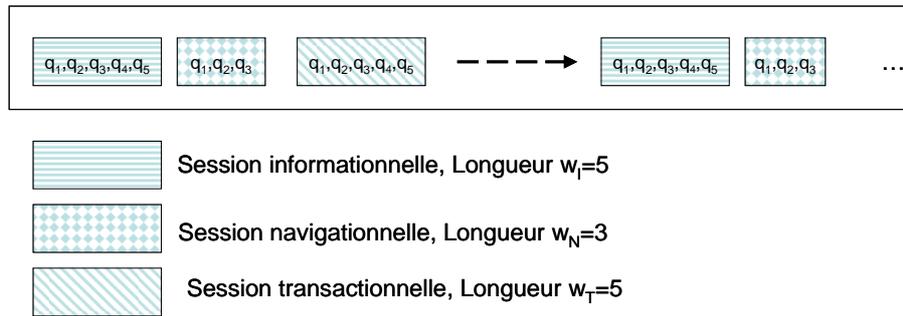


FIG. 4.2 – Exemple d’une séquence des sessions de type informationnel, navigationnel et transactionnel

$$R(c) = \frac{\#classification\ correcte(c)}{\#requêtes(c)}$$

où c est une catégorie d’une requête ayant une valeur dans $\{I, N, T\}$, $\#requêtes(c)$ est le nombre de requêtes de test de type c et $\#total$ des requêtes est le nombre total des requêtes de test. Notons que la mesure de rappel est la mesure la plus appropriée pour évaluer la robustesse d’un classifieur.

4.4.4 Résultats de la classification

Nous présentons dans le tableau 4.4.4 le pourcentage de la précision et du rappel de la classification des requêtes de test informationnelles (Info), navigationnelles (Nav) et transactionnelles (Trans) selon les trois méthodes de classification : la baseline TiMBL, notre méthode de classification basée seulement sur la morphologie de la requête (QF) et notre méthode de classification combinée (QF+QP). Nous avons également calculé la moyenne en termes de la précision et du rappel obtenues pour chacune des trois méthodes de classification.

En comparant notre approche (QF) à la classification de la baseline de TiMBL, nous pouvons constater qu’en moyenne, notre méthode (QF) donne une précision légèrement plus élevée (18.49%) que TiMBL (17.88%) et un rappel légèrement proche de TiMBL pour tous les types des requêtes. Plus précisément :

- notre méthode de classification (QF) donne une valeur de précision de classification et de rappel des requêtes navigationnelles plus élevée que TiMBL.
- TiMBL donne une précision de classification des requêtes informationnelles plus élevée que notre méthode (QF) et une précision de classification des requêtes transactionnelles légèrement plus élevée que notre méthode QF.

Nous pouvons dire que la mauvaise classification des requêtes informationnelles selon la morphologie est du probablement à leur classification comme étant le type par défaut. Cependant, sur la base de ces résultats, nous pouvons confirmer la validité de notre méthode d'estimation du type de besoin selon la morphologie (QF) comparativement à un classifieur basique d'une part et que les probabilités calculées pour chaque type de besoin selon cette méthode (QF) servent de base à l'estimation du type de besoin intégrant le profil de la session.

En comparant notre méthode de classification combinée (QF+QP) à la méthode QF et à la baseline de TiMBL, nous constatons le suivant :

- Notre méthode de classification contextuelle (QF+QP) donne la meilleure précision moyenne et rappel moyen de classification pour tous les types de requêtes comparativement à la méthode (QF) et TiMBL.
- L'amélioration par rapport à TiMBL correspond à un rappel plus élevé pour les requêtes navigationnelles et transactionnelles et une précision plus élevée pour les requêtes informationnelles et transactionnelles.

En se basant sur le rappel comme une mesure de base de comparaison entre les trois méthodes de classification, nous pouvons confirmer que l'intégration du profil de la session comme étant le type des requêtes associées permet de mieux prédire le type d'une nouvelle requête émise par l'utilisateur spécialement pour le type navigationnel et transactionnel.

En conclusion, la prédiction du type de besoin informationnel est la plus difficile vu que ce type est considéré comme étant le type par défaut et qu'il ne possède pas des caractéristiques spécifiques. L'amélioration de la qualité de la classification revient à augmenter la précision de la méthode de classification (QF) en décrivant plus des caractéristiques spécifiques liées aux différents types de besoins, spécialement pour le type informationnel.

	Info		Nav		Trans		Moyenne	
<i>Mesure</i>	<i>P</i>	<i>R</i>	<i>P</i>	<i>R</i>	<i>P</i>	<i>R</i>	<i>P</i>	<i>R</i>
Timbl (%)	10.08	50.0	29.03	72.0	14.52	36.73	17.88	52.91
QF (%)	3.24	16.33	38.87	96.0	13.36	33.67	18.49	48.67
QF+QP (%)	15.91	42.86	24.24	96.97	23.48	62.0	21.21	67.28

TAB. 4.4 – Résultats de classification des requêtes de test

Chapitre 5

Construction et évolution du profil utilisateur à base d'ontologie pour une recherche d'information personnalisée

5.1 Introduction

L'objectif de la RI personnalisée est d'intégrer le profil de l'utilisateur dans le processus de RI afin de mieux répondre à ses besoins en information spécifiques tenant compte de ses intérêts et de ses intentions de recherche.

Le chapitre 2 de l'état de l'art présenté dans la première partie de cette thèse a permis de cerner le domaine de l'accès contextuel à l'information guidé par le profil de l'utilisateur, la notion autour de laquelle s'articule la RI personnalisée. Notre contribution dans le domaine de l'accès personnalisé à l'information porte sur la conception d'un modèle d'accès personnalisé à l'information intégrant la construction de profils sémantiques des utilisateurs dans des tâches de recherche informationnelles [3, 4, 5]. Le profil est représenté par une liste de concepts ou un graphe de concepts issus de l'ontologie de l'ODP. Le profil utilisateur à court terme traduit généralement un centre d'intérêt de l'utilisateur construit sur la base d'une ou plusieurs sessions de recherche traitant d'un même besoin en information. Le profil à long terme de l'utilisateur traduit des centres d'intérêts persistants construits sur la base de récurrence des centres d'intérêts au cours des sessions de recherche de l'utilisateur. Finalement, la personnalisation du processus de RI est basée sur l'exploitation combinée du profil utilisateur à court ou à long terme dans un ordonnancement personnalisé des résultats retournés en réponse à des requêtes corrélées au sujet du profil.

Ce chapitre est organisé comme suit. La section 5.2 présente les problématiques et les motivations concernant le modèle de représentation du profil ainsi que le mode de dérivation du profil à court et à long terme. La section 5.3 présente l'architecture générale de notre modèle et la section 5.4 présente la terminologie et les notations utilisées. La section 5.5 présente les modèles de représentation du profil et le principe de délimitation des sessions de recherche permettant l'évolution du profil à court terme ainsi que le principe de définition du profil utilisateur sémantique à court et à long terme. La section 5.6 décrit la construction et l'évolution du profil de l'utilisateur à court terme et la section 5.7 décrit celle du du profil utilisateur à long terme. Finalement, l'exploitation du profil de l'utilisateur à court et à long terme par le réordonnancement des résultats de recherche est présentée dans la section 5.8. La dernière section conclut le chapitre.

5.2 Problématique et motivations

Les défis majeurs en RI personnalisée concernent principalement le modèle de représentation du profil et la gestion de l'évolution du profil. Nos objectifs dans le cadre de notre approche de personnalisation s'articulent autour de la définition d'un profil utilisateur ayant une représentation sémantique, et la construction et l'évolution de ce profil à court et à long terme dans le but de répondre à la fois à des besoins en information spécifiques ou récurrents au cours de ses sessions de recherche. Nos motivations concernant les différents éléments de notre contribution s'articulent sur les points suivants :

- Représenter le profil utilisateur par un paquet des concepts sémantiquement reliés traduisant ainsi un centre d'intérêt unique. L'objectif est dégager une représentation unique du besoin en information traitée par un ensemble de requêtes corrélées.
- Construire et faire évoluer le profil à court terme selon des mécanismes de collecte de l'information pertinente issue de l'évaluation des requêtes traitant du même besoin en information et de déploiement de ces informations au niveau spécifique du besoin de l'utilisateur dans une tâche de recherche donnée.
- Construire et faire évoluer le profil à long terme selon des mécanismes de gestion de la diversité des centres d'intérêts de l'utilisateur et l'exploiter pour répondre à des requêtes traitant d'un besoin en information récurrent. L'exploitation du profil à long terme est à la base de l'identification du centre d'intérêt le plus adéquat à une tâche de recherche initiée par une requête récurrente donnée.

Plus précisément, nos contributions portent principalement sur la proposition de deux modèles de représentation sémantiques du profil de l'utilisateur,

un mécanisme de délimitation des sessions de recherche pour l'évolution du profil à court terme dans une même session de recherche et l'utilisation d'une mesure de similarité basé-graphe pour l'identification d'un profil à long terme adéquat à une requête récurrente. Nous résumons notre orientation générale et la spécificité de notre contribution comparativement aux approches développées dans le domaine selon les deux aspects suivants :

1. Modèle de représentation du profil utilisateur :

notre approche intègre un modèle de représentation sémantique du profil de l'utilisateur basée sur l'ontologie de l'ODP. Dans les approches existantes, le profil utilisateur est représenté par une instance de l'ontologie décrivant un ou plusieurs centres d'intérêts [164, 67, 124], des paires des concepts relativement spécifiques aux requêtes de l'utilisateur, et où la hiérarchisation principale de l'ontologie est ignorée [165], ou alors un ensemble de concepts issus des trois premiers niveaux de l'ontologie utilisée invariablement entre tous les utilisateurs [123]. Ces modèles de représentation sont dans quelques approches très spécifiques à la requête en cours d'évaluation [165, 164], soient contiennent une diversité des centres d'intérêts représentés de manière éparse dans l'ontologie [67, 124, 164]. Comparativement à ces approches, nous proposons une représentation sémantique du profil de l'utilisateur basée sur une liste de concepts ou un graphe de concepts sémantiquement corrélés à un même centre d'intérêt et issus d'une ontologie de domaine largement utilisée, notamment l'ODP (*Open Directory Project*).

- (a) Le profil représenté par une liste de concepts de l'ontologie présente l'avantage de personnaliser la recherche à un niveau de spécificité commun pour l'ensemble des requêtes appartenant à une même session de recherche [6, 7, 8].
- (b) Le profil représenté par un graphe de concepts permet de grouper, dans une même structure, des concepts sémantiquement reliés et issus de différentes portions de l'ontologie, disposant ainsi de plusieurs niveaux de spécificité/généralité du profil [9, 10, 4]. Notre intuition de base concernant ce modèle consiste à renforcer la corrélation entre les concepts produisant ainsi un profil de précision élevée et répondant à des besoins en information spécifiques de l'utilisateur.

2. Évolution du profil utilisateur : notre modèle intègre la gestion de l'évolution du profil à court et à long terme.

- (a) L'évolution du profil utilisateur à court terme consiste à enrichir sa représentation sur la base des requêtes corrélées et elle est conjointement liée à un mécanisme de délimitation des sessions de recherche permettant de scruter le changement des centres d'intérêts au cours du temps.
- (b) L'évolution du profil à long terme est basée sur l'hypothèse de récurrence des besoins en information au cours des sessions de recherche[179],

qui permet de traduire une diversité des centres d'intérêts de l'utilisateur. Selon cette hypothèse, nous adoptons le principe de l'évolution du profil à long terme basée sur son augmentation par des profils construits à court terme.

Les seules approches dans le domaine qui proposent une évolution du profil à court terme dans une même session de recherche sont basées sur une mesure de similarité sémantique utilisant l'information mutuelle (*Jensen-shanon similarity*) entre des requêtes successives [162] ou une mesure statistique de changements de rangs des termes représentatifs entre des contextes de requêtes successifs en utilisant la mesure de *kendall* [179]. Quant à l'évolution du du profil à long terme, la plupart des approches d'évolution sont basées souvent sur l'accumulation de l'historique de recherche, comme dans *Google Psearch*¹ et [181], ou à la mise à jour des contextes (paire de concepts) qui sont relativement spécifiques à l'historique des requêtes soumises [165].

Comparativement à ces approches, notre méthode d'évolution du profil se distingue par les points suivants :

- (a) La détection du basculement des centres d'intérêts au cours des session de recherche est basée sur une similarité conceptuelle entre requêtes successives qui permet de scruter le changement de l'importance des concepts récurrents d'une requête à une autre. Une méthode statistique est déployée pour cela afin de détecter, en cours du temps, les différents changements des centres d'intérêts.
- (b) L'initialisation du profil utilisateur à court terme adéquat à une requête utilisateur est à la base de l'utilisation d'une mesure de similarité basée-graphe entre la requête et un profil utilisateur appris à long terme. L'utilisation de telles mesures permet de dépasser les limites de la similarité basée-terme.

5.3 Architecture du système

L'architecture générale de notre approche de personnalisation est décrite dans la figure 5.1. Cette architecture est basée sur la modélisation et l'exploitation du profil utilisateur sémantique décrivant ses centres d'intérêts à partir du *feedback* de l'utilisateur et l'ontologie de domaine de l'ODP. Nous identifions dans cette architecture les composantes suivantes : la délimitation des sessions de recherche, la construction et l'évolution du profil utilisateur à court et celui à long terme et l'accès personnalisé à l'information.

1. Délimitation des sessions de recherche : le scénario d'utilisation

¹<http://www.google.com/psearch>

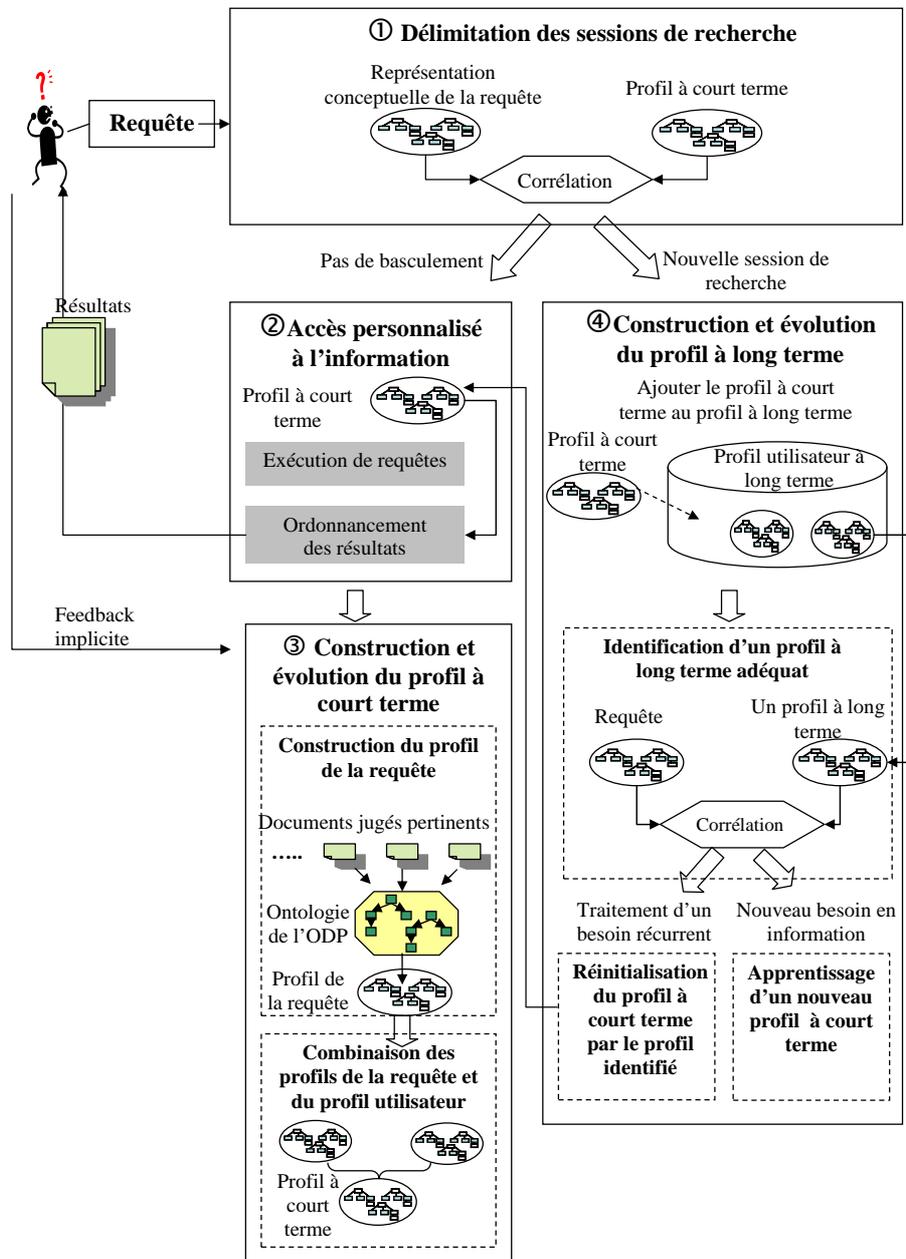


FIG. 5.1 – Architecture générale de notre modèle

du système est basé sur l'hypothèse qu'un utilisateur a divers centres d'intérêt et qu'il peut basculer d'un centre d'intérêt vers un autre au cours de ses sessions de recherche. Le basculement entre les centres d'intérêts est détecté via un mécanisme de délimitation des sessions de recherche basé sur une mesure de corrélation conceptuelle entre la requête et le profil de l'utilisateur à court terme construit dans la session de recherche courante.

2. **Construction et évolution du profil à court terme** : la construction du profil de l'utilisateur à court terme est basée sur l'exploitation

de deux sources d'évidence représentées par les documents jugés pertinents implicitement pour une requête donnée et l'ontologie de domaine, plus précisément l'ODP. Ces deux sources d'évidence sont utilisés pour construire le profil de la requête et le faire évoluer dans une même session de recherche définissant ainsi le profil à court terme.

3. **Construction et évolution du profil à long terme** : la construction du profil utilisateur à long terme consiste à agréger les profils utilisateurs à court terme construits sur la base de plusieurs sessions de recherche. Un profil à court terme peut concerner le sujet de recherche de plusieurs sessions de recherche et donc possède une représentation unique à long terme.
4. **Accès personnalisé à l'information** : le profil à court terme ou à long terme est exploité dans le processus d'ordonnancement personnalisé des résultats de recherche des requêtes traitant le même domaine d'intérêt porté par ce profil. Il s'agit d'exploiter le profil à court terme pour des requêtes inscrites dans une même session de recherche ou le profil à long terme le plus proche d'une requête initiant une nouvelle session de recherche.

5.4 Définitions de base et notations

Nous définissons dans cette section des notations utilisées dans notre approche de personnalisation. Nous nous intéressons à la définition de la notion d'une itération de recherche, une session de recherche, le profil de la requête, le profil de l'utilisateur à court terme ou à long terme ainsi que les relations entre ces différentes définitions.

– *Itération de recherche*

Une itération de recherche est définie par un ensemble d'actions impliquant différents événements permettant d'accomplir une tâche de recherche. Ces événements sont définis par la formulation d'une requête par l'utilisateur, la sélection de l'information via un processus de recherche et les interactions de l'utilisateur sur l'ensemble de résultats retournés en réponse à la requête. Les éléments définissant une itération de recherche sont les suivants :

1. la requête q^s soumise à un instant s par un utilisateur u ,
2. la liste de résultats D^s retournés par le système correspondant à la requête q^s ,
3. la sous-liste de résultats D_r^s jugés pertinents implicitement par l'utilisateur.

Un document est considéré comme pertinent s'il a été ainsi jugé par l'utilisateur de manière implicite².

– **Session de recherche**

Une session de recherche est définie par une séquence de requêtes liées à un même besoin en information. Formellement, nous définissons une session de recherche S à l'instant s par une séquence de requêtes $\{q^0, \dots, q^{s-1}, q^s\}$ émises respectivement à l'instant $\{0, \dots, s-1, s\}$.

– **Profil de la requête**

Le profil de la requête traduit les concepts d'intérêts de l'utilisateur correspondant à une requête donnée. Il est représenté à l'instant s , selon le modèle de représentation du profil, par un ensemble de concepts au niveau i d'une ontologie prédéfinie, plus précisément l'ODP, $P_{q^s} = \{C_i^1, C_i^2, \dots, C_i^n\}$ ou alors un graphe G_q^s de concepts sémantiquement reliés et issu de l'ontologie. Ce profil est construit à partir des documents jugés pertinents D_r^s retournés par le système et correspondant à la requête q^s .

– **Profil de l'utilisateur à court terme**

Le profil de l'utilisateur à court terme traduit un domaine d'intérêt de l'utilisateur construit sur la base d'une ou plusieurs sessions de recherche traitant un même besoin en information. Il est également représenté à l'instant s , par un ensemble de concepts au niveau i d'une ontologie prédéfinie, $P_{u^s} = \{C_i^1, C_i^2, \dots, C_i^n\}$ ou alors un graphe G_u^s de concepts sémantiquement reliés et issu de l'ontologie.

– **Profil de l'utilisateur à long terme**

Le profil de l'utilisateur à long terme traduit les domaines d'intérêts récurrents de l'utilisateur issus à partir de ses sessions de recherche. Ce profil comprend l'ensemble de profils récurrents construits à court terme, représentant chacun un domaine d'intérêt de l'utilisateur, noté $U = \{P_u^1, P_u^2, \dots, P_u^n\}$.

5.5 Définition du profil sémantique d'un utilisateur de SRI : approche basée sur la récurrence des sujets véhiculés par les sessions de recherche

L'approche générale de notre modèle d'accès personnalisé à l'information intègre la définition d'un profil utilisateur sémantique basée sur l'exploitation d'une ontologie de domaines, notamment l'ODP. Cette définition met en

²Documents sauvegardés et/ou imprimés et/ou satisfaisant des mesures telles que le taux de clics, le temps de lecture, etc.

évidence le modèle de représentation du profil et le principe de construction et d'évolution du profil à court et à long terme.

Nous abordons dans cette section la représentation sémantique du profil utilisateur, le principe de délimitation des sessions de recherche inhérente à l'évolution du profil à court terme dans une même session de recherche et le principe général de définition du profil à court et à long terme, détaillé selon un algorithme approprié et un exemple illustratif.

5.5.1 Représentation sémantique du profil utilisateur

Notre approche intègre deux modèles de représentation sémantique du profil utilisateur basés sur l'exploitation d'une ontologie de domaines prédéfinies, notamment l'ODP comme étant la ressource de base des domaines sémantiques. Le premier modèle est basé sur une liste de concepts et le deuxième est basé sur un graphe de concepts reliés par des liens issus de l'ontologie. Nous présentons dans cette section la représentation de l'ontologie de l'ODP et la représentation sémantique de chacun des modèles du profil utilisateur.

5.5.1.1 Représentation de la ressource sémantique : l'ontologie de l'ODP

Il existe plusieurs hiérarchies de concepts ou ontologies de domaines conçues dans le but de répertorier le contenu des pages web pour une navigation facile par les utilisateurs. On cite les portails en ligne tels que "Yahoo", "Mmagellan", "Lycos", et l'"ODP". Vu que l'ODP est le plus grand et le plus complet des répertoires du Web édités par des êtres humains ³, on l'utilise comme une source de connaissance sémantique dans le processus de construction du profil utilisateur. Notre objectif dans cette section est de représenter chaque concept de l'ODP par un ensemble de termes servant ainsi ultérieurement à la dérivation de la représentation sémantique du profil de l'utilisateur.

La figure 5.2 montre l'interface de l'ODP qui montre en gras les concepts du plus haut niveau de l'ontologie. Les données de l'ODP sont disponibles dans deux fichiers de type "RDF" : le premier contient la structure arborescente de l'ontologie de l'ODP et le deuxième liste les ressources ou les pages web associées à chacune des concepts.

- **Les concepts** : chaque concept de l'ODP représente un domaine d'intérêt des utilisateurs du web et est associée manuellement par des éditeurs à des

³<http://www.dmoz.org/World/Français/about.html>

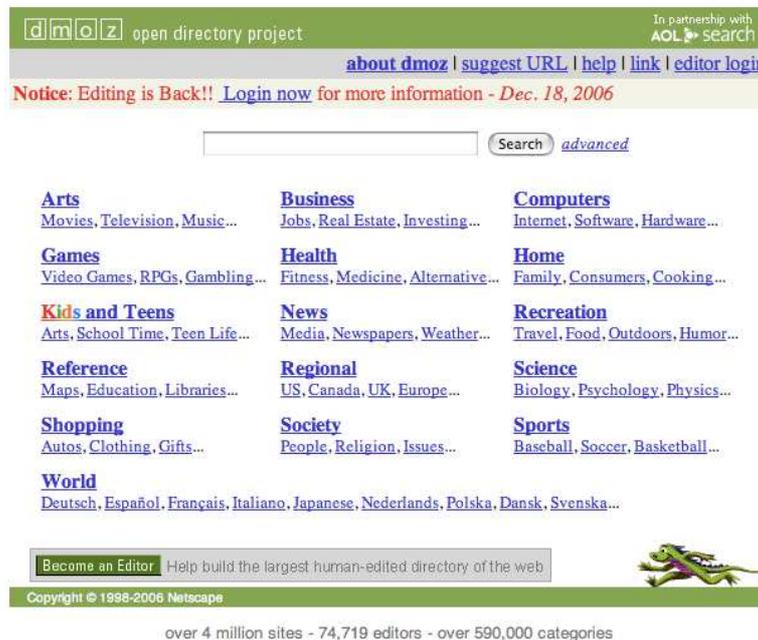


FIG. 5.2 – L’interface de l’ontologie de domaine de l’ODP

pages web dont le contenu correspond au concept. Les concepts sont organisés hiérarchiquement où les concepts de haut niveau représentent des concepts généraux et les concepts de bas niveau représentent les concepts spécifiques. Chaque concept de l’ODP est représentée par un titre et une description décrivant en général le contenu des pages web associées, et chaque page web est associée également à un titre et une description décrivant son contenu.

- **Les liens entre concepts** : Les concepts de l’ontologie sont reliés avec des relations de différents types tels que "is-a", "symbolic" et "related" ; les liens de type "is a" permettent de passer hiérarchiquement des concepts génériques aux concepts plus spécifiques. Les liens de type "symbolic" supportent la multiclassification des pages dans plusieurs concepts et permettent à l’utilisateur de naviguer entre des concepts sémantiquement liés sans avoir recours à des concepts généraux. Les liens de type "related" sont libellés par "see also" permettent de pointer vers des concepts traitant de la même thématique sans disposer des pages web en commun.
- **Représentation des concepts** : la représentation des concepts peut être obtenue selon deux méthodes :
 1. Représenter chaque concept par un vecteur de termes pondérés dont les termes sont issus des pages web associées. Cette représentation est à la base de l’utilisation d’une quantité d’information suffisante dans la représentation de chaque concept pour une bonne classification.
 2. Représenter chaque concept par le titre associé dans l’ontologie.

Les problématiques qui se posent concernant ces deux modes de

représentation sont les suivantes :

1. Dans la première méthode de représentation des concepts, certains concepts des hauts niveaux de l'ontologie ne sont pas associés à des pages Web. Le titre et la description du concept servent seulement à le décrire.
2. La deuxième méthode de représentation souffre de l'insuffisance de termes représentatifs de chaque concept (2 ou 3 mots) et nécessite une méthode de désambiguïsation efficace qui résoud les problèmes de polysémie de termes [124].

Nous adoptons la première méthode de représentation des concepts en exploitant assez bien de données pour la représentation d'un concept de l'ODP d'une part et en proposant une méthode de représentation des concepts généraux à partir des concepts spécifiques.

1. Nous exploitons pour la représentation d'un concept les 60 premiers titres et descriptions des pages Web associées. Pour valider ce choix, l'étude dans [161] a montré que l'utilisation des titres et des descriptions composés manuellement dans le répertoire du web "Looks-mart" permet d'avoir une précision de classification plus élevée que l'utilisation du contenu des pages.
2. Nous appliquons une méthode de propagation de termes à partir des concepts de bas niveaux ou spécifiques vers les concepts de haut niveau ou généraux. Cette propagation garde l'importance des termes issus des pages *Web* classifiées directement sous le concept à représenter par rapport aux termes propagés de ses concepts descendantes.

A cet effet, la représentation des concepts est obtenu selon la procédure suivante :

1. concaténer les titres et descriptions des 60 premiers liens url associées à chaque concept c_j de l'ODP dans un super-document sd_j . Nous obtenons ainsi une collection de super-documents, un par concept,
2. lemmatiser les super-documents à l'aide de l'algorithme de porter,
3. calculer la fréquence de chaque terme t_i dans le super-document Sd_j , noté tf_{ij} ,
4. appliquer une méthode de propagation des termes et de leurs scores à partir des sous-concepts vers le concept parent. Étant donné un concept c_j ayant n sous-concepts associés, noté chacun c_k , nous recalculons la fréquence d'un terme existant ou propagé t_i dans un super-document Sd_j , notée ttf_{ij} , selon la formule suivante :

$$ttf_{ij} = \left[(tf_{ij} + \sum_{k=1..n} tf_{ik}) \right] / (n + 1) \quad (5.1)$$

5. représenter chaque super-document noté sd_j par un vecteur \vec{c}_j selon le modèle vectoriel où le poids w_{ij} du terme t_i dans le super-document sd_j est calculé comme suit :

$$w_{ij} = tf_{ij} * \log\left(\frac{N}{N_i}\right) \quad (5.2)$$

Où

N =le nombre de super-documents de la collection

N_i =le nombre de super-documents contenant le terme t_i

5.5.1.2 Le profil comme une liste de concepts

Nous proposons un modèle de représentation préliminaire du profil de l'utilisateur défini au niveau i de l'ontologie par une liste de concepts, noté $P_u = \{C_i^1, C_i^2, \dots, C_i^n\}$. Notre première intuition derrière ce modèle de représentation consiste à représenter le profil utilisateur par un ensemble de concepts ayant un niveau de spécificité moyen de l'ontologie de l'ODP et à la fois efficace dans le sens où il permet d'améliorer la précision de recherche des requêtes corrélées à un même centre d'intérêt.

Bien que ce modèle de représentation du profil permet d'avoir un niveau de personnalisation efficace au niveau de la session, nous proposons d'enrichir cette représentation dans le but de recouvrir deux aspects liés principalement à la corrélation des concepts du profil et à sa spécificité.

5.5.1.3 Le profil comme un graphe de concepts

Le but fondamental de la représentation du profil de l'utilisateur sous forme d'un graphe est de décrire le profil sous forme d'un ensemble de concepts sémantiquement liés à travers l'ontologie de l'ODP et comprenant des concepts de plusieurs niveaux spécifiques.

Formellement, la structure du graphe $G=(V,E)$ est constituée d'une composante hiérarchique formée par les liens de type "is-a" et une composante non hiérarchique formée par des liens de différents types prédéfinis dans l'ontologie de l'ODP, où :

- V est un ensemble de nœuds pondérés, représentant les concepts d'intérêts de l'utilisateur,
- E est un ensemble d'arcs entre les nœuds du graphe V , partitionné en trois sous-ensembles T , S et R , tel que :

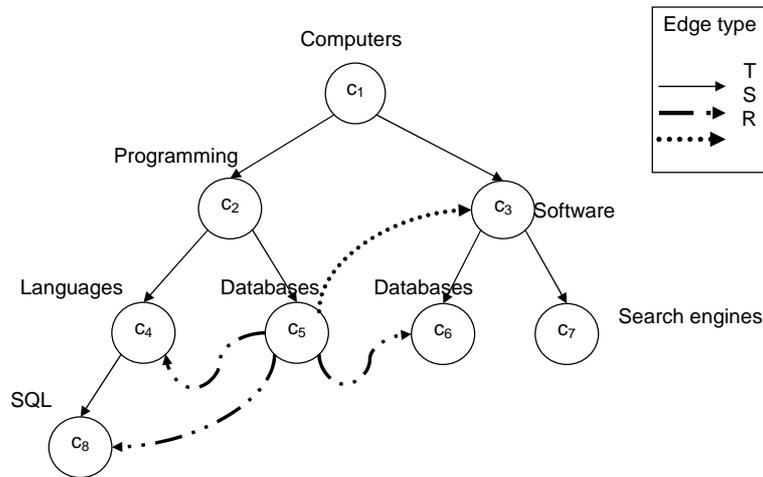


FIG. 5.3 – Une portion d’un profil utilisateur représenté sous forme d’un graphe issu de l’ODP.

- T correspond à la composante hiérarchique du profil utilisateur contenant les liens de type "is-a",
- S correspond à la composante non-hiérarchique contenant les liens de type "symbolic",
- R correspond à la composante non-hiérarchique contenant les liens de type "related".

La figure 5.3 illustre un exemple d’un profil utilisateur dérivé de l’ontologie de l’ODP et correspondant à la recherche dans le domaine *computer language programming*. Dans cet exemple, le profil utilisateur G est défini par les ensembles suivants :

$$\begin{aligned}
 V &= \{(c_1, score(c_1)), (c_2, score(c_2)), \dots, (c_8, score(c_8))\}, \\
 S &= \{(c_5, c_4), (c_5, c_8), (c_5, c_6)\}, \\
 T &= \{(c_1, c_2), (c_1, c_3), (c_2, c_4), (c_2, c_5), (c_3, c_6), (c_3, c_7), (c_4, c_8)\}, \\
 R &= \{(c_5, c_3)\}.
 \end{aligned}$$

L’exemple suivant permet de valider le modèle de représentation du profil de l’utilisateur : soit un utilisateur effectuant sa recherche s’intéresse au "soccer video games". Ce centre d’intérêt peut activer des concepts pertinents sémantiquement liés et qui peuvent être localisés dans différentes portions de l’ontologie tels que des concepts sous les concepts de niveau général *Sports and Games*. Nous constatons que les concepts non pertinents qui peuvent être obtenus parmi la liste des concepts associés à ce centre d’intérêt sont exclus vu qu’il ne disposent pas des liens sémantiques avec les concepts pertinents.

5.5.2 Principe de délimitation des sessions de recherche

Notre approche de dérivation du profil utilisateur est conjointement liée à un mécanisme de délimitation des sessions de recherche qui permet de détecter le basculement dans le sujet de la requête au cours des sessions de recherche.

La majorité des algorithmes de délimitation des sessions de recherche regroupent les requêtes selon des approches basées sur le temps [77, 87], basé contenu [73] ou sémantiques [196, 174]. Contrairement à ces approches, notre approche est basée sur l'hypothèse suivante : *la délimitation des sessions de recherche est décelée via l'observation du changement des concepts dominants d'une requête à une autre dans la session*. Sur la base de cette hypothèse, nous utilisons une représentation conceptuelle de la requête et calculons la corrélation conceptuelle avec le profil utilisateur construit dans la session de recherche. Nous constatons qu'une valeur de corrélation, notée ΔI , calculée entre une requête q^{s+1} soumise à l'instant $s + 1$ et le profil utilisateur P_u^s construit à l'instant s ($\Delta I = q^{s+1} \circ P_u^s < \sigma^*$) inférieure à un seuil de corrélation optimal prédéfini σ^* , signifie un basculement dans le sujet des requêtes. Sinon, les requêtes traitent le même centre d'intérêt général.

5.5.2.1 Représentation conceptuelle de la requête

Nous définissons une représentation duale de la requête q^{s+1} soumise à l'instant $s + 1$. La première représentation est basée sur un vecteur de termes pondérés, noté \bar{q}_t^{s+1} , où le poids d'un terme désigne sa fréquence d'apparition dans la requête.

$$\bar{q}_t^{s+1} = (t_1, t_2, t_3, \dots, t_n)$$

La deuxième représentation est basée sur un vecteur de concepts \bar{q}_c^{s+1} issu de la projection de la requête \bar{q}_t^{s+1} sur l'ontologie :

$$\bar{q}_c^{s+1} = (c_1, c_2, c_3, \dots, c_n)$$

La représentation du vecteur conceptuel de la requête est basée sur l'augmentation des concepts associés à la requête par activation des concepts reliés et leur désambiguïsation en exploitant un schéma de pondération contextuelle. Notre intuition concernant ces deux étapes est la suivant :

- L'augmentation des concepts associés à la requête est nécessaire dans le but d'activer le plus de concepts pouvant être similaires aux concepts du profil et produire ainsi une valeur de corrélation plus élevée.
- La désambiguïsation est nécessaire pour des requêtes qui peuvent être associées à la fois à des concepts pertinents et non pertinents de l'ontologie. Nous adoptons un schéma de pondération contextuelle qui permet de

désambiguïser l'ensemble des concepts associés à la requête en fonction du profil de l'utilisateur construit dans la session. Cette pondération a un double objectif :

1. Attribuer des poids plus élevés aux concepts associés à la requête et qui sont récurrents dans la session. La récurrence d'un concept est mesurée par le nombre d'apparition de ce concept dans le profil utilisateur à ses différents stades d'évolution dans la session de recherche.
2. Exclure les concepts non pertinents associés à la requête par l'atténuation de leur score.

Dans le but d'atteindre ces objectifs, la procédure du calcul du vecteur conceptuel de la requête est résumée par les étapes suivantes :

1. **Extraction des concepts de la requête** : nous projetons la requête \vec{q}_t^{s+1} sur l'ontologie selon la mesure du cosinus. Nous obtenons un ensemble de concepts c_i dont le poids est calculé comme suit :

$$CW(q_t^{s+1}, c_i) = \cos(\vec{q}_t^{s+1}, \vec{c}_i) \quad (5.3)$$

2. **Activation des concepts reliés** : nous appliquons un procédure de propagation de scores par activation des liens de référence dans l'ontologie. Cette procédure est similaire à celle détaillée dans la section 5.6.2.2.
3. **Pondération contextuelle des concepts de la requête** : les poids finaux des concepts sont recalculés selon un schéma de pondération qui introduit la fréquence des requêtes de la session indexées par un concept donné. Plus particulièrement, dans le vecteur conceptuel de la requête, noté $\vec{q}_c^{s+1} = \langle w_1, w_2, \dots, w_i, \dots \rangle$ le poids w_i d'un concept c_i est calculé en utilisant le score $CW(q_t^{s+1}, c_i)$ et la fréquence des requêtes $QF(c_i)$ indexés par le concept c_i comme suit :

$$w_i = CW(q_t^{s+1}, c_i) * QF(c_i) \quad (5.4)$$

où la fréquence des requêtes (QF) est défini formellement comme suit :

$$QF(c_i) = \frac{|\vec{q}|^S}{\langle |\vec{q}|^S, c_i \rangle} \quad (5.5)$$

$|\vec{q}|^S$ désigne le nombre total des requêtes soumises dans la session de recherche courante, notée S , $\langle |\vec{q}|^S, c_i \rangle$ désigne le nombre total des requêtes de la session ayant contribué à la considération du concept c_i dans le représentation du profil utilisateur $\{P_u^0, \dots, P_u^s\}$ à différents stades d'évolution de la session S .

5.5.2.2 Mesures de délimitation des sessions de recherche

Cette section présente deux mesures de corrélation utilisées dans notre méthode de délimitation des sessions de recherche, la mesure de *Kendall* et *WebJaccard*.

- La mesure de *Kendall* est une mesure statistique de corrélation des rangs des concepts dominants entre la représentation conceptuelle de la requête et le profil construit dans la session.
- La mesure de *WebJaccard* est basée sur le nombre de concepts communs entre la requête et le profil utilisateur.

Notre choix concernant ces deux mesures consiste à montrer le degré d'importance du rang des concepts utilisé dans *Kendall* par rapport au nombre de concepts communs utilisé dans *WebJaccard* pour scruter le changement dans le sujet de recherche dans la session.

5.5.2.2.1 Mesure de corrélation basée-Kendall Le but de cette mesure consiste à scruter le basculement des centres d'intérêts de l'utilisateur au cours des sessions de recherche en se basant sur le changement de l'importance des concepts d'une requête à une autre dans la session. Ce changement est observé par la détection d'une différence significative des rangs des concepts entre le profil de l'utilisateur et la nouvelle requête soumise. Nous définissons le degré de corrélation conceptuelle ΔI entre une nouvelle requête \vec{q}_c^{s+1} et le profil de l'utilisateur \vec{P}_u^s comme suit :

$$\Delta I = Kendall(\vec{q}_c^{s+1}, \vec{P}_u^s) = \frac{\sum_{c_i} \sum_{c_j} S_{c_i c_j}(\vec{q}_c^{s+1}) \times S_{c_i c_j}(\vec{P}_u^s)}{\sqrt{\sum_{c_i} \sum_{c_j} S_{c_i c_j}^2(\vec{q}_c^{s+1}) \times \sum_{c_i} \sum_{c_j} S_{c_i c_j}^2(\vec{P}_u^s)}} \quad (5.6)$$

$$S_{c_i c_j}(\vec{q}_c^{s+1}) = sign(\vec{q}_c^{s+1}(c_i) - \vec{q}_c^{s+1}(c_j)) = \frac{\vec{q}_c^{s+1}(c_i) - \vec{q}_c^{s+1}(c_j)}{|\vec{q}_c^{s+1}(c_i) - \vec{q}_c^{s+1}(c_j)|}$$

$$S_{c_i c_j}(\vec{P}_u^s) = sign(\vec{P}_u^s(c_i) - \vec{P}_u^s(c_j)) = \frac{\vec{P}_u^s(c_i) - \vec{P}_u^s(c_j)}{|\vec{P}_u^s(c_i) - \vec{P}_u^s(c_j)|}$$

où c_i et c_j sont deux concepts issus respectivement de la requête et du profil utilisateur, $\vec{q}_c^{s+1}(c_i)$ est le poids du concept c_i dans la requête \vec{q}_c^{s+1} , et $\vec{P}_u^s(c_i)$ est le poids du concept c_i dans le profil utilisateur \vec{P}_u^s . Pour un couple de concepts c_i et c_j , l'ordre de ces concepts l'un par rapport à l'autre est défini par $S_{c_i c_j}(\vec{q}_c^{s+1})$ et $S_{c_i c_j}(\vec{P}_u^s)$ respectivement dans la requête et le profil de l'utilisateur.

Le concept qui apparaît dans un seul vecteur aura un poids nul dans le deuxième vecteur. Par définition, cette formule permet de mesurer le degré de correspondance entre deux listes ordonnées d'objets et évalue la significativité de cette correspondance.

1. Si l'ensemble des concepts ont le même ordre dans les deux représentations, alors la correspondance entre les deux listes est parfaite (les deux listes sont similaires) et le coefficient de *Kendall* vaut 1.
2. Par opposition, si une liste a l'ordre inverse de l'autre, alors le coefficient de *Kendall* vaut -1.
3. Pour tous les autres cas, la valeur de corrélation appartient à l'intervalle de valeurs [-1 1], où des valeurs proches de 1, implique une corrélation plus significative entre les deux représentations.

5.5.2.2.2 Mesure de similarité basée-WebJaccard Nous utilisons la mesure de *WebJaccard* [76] dans le but de mesurer la similarité conceptuelle ΔI en se basant sur le nombre de concepts communs entre les concepts associés à une nouvelle requête q_c^{s+1} soumise à l'instant de recherche $s+1$ dans la session et le profil de l'utilisateur P_u^s construit jusqu'à l'instant s .

La mesure *WebJaccard* est utilisée classiquement dans la délimitation des sessions de recherche en se basant sur le feedback utilisateur [63]. Elle permet de mesurer le nombre de pages Web communs visités entre deux ensembles de résultats associés à des itérations de recherche successives. Vu que nous avons à mesurer la correspondance conceptuelle entre la requête et le profil, nous avons réutilisé la mesure *WebJaccard* en remplaçant la notion des pages Web communs par le nombre de concepts communs. Le degré de similarité conceptuelle ΔI selon *WebJaccard* est calculée comme suit :

$$\Delta I = \text{WebJaccard}(q^{s+1}, P_u^s) = \frac{H(q_c^{s+1} \cap P_u^s)}{H(q_c^{s+1}) + H(P_u^s) - H(q_c^{s+1} \cap P_u^s)} \quad (5.7)$$

où $H(q_c^{s+1} \cap P_u^s)$ est le nombre de concepts communs entre la requête et le profil, $H(q_c^{s+1})$ est le nombre de concepts associés à la requête dans q_c^{s+1} et $H(P_u^s)$ est le nombre de concepts retrouvés dans le profil P_u^s . La valeur de similarité obtenue appartient à l'intervalle [0 1], où une valeur égale à 0 implique que la requête et le profil ne sont pas similaires, et une valeur proche de 1 implique que la requête et le profil sont proches l'un de l'autre.

5.5.3 Principe de définition du profil utilisateur à court et à long terme

Nous présentons dans cette section l'algorithme générique de définition du profil à court et à long terme ainsi qu'un exemple illustratif du déroulement de l'approche de personnalisation dans une session de recherche.

5.5.3.1 Algorithme générique de définition du profil à court et à long terme

Nous détaillons le principe général de définition du profil à court et à long terme dans l'algorithme 5.5.3.1. Il s'agit de :

1. l'initialisation du profil à court terme au début d'une session de recherche traitant un nouveau centre d'intérêt.
2. la délimitation des sessions de recherche effectuée à la soumission d'une nouvelle requête. Ce test engendre un des deux cas possibles :
 - un basculement vers un nouveau centre d'intérêt à apprendre,
 - un basculement vers un centre d'intérêt récurrent.

Dans cet algorithme, le système implique le scénario suivant pour chaque requête soumise par l'utilisateur : un utilisateur u soumet une requête q^s à l'instant s au moteur de recherche ; ce dernier retourne une liste de résultats D^s parmi lesquels l'utilisateur clique sur un ensemble de résultats D_r^s qui lui semblent pertinents. Partant de ces documents, le système construit le profil de la requête selon un ensemble de concepts issus de l'ontologie de l'ODP. Nous considérons que le profil de l'utilisateur à long terme est représenté par un ensemble de profils unitaires, noté $U = \{P_u^1, P_u^2, \dots, P_u^n\}$. Le seuil de délimitation des sessions de recherche est fixé à σ^* , impliquant que deux requêtes successives sont inscrites dans la même session si la corrélation est supérieure au seuil optimal.

1. **Initialisation du profil à court terme** : A l'issue de la soumission d'une première requête q^0 d'une session traitant d'un nouveau besoin en information, le système construit le profil de la requête à partir des documents pertinents D_r^s et initialise le profil de l'utilisateur par le profil de cette première requête.
2. **Test de délimitation des sessions de recherche** : le système traite ensuite chaque nouvelle requête q^{s+1} soumise à un instant $s + 1$ dans un mécanisme de délimitation des sessions de recherche. Ce dernier est basé sur la mesure de corrélation conceptuelle ΔI entre les concepts du profil utilisateur construit jusqu'à l'instant s , noté P_u^s et les concepts associés à la nouvelle requête q^{s+1} . Deux cas peuvent être envisagés : la nouvelle requête traite le même centre d'intérêt que la session courante résultant ainsi l'évolution du profil à court terme, ou elle initie une nouvelle session de recherche résultant ainsi l'évolution du profil à long terme.
 - **la requête est inscrite dans la session courante** : si la corrélation ΔI est supérieure à un seuil optimal σ^* , on considère que la requête q^{s+1} est considérée liée au profil utilisateur. Ceci induit les deux mécanismes suivants :
 - (a) **Personnalisation du processus de recherche** : le profil utilisateur est par la suite utilisé dans le ré-ordonnement de ses résultats de recherche de la requête en cours d'évaluation,

Algorithme 1 Processus général de RI personnalisée intégrant un profil utilisateur sémantique

Notations :

$U = \{P_u^1, P_u^2, \dots, P_u^n\}$ le profil de l'utilisateur u à long terme,

σ^* le seuil de corrélation optimal de délimitation des sessions de recherche,

d^* est le seuil optimal de la distance sémantique requête-profil.

Traitement d'une nouvelle session de recherche S^i initiée par la requête q^0 :

si $U \neq \emptyset$ alors

* Extraire les concepts associés à la requête à partir de l'ontologie de l'ODP, noté q_c^0 ,

* sélectionner $P^* = \operatorname{argmax}_{P_u^i \in U} d_g(q_c^0, P_u^i)$,

si $d_g(q_c^0, P^*) < d^*$ alors

Il s'agit d'un profil préalablement découvert :

* Réinitialiser le profil utilisateur dans la session par le profil identifié :

$P_u^0 = P^*$,

* réordonner les résultats de recherche de la requête q^{s+1} avec le profil

P_u^0 ,

sinon

Il s'agit d'un nouveau profil à apprendre :

* Construire le profil de la requête P_{q^0}

* Initialiser le profil de l'utilisateur par le profil de cette requête : $P_u^0 =$

P_{q^0}

finsi

finsi

pour Chaque nouvelle requête q^{s+1} émise par l'utilisateur **faire**

Test de délimitation des sessions de recherche :

* calcul de la corrélation conceptuelle entre la requête et le profil $\Delta I = (q^{s+1} \circ P_u^s)$

si $\Delta I \geq \sigma$ alors

La requête q^{s+1} est inscrite dans la même session

* Réordonner les résultats de recherche de la nouvelle requête q^{s+1} en utilisant le profil utilisateur P_u^s

* Construire le profil de la requête $P_{q^{s+1}}$

* Mise à jour du profil utilisateur : $P_u^{s+1} = P_u^s \cup P_{q^{s+1}}$

sinon

Détection d'une nouvelle session S^{i+1} :

* Augmenter le profil à long terme par le profil à court terme créé jusqu'à l'instant s : $U = U \cup P_u^{S^s}$

Repartir au début du fonctionnement de l'algorithme.

finsi

fin pour

- (b) **Évolution du profil à court terme** : A partir du *feedback* de l'utilisateur, le système construit le profil de la nouvelle requête $P_{q^{s+1}}$ et est utilisé dans le processus d'évolution du profil utilisateur P_u^s .
- **Détection d'une nouvelle session de recherche** : si la corrélation $\Delta I < \sigma^*$, alors la requête initie une nouvelle session de recherche. Ceci contribue à l'évolution du profil à long terme par addition du profil appris P_u^s au profil à long terme et l'identification de l'une des deux cas suivants : la requête traite d'un centre d'intérêt récurrent à long terme ou un nouveau besoin en information à apprendre. L'identification de ces deux figures de cas se fait sur la base d'une mesure de similarité sémantique entre la requête et les profils préalablement appris dans U . Nous fixons un seuil de similarité optimal, noté d^* au delà duquel le profil utilisateur et la requête seront considérés non similaires.
- (a) **Traitement d'un centre d'intérêt récurrent** : si la requête est corrélée à un profil utilisateur à long terme déjà appris P^* , celui-ci sert à l'initialisation du profil de l'utilisateur à court terme dans cette nouvelle session et est utilisé dans le ré-ordonnement des résultats de recherche de cette requête.
- (b) **Traitement d'un nouveau centre d'intérêt** : si la requête n'est corrélée à aucun des profils utilisateurs déjà appris, nous constatons qu'un nouveau besoin en information est à apprendre dans la nouvelle session de recherche S^1 en repartant à l'initiation d'une session de recherche traité au début de l'algorithme.

La gestion du cycle de vie du profil à long terme est à la base de la récurrence des centres d'intérêts de l'utilisateur au cours de ses sessions de recherche d'une part et de leur fraîcheur d'autre part. Ces deux critères peuvent être pris en compte en calculant le nombre de sessions de recherche traitant un centre d'intérêt donné dans un intervalle de temps précis.

5.5.3.2 Illustration

Dans le but d'illustrer l'algorithme de définition du profil, on suppose un scénario de recherche comportant trois itérations de recherche effectuées par un utilisateur u sur une collection de documents. Nous considérons deux requêtes successives traitant du même besoin en information suivies par une requête traitant un besoin en information différent. Selon ce scénario de recherche, nous détaillons le principe de définition du profil à court terme, à partir des requêtes corrélées ainsi que le principe de base du mécanisme de délimitation des sessions de recherche.

Première itération de recherche :

La première itération de recherche est définie par la soumission à l’instant s_0 de la première requête $q^{s_0} = \text{”encephal mosquito spray”}$ dans une session de recherche. Dans cette première itération de recherche, le fonctionnement du système est décrit par les actions suivantes :

- Construction du profil de la requête q^{s_0} :
 1. A partir de l’ensemble de documents jugés pertinents par l’utilisateur, un vecteur de termes pondérés est construit, dont les termes les mieux pondérés sont représentatifs dans l’ensemble de ces documents.
 $K_{q^{s_0}} = \{\text{encephal } 5.09, \text{ mosquito } 4.12, \text{ spray } 3.51, \text{ insecticid } 2.37, \text{ malathion } 2.36, \text{ state } 2.16, \text{ louy } 2.11, \text{ fever } 2.07, \text{ outbreak } 1.83, \text{ lethargy } 1.78, \text{ headach } 1.64\}$.
 2. Le profil de la requête q^{s_0} associé est obtenu en projetant le vecteur $K_{q^{s_0}}$ sur l’ontologie de l’ODP suivi par une étape de désambiguïsation. La racine de l’ontologie est nommée ”Top”, sous laquelle se classifient les concepts hiérarchiquement. Le profil associé à la requête q^{s_0} , nommé $P_q^{s_0}$ est composé des concepts pondérés parmi lesquels nous présentons les six premiers dans le tableau 5.1.
- Initialisation du profil de l’utilisateur : Le profil de l’utilisateur est initialisé par le profil de cette requête q^{s_0} , nommé $P_u^{s_0}$,

Concept	Poids
Top/ Health/ Conditions and Diseases/ Infectious Diseases/ Viral/ West Nile Virus	0.30
Top/ Health/ Conditions and Diseases/ Neurological Disorders/Infections/Encephalitis	0.23
Top/ Health/ Conditions and Diseases/ Infectious Diseases/ Vector Borne	0.22
Top/ Recreation/ Outdoors/ Equipment/ Insect Repellent	0.22
Top/ Health/ Conditions and Diseases/ Infectious Diseases	0.22
Top/ Health/ Conditions and Diseases/ Infectious Diseases/ Viral	0.19

TAB. 5.1 – Les premiers concepts associés à la requête *encephal mosquito spray*

Deuxième itération de recherche :

La deuxième itération de recherche est définie par la soumission d’une deuxième requête $q^{s_1} = \text{”state encephal nile”}$ à l’instant s_1 traitant le même besoin en information que la requête précédente. Dans cette itération, le fonctionnement du système est décrit par les actions suivantes :

- Test de délimitation des sessions de recherche : Dans le but de tester si la requête appartient à la session de recherche courante, le système projette la requête q^{s_1} sur l’ontologie de l’ODP afin d’extraire les concepts les plus similaires à la requête et les comparer à ceux du profil. La liste des

premiers concepts pondérés les plus similaires à la requête sont présentés dans le tableau 5.2.

Concept	Poids
Top/ Health/Conditions and Diseases/ Infectious Diseases/ Viral/ West Nile Virus	0.54
Top/ Health/ Conditions and Diseases/ Neurological Disorders/ Infections/ Encephalitis	0.30
Top/ Health/ Conditions and Diseases/ Neurological Disorders/ Dementia/ AIDS Dementia Complex	0.28
Top/ Recreation/ Outdoors/ Equipment/ Insect Repellent	0.27
Top/ Science/ Biology/ Flora and Fauna/ Animalia/ Arthropoda/ Insecta/ Diptera/ Mosquitoes	0.23
Top/ Health/ Conditions and Diseases/ Infectious Diseases/ Vector Borne	0.22

TAB. 5.2 – Les premiers concepts associés à la requête *state encephal nile*

La valeur de corrélation (selon la mesure statistique de *Kendall* entre la requête q^{s_1} et le profil de l'utilisateur $P_u^{s_0}$ est de 0.10. Étant donné un seuil de délimitation des sessions de recherche égale à -0.34 , la requête est inscrite dans la session de recherche courante.

- Personnalisation de la requête q^{s_1} : Le profil de l'utilisateur $P_u^{s_0}$ est utilisé dans le réordonnancement des résultats de recherche de la requête q^{s_1} . Pour chaque document résultat de recherche, un score personnalisé est calculé basé sur la similarité du document avec les concepts du profil. Ce score est ensuite combiné avec le score original du document produisant ainsi le score final.
- Construction du profil de la requête q^{s_1} :
 1. A partir de l'ensemble de documents jugés pertinents par l'utilisateur restitués pour la requête q^{s_1} , un vecteur de termes pondérés est construit contenant les termes suivants :
 $K_{q^{s_1}} = \{\text{state } 7.38, \text{ encephal } 5.24, \text{ nile } 3.94, \text{ mosquito } 3.68, \text{ cdc } 3.22, \text{ viru } 2.86, \text{ strain } 2.41, \text{ spray } 2.33, \text{ kunjin } 2.22, \text{ pesticid } 2.21, \text{ louy } 2.05\}$.
 2. le profil de la requête q^{s_1} associé est obtenu en projetant le vecteur $K_{q^{s_1}}$ sur l'ontologie de l'ODP suivi par une étape de désambiguïsation. Le profil obtenu, nommé $P_q^{s_1}$ est composé des concepts pondérés parmi lesquels nous présentons les premiers dans le tableau 5.3.
- Évolution du profil de l'utilisateur : Cette évolution se fait par combinaison du profil de la requête q^{s_1} avec le profil $P_u^{s_0}$ produisant ainsi le profil e l'utilisateur à l'instant s_1 , noté $P_u^{s_1}$. La combinaison de ces deux profils est à la base de l'accumulation des poids des concepts communs et l'addition des nouveaux concepts issus du profil de la requête au profil de l'utilisateur $P_u^{s_1}$ (cf. sect.5.6.3).

Concept	Poids
Top/ Health/ Conditions and Diseases/ Infectious Diseases/ Viral/ West Nile Virus	0.36
Top/ Health/ Conditions and Diseases/ Infectious Diseases/ Vector Borne	0.2535
Top/ Health/ Conditions and Diseases/ Infectious Diseases/ Viral	0.2534
Top/ Health/ Conditions and Diseases/ Infectious Diseases	0.2516
Top/ Recreation/ Outdoors/ Equipment/ Insect Repellent	0.21
Top/ Health/ Conditions and Diseases/ Neurological Disorders/ Infections/ Encephalitis	0.2049
Top/ Health/ Animal/ Mammals/ Equine/ Conditions and Diseases	0.2043

TAB. 5.3 – Les premiers concepts associés au profil de la requête *state encephal nile*

Troisième itération de recherche :

La troisième itération de recherche est définie par la soumission d’une troisième requête q^{s_3} = “*earthquakes state*” à l’instant s_3 par l’utilisateur traitant un nouveau centre d’intérêt. Dans cette itération, le fonctionnement du système est décrit par les actions suivantes :

- Test de délimitation des sessions de recherche :
le système projette la requête q^{s_3} sur l’ontologie de l’ODP afin d’extraire les concepts associés. La liste des premiers concepts obtenus sont présentés dans le tableau 5.4.

Concept	Poids
Top/ Science/ Earth Sciences/ Geophysics/ Earthquakes	0.0865
Top/ Shopping/ Home and Garden/ Emergency Supplies/ Natural Disasters	0.0861
Top/ Kids and Teens/ School Time/ Science/ The Earth/ Geology/Earthquakes	0.082
Top/ Science/ Earth Sciences/ Geophysics/ Earthquakes/ Seismicity Reports/ United States	0.081
Top/ Science/ Earth Sciences/ Geophysics/ Earthquakes/ Past Earthquakes	0.079
Top/ Science/ Earth Sciences/ Geophysics/ Earthquakes/ Seismicity Reports/ United States	0.081
Top/ Science/ Earth Sciences/ Geophysics/ Earthquakes/ Past Earthquakes/ Kashmir 2005	0.078

TAB. 5.4 – Les premiers concepts associés à la requête *earthquakes state*

La valeur de corrélation de Kendall entre la requête q^{s_2} et le profil de l’utilisateur $P_u^{s_1}$ est de -0.46 . Étant donné un seuil de délimitation des

sessions de recherche égale à -0.34 , la requête initie une nouvelle session de recherche traitant un nouveau besoin en information.

- Apprentissage d'un nouveau centre d'intérêt ou identification d'un centre d'intérêt récurrent adéquat à la requête.

5.6 Construction et évolution du profil de l'utilisateur à court terme

L'approche générale de dérivation du profil utilisateur à court terme consiste à construire le profil sémantique de la requête, le désambigüiser et la faire évoluer dans une même session de recherche. La dérivation du profil à court terme est basée sur les étapes suivantes :

1. Construction du profil sémantique de la requête : la construction du profil de la requête est à la base de l'exploitation des documents jugés pertinents implicitement par l'utilisateur dans une session de recherche et une ontologie de référence, notamment l'ODP.
2. Désambigüisation du profil de la requête : cette étape consiste à extraire les concepts les plus pertinents à la requête ou augmenter leur importance dans la représentation finale de la requête.
3. Évolution du profil au cours d'une session : l'évolution du profil utilisateur à court terme est à la base de la combinaison des profils des requêtes associées dans le but d'augmenter l'importance des concepts récurrents dans la session.

Nous présentons dans cette section le principe de dérivation du profil utilisateur à court terme, selon les deux modèles de représentation, notamment le profil comme une liste de concepts et le profil comme un graphe de concepts issu de l'ontologie de l'ODP.

5.6.1 Construction du profil sémantique de la requête

Le profil de la requête est représenté selon le même modèle de représentation du profil utilisateur par une liste de concepts ou un graphe de concepts reliés. La construction du profil de la requête est basé sur deux principales étapes :

1. Extraction du vecteur de pertinence de la requête comme étant l'ensemble de termes dominants issus des documents jugés pertinents à la requête.
2. Projection du vecteur de pertinence sur l'ontologie de l'ODP et extraire les concepts les plus similaires à la requête.

- **Extraction du vecteur de pertinence de la requête** : formellement, chaque requête q^s soumise à l’instant s est associée à un ensemble de documents D^s retournés par le système et un ensemble de documents pertinents D_p^s jugés implicitement par l’utilisateur. Nous dérivons tout d’abord le vecteur de pertinence de la requête, noté K^s , comme étant un vecteur de termes les plus représentatifs dans les documents pertinents associés D_p^s . Le poids d’un terme dans le vecteur est calculé selon la formule suivante :

$$K^s(t) = \frac{1}{|D_p^s|} \sum_{d \in D_p^s} w_{td} \quad (5.8)$$

Où $w_{td} = tf_d * \log(n/n_t)$, tf_d est la fréquence du terme t dans le document d , n est le nombre total de documents dans la collection de test et n_t est le nombre de documents contenant le terme t .

- **Projection du vecteur de pertinence sur l’ontologie** : nous projetons ce vecteur sur l’ontologie de l’ODP selon la mesure de similarité vectorielle *cosinus* comme suit :

$$score(c_j) = \cos(\vec{c}_j, \vec{K}^s) \quad (5.9)$$

Nous obtenons ainsi un ensemble $\theta^s = \{(c_1, score(c_1)), \dots, (c_i, score(c_i)), \dots\}$ de concepts pondérés associés à la requête.

5.6.2 Désambiguïisation du profil de la requête

L’objectif principal de la désambiguïisation du profil de la requête est d’augmenter l’importance des concepts pertinents dans la représentation finale du profil de la requête. Nous proposons deux méthodes de désambiguïisation du profil de la requête :

1. Désambiguïisation par activation des liens hiérarchiques : cette méthode produit un profil de la requête q^s émis à l’instant s , représenté par une liste de concepts pondérés $P_{q^s} = \{(c_1, sw(c_1)), (c_2, sw(c_2)), \dots, (c_n, sw(c_n))\}$.
2. Désambiguïisation par activation des liens de référence : cette méthode produit un profil de la requête q^s émis à l’instant s , représenté par un graphe de concepts issus de l’ontologie, noté $G_q^s = (Vs_q, Es_q)$ où Vs_q représente les nœuds du graphe et Es_q représente les arcs du graphe.

5.6.2.1 Désambiguïisation par activation des liens hiérarchiques

La désambiguïisation par activation des liens hiérarchiques consiste à pondérer les concepts associés à la requête sur la base de l’hypothèse suivante : *l’importance d’un concept dans la représentativité du profil de la requête est*

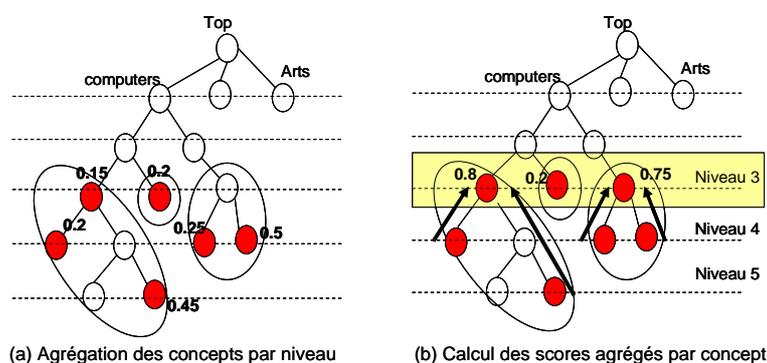


FIG. 5.4 – Désambiguïsation du profil de la requête par activation des liens hiérarchiques

déterminée récursivement par celle de ses sous-concepts via la relation "est un".

Selon cette hypothèse, nous privilégions l'apparition de groupes de concepts liés sémantiquement plutôt que des concepts éparses dans l'ontologie. A cet effet, nous adoptons un schéma d'agrégation des sous-concepts et considérons qu'un concept d'un certain niveau i de l'ontologie est plus pertinent s'il possède un plus grand groupe de sous-concepts pondérés issus de θ^s . Comme illustré sur la figure 5.4(a), nous opérons un groupage des sous-concepts sous les concepts parents du niveau 3 de l'ontologie, puis appliquons une propagation des scores jusqu'aux concepts parents (voir figure 5.4(b)). Le score de pertinence d'un parent c_j ayant un ensemble de n sous-concepts, noté $S(c_j)$, est calculé comme suit :

$$sw(c_j) = \frac{1}{n} \cdot \sum_{1 \leq k \leq n \wedge c_k \in S(c_j)} sw(c_k) \quad (5.10)$$

Finalement, le profil de la requête à l'instant s est représenté par une liste de concepts pondérés du niveau 3 de l'ontologie, noté

$$P_{q^s} = \{(c_1, sw(c_1)), (c_2, sw(c_2)), \dots, (c_n, sw(c_n))\}$$

5.6.2.2 Désambiguïsation par activation des liens de référence

Le but de cette étape est d'extraire un graphe de concepts sémantiquement liés à partir de l'ensemble des concepts associés à la requête tout en respectant la topologie du graphe de l'ontologie de l'ODP. Le principe de la désambiguïsation du profil de requête est basé sur l'hypothèse suivante : *le profil sémantique d'une requête q est un graphe de concepts reliés sans être forcément une sous-partition unique de l'ontologie, de manière à traduire la perception individuelle*

de l'utilisateur du centre d'intérêt général en cours d'identification à travers la requête q .

Le principe de cette désambiguïsation est basée sur la propagation de scores de concepts par activation des liens de référence dans l'ontologie à partir de l'ensemble de concepts associés à la requête, noté θ^s décrite dans l'algorithme 2. Nous distinguons le rôle de différents types de liens dans l'activation des liens de référence. Cette distinction a pour objectif de pondérer les concepts activés via les liens sémantiques selon leur degré de similitude avec les concepts de base dans θ^s . En effet,

1. Les liens de type *symbolic* (S) servent à la multi-classification d'une page sous plusieurs concepts. Ces liens relient des concepts effectivement corrélés et par conséquent sont traités au même niveau que les liens de type "is-a" dans l'ontologie de l'ODP.
2. Les liens de type *related* (R, libellé par "see also") relient des concepts traitant des thématiques assez différents.

A cet effet, nous utilisons la pondération des arcs adoptée dans [125]. Le poids d'un arc e_{jk} reliant le concept c_j au concept c_k est donné par $\alpha_S = 1$ si $e_{jk} \in S$ ou par $\alpha_R = 0.5$ si $e_{jk} \in R$.

Les étapes de désambiguïsation du profil de la requête sont les suivantes :

- **Augmentation des concepts par activation des liens de référence** : chaque concept c_i de θ^s propage son poids aux concepts auxquels il est lié sémantiquement, noté chacun c_k , (via les liens de type "related" et "symbolic") extraits dans une liste ℓ_j .
- **Calcul du score des concepts activés** : le poids d'un concept c_k est calculé par accumulation des poids propagés comme suit :

$$score(c_k) = \begin{cases} score(c_k) + \sum_{c_j} (\alpha_R * score(c_j)) / (\alpha_R + 1) & \text{si } e_{jk} \in R \\ score(c_k) + \sum_{c_j} (\alpha_S * score(c_j)) / (\alpha_S + 1) & \text{si } e_{jk} \in S \end{cases}$$

- **Extraction des graphes de concepts** : à partir de l'ensemble des concepts initiaux dans θ^s et activés du profil de la requête, nous reconstruisons un ou plusieurs graphes de concepts non reliés en exploitant les liens de l'ontologie. L'extraction de chaque graphe est basé sur un processus incrémental qui initie un graphe G_i par un concept initial ou activé c_i et lui rajoute parmi les concepts restants ceux qui lui sont reliés avec les arcs associés. Vu que la propagation des scores se fait à un seul saut dans l'ontologie, nous combinons ces graphes en rajoutant leurs nœuds et leurs arcs ainsi que leurs poids.
- **Sélection du graphe maximal** : finalement le profil de la requête G_q^s associé à la requête q^s est représenté par le graphe ayant le poids le plus élevé parmi les graphes créés. Nous définissons le poids $w(G_i)$ d'un graphe G_i par la somme des poids de ses nœuds :

$$w(G_i) = \sum_{c_k \in G_i} score(c_k)$$

Algorithme 2 Algorithme de propagation des scores des concepts

Entrée : θ^s est l'ensemble initial de concepts activés par l'information agrégée issue de l'évaluation de q

Sortie : $G_q^s = (V_{s_q}, E_{s_q})$ le graphe sémantique résultat
 $\theta^s = \{c_1, c_2, \dots, c_n\}$, $ListGraphs = \emptyset$

pour chaque concept $c_i \in \theta^s$ **faire**

$Queue_i = \{c_i\}$ {initialisation du graphe induit par c_i }

//initialisation du graphe induit par c_i

$G_i = (V_i, E_i)$, $V_i = V_i \cup \{c_i\}$, $E_i = \emptyset$, $w(G_i) = score(c_i)$

tantque $Queue_i.HasElement()$ **faire**

$c_j = Queue_i.PopElement()$

//extraire les liens (*is-a*, *symbolic*, *related*)

$\ell_j = GetLinkedConcepts(c_j)$

pour chaque concept $c_k \in \ell_j$ **faire**

si $e_{jk} \in S$ **alors**

$\alpha = \alpha_S$ // arc de type *symbolic*

sinon si $e_{jk} \in R$ **alors**

$\alpha = \alpha_R$ // arc de type *related*

finsi

//propagation de scores pour tous les concepts reliés

$score(c_k) = (\alpha * score(c_j) + score(c_k)) / (\alpha + 1)$

$V_i = V_i \cup c_k$, $E_i = E_i \cup e_{jk}$, $w(G_i) = w(G_i) + score(c_k)$

si $c_k \in \theta^s$ **alors**

$\theta^s = \theta^s - \{c_k\}$

$Queue.PushElement(c_k)$

finsi

fin pour

fin tantque

$ListGraphs = ListGraphs \cup \{G_i\}$

fin pour

//si deux graphes induits G_m, G_n ont des concepts communs

pour chaque $G_m, G_n \in ListGraphs$ **faire**

si $V_m \cap V_n \neq \emptyset$ **alors**

$E_m = E_m \cup E_n$, $V_m = V_m \cup V_n$, $w(G_m) = w(G_m) + w(G_n)$ // fusionner les graphes

finsi

fin pour

$G_q^s = argmax_{ListGraphs(G_i)}(w(G_i));$

5.6.3 Adaptation et évolution du profil utilisateur à court terme

L'évolution du profil utilisateur à court terme est à la base de la combinaison des profils des requêtes inscrites dans une même session de recherche. Nous proposons deux méthodes d'évolution du profil à court terme :

1. évolution du profil par combinaison linéaire des listes de concepts : le profil en question est alors représenté par une liste de concepts.
2. évolution du profil par combinaison des graphes de concepts : le profil en question est alors représenté par un graphe de concepts.

5.6.3.1 Évolution du profil par combinaison linéaire des listes de concepts

L'évolution du profil de l'utilisateur par combinaison linéaire des listes de concepts consiste à privilégier les concepts récurrents dans la session et issus de l'évaluation de la requête. Selon ce principe, l'évolution du profil à court terme consiste à définir une fonction linéaire qui consiste à :

1. augmenter le poids des concepts de la requête, récurrents dans la session via le profil à court terme,
2. atténuer le poids des concepts non récurrents.

soient P_u^{s-1} et P_q^s les représentations sémantiques associées respectivement au profil de l'utilisateur et de la requête courante, on calcule le nouveau poids d'un concept c_j dans le profil de l'utilisateur P_u^s comme suit :

$$score_{P_u^s}(c_j) = \begin{cases} \beta * score_{P_q^s}(c_j) + (1 - \beta) * score_{P_u^{s-1}}(c_j) & \text{si } c_j \in P_u^{s-1} \\ \beta * score_{P_q^s}(c_j) & \text{sinon} \end{cases} \quad (5.11)$$

où $score_{P_u^{s-1}}(c_j)$ est le poids du concept c_j dans le profil de l'utilisateur P_u^{s-1} , $score_{P_q^s}(c_j)$ est le poids du concept c_j dans le profil de la requête P_q^s .

5.6.3.2 Évolution du profil par combinaison des graphes de concepts

L'évolution du profil de l'utilisateur à court terme représenté par un graphe de concepts est basée également sur la récurrence du sujet de la requête à travers une session de recherche. Le principe de l'évolution à ce niveau consiste à rajouter au profil de l'utilisateur des nouveaux concepts d'intérêt issus du profil de la requête et renforcer le degré d'intérêt des concepts récurrents dans la session.

Soient $G_u^s = (V_u^s, E_u^s)$ et $G_q^{s+1} = (V_q^{s+1}, E_q^{s+1})$ le profil de l'utilisateur construit jusqu'à l'instant s et le profil de la requête soumise à l'instant $s + 1$, l'évolution du profil à court terme est basée sur la combinaison de ces deux profils sur la base des deux principes suivants :

- accumuler les poids des concepts communs c_i entre le profil de la requête et le profil utilisateur. Ceci permet de pondérer mieux les concepts récurrents de la session dans la représentation du profil utilisateur.

$$G_u^{s+1}(c_i) = \vec{G}_u^s(c_i) + \vec{G}_q^{s+1}(c_i)$$

où $\vec{G}_u^s(c_i)$ est le poids du concept c_i dans le profil utilisateur, $\vec{G}_q^{s+1}(c_i)$ est le poids du concept c_i dans le profil de la requête G_q^{s+1} .

- Combiner les nœuds et les arcs du profil utilisateur avec le profil de la requête dans le but de garder tous les concepts de la session ayant des degrés d'intérêt significatifs dans la représentation du profil de l'utilisateur. :

$$V_u^{s+1} = V_u^s \cup V_q^{s+1}, E_u^{s+1} = E_u^s \cup E_q^{s+1}$$

La figure 5.5 montre un exemple du profil de l'utilisateur construit dans une session et le profil d'une requête soumise par l'utilisateur et lié au même sujet de la session, notamment "computer programming languages". L'évolution du profil de l'utilisateur est basée sur l'ajout des nouveaux concepts issus du profil de la nouvelle requête et l'affectation des poids des concepts communs avec le profil de l'utilisateur.

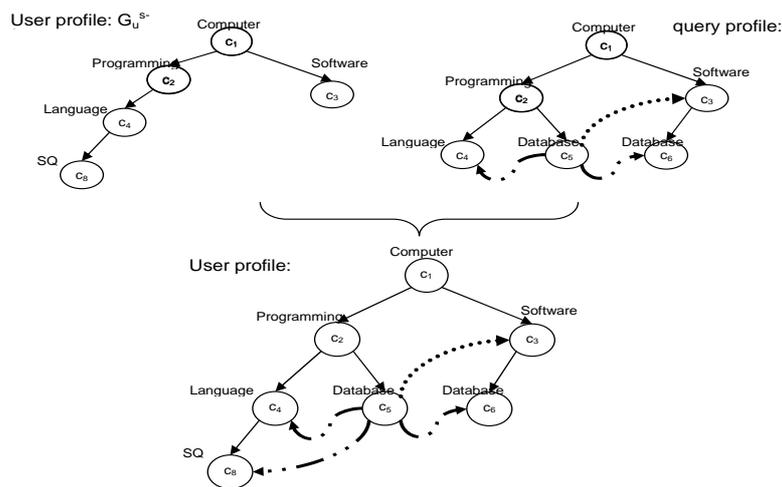


FIG. 5.5 – Évolution du profil de l'utilisateur à court terme

5.7 Construction et évolution du profil utilisateur à long terme

Le profil de l'utilisateur à long terme traduit généralement des besoins en information récurrents de l'utilisateur qualifiés par la persistance et la généralité. L'hypothèse de base de construction du profil à long terme est la suivante : *le profil à long terme contient les centres d'intérêt de l'utilisateur identifiés à partir des sujets de recherche récurrents véhiculés par les sessions de recherche.*

Sur la base de cette hypothèse, le principe de construction et d'évolution du profil à long terme est basée principalement sur l'agrégation des profils utilisateurs récurrents au cours des sessions de recherche et la gestion du cycle de vie de chacun. Plus précisément, il s'agit de trois étapes complémentaires suivantes :

1. L'addition d'un nouveau profil traduisant un nouveau besoin en information. Elle se déclenche quand l'utilisateur bascule vers un nouveau centre d'intérêt d'une session à une autre.
2. La mise à jour d'un profil à long terme préalablement appris. Elle se déclenche quand l'utilisateur bascule vers un centre d'intérêt récurrent d'une session à une autre.
3. La gestion du cycle de vie d'un profil à long terme. Il s'agit de définir une variable qui traduit la récurrence du profil à long terme au cours des sessions de recherche de l'utilisateur ainsi que sa fraîcheur.

Formellement, nous représentons le profil à long terme d'un utilisateur u par un ensemble des profils récurrents à court terme, noté $U = \{G_u^1, G_u^2, \dots, G_u^m\}$. Chaque profil à long terme G_u^i a une représentation sémantique selon un graphe de concepts.

Nous présentons dans cette section les principes généraux de construction et d'évolution du profil à long terme basé-graphe, un aperçu des métriques de distance sémantique basé-graphe et la proposition d'une mesure de distance sémantique entre graphes de profils.

5.7.1 Principes généraux

Le principe général de construction et d'évolution du profil à long terme est décrit dans l'algorithme 12. Nous traitons seulement le cas d'un basculement des centres d'intérêts d'une session à une autre permettant ainsi l'évolution du profil à long terme. La mise à jour du profil à long terme intègre au début une étape d'identification d'un profil adéquat au sujet de recherche traité par une requête récurrente. Cette identification est fondée sur l'hypothèse suivante : *une*

Algorithme 3 Gestion de l'évolution du profil à long terme

Notations :

$U = \{P_u^1, P_u^2, \dots, P_u^n\}$ le profil de l'utilisateur U , soit d^* est le seuil optimal de la distance sémantique requête-profil,

Initialisation du profil à long terme

Soit la première session de recherche S^0 :

* Construire le profil à court terme P_{S^0} sur la session S^0

* Initialiser le profil à long terme par le profil de la session $P_u^0 = P_{S^0}$

Détection d'une nouvelle session

Soient S^n et S^{n+1} deux sessions de recherche successives, S^{n+1} est initiée par la requête q^{n+1} , P_{S^n} le profil construit sur S^n

* Addition du nouveau profil P_{S^n} à long terme : $U = U \cup P_{S^n}$

* Identification d'un profil à long terme adéquat à la session S^{n+1} :

1. Calcul de la distance sémantique basé-graphe requête-profil : $G^* = \operatorname{argmax}_{P_u^i \in U} d_g(q_c^{n+1}, P_u^i) \wedge d_g(q_c^{n+1}, P_u^i) \prec d^*$,

2. Initialisation du profil à court terme dans la session S^{n+1} : $P_u^{S^{n+1}} = P^*$,

requête est similaire à un profil utilisateur à long terme si ce dernier recouvre les concepts associés à la requête à un niveau de généralité approximatif. Sur la base de cette hypothèse, nous nous basons sur une mesure de similarité basé-graphe appliquée entre le profil à long terme et la requête en se basant sur l'ontologie de l'ODP comme référentiel de base :

- Chaque profil à long terme de l'utilisateur est représenté par un graphe de concepts issu de l'ontologie de l'ODP.
- La requête est représentée également par un graphe de concepts issu de l'ontologie de l'ODP.

1. Le vecteur basé-terme de la requête \vec{q}_i est projetée sur l'ontologie de l'ODP.
2. A partir de l'ensemble des concepts obtenus, nous construisons le super-graphe de l'ontologie qui contient les concepts pondérés de la requête. Cette construction est à la base de l'activation hiérarchique ascendante des ancêtres de deux concepts de la requête formant ainsi un graphe comme étant le sous-arbre minimal qui contient tous les concepts associés à la requête.
3. Calculer une mesure de similarité basé-graphe entre chacun des profils à long terme et la requête.
4. Trier les profils à long terme dans l'ordre décroissant de leur similarité avec la requête et sélectionner le profil qui excède un seuil de similarité prédéfini.

5.7.2 Aperçu des métriques de distance basé-graphe

L'évaluation de la similarité d'objets fait le sujet de nombreuses applications comme, par exemple, la recherche d'information, la reconnaissance de formes ou le raisonnement à partir de cas. On s'intéressera dans notre travail au cas où les objets sont décrits par des graphes.

Les mesures ou distances permettant d'évaluer la similarité de graphes sont des mesures basées sur des appariements univoques, où chaque sommet est apparié à au plus un sommet de l'autre graphe (isomorphisme de (sous-)graphes, plus grand sous-graphe commun, distance d'édition de graphes).

Les premières approches d'appariement entre graphes sont limitées à trouver un isomorphisme de graphes ou sous-graphes entre deux graphes [188]. Un isomorphisme entre deux graphes est une projection bijective entre les nœuds de deux graphes qui ont le même nombre de nœuds et la même structure des arcs. Similairement, un isomorphisme de sous-graphes entre deux graphes g_1 et g_2 , est un isomorphisme entre g_1 et un sous-graphe de g_2 . L'isomorphisme de (sous-)graphes sont utiles pour tester si deux graphes sont similaires ou si un objet est présent dans un groupe de plusieurs objets. Toutefois, l'isomorphisme de (sous-)graphes est limité dans leur applicabilité dans le monde de comparaison entre objets. En effet, les objets réels sont souvent affectés par un bruit dans le sens où deux représentations d'un même objet peuvent être non similaires.

Dans le but de pallier les limitations de l'isomorphisme de (sous-)graphes, les approches ont recours à l'utilisation du plus grand sous-graphe commun entre deux graphes comme une mesure de similarité [117, 128], le plus petit super-graphe [38] et la distance d'édition [60]. La distance d'édition n'est pas seulement une mesure de similarité entre les deux graphes mais implique aussi une projection qui corrige les distorsions des graphes avec un coût minimal. Nous décrivons dans la suite les mesures de distance sémantique les plus utilisées entre graphes.

5.7.2.1 Distance basée sur le plus grand sous-graphe commun (mcs)

Le plus grand sous-graphe commun g de deux graphes g_1 et g_2 est un sous-graphe de g_1 et de g_2 et qui possède parmi tous les sous-graphes communs, le plus grand nombre de nœuds. La mesure de similarité selon le plus grand sous-graphe commun est donnée par la formule suivante :

$$d(g_1, g_2) = 1 - \frac{mcs(g_1, g_2)}{\max(|g_1|, |g_2|)} \quad (5.12)$$

Dans cette formule, $mcs(g_1, g_2)$ est le plus grand sous-graphe commun entre g_1 et g_2 , $|g_1|$ désigne le nombre de nœuds dans g_1 . Cette équation implique que

plus g_1 et g_2 sont similaires, plus large est le plus grand sous-graphe commun.

5.7.2.2 Distance basée sur le plus petit super-graphe (MCS)

Plus récemment, une nouvelle mesure de similarité entre graphes a été introduite tenant compte du plus petit super-graphe commun [38]. Un super-graphe g de deux graphes g_1 et g_2 , est le plus petit graphe qui contient g_1 et g_2 comme sous-graphes. La mesure basée sur le super-graphe entre deux graphes peut être vue comme la plus petite super-arbre de deux arbres [72]. La distance entre deux graphes g_1 et g_2 selon cette mesure est donnée par la formule suivante :

$$d(g_1, g_2) = 1 - \frac{|g_1| + |g_2| - |MCS(g_1, g_2)|}{\max(|g_1|, |g_2|)} \quad (5.13)$$

Il a été démontré que le calcul de cette mesure de similarité peut être déduite du calcul du plus grand sous-graphe commun de deux graphes [37].

5.7.2.3 Distance basée sur la combinaison du mcs et MCS

Fernandez et Valiente [62] ont proposé une mesure de distance basée sur le plus grand sous-graphe commun de deux graphes et le plus petit super-graphe de deux graphes selon la formule suivante :

$$d_{MCS}(g_1, g_2) = |MCS(g_1, g_2)| - |mcs(g_1, g_2)| \quad (5.14)$$

Dans cette mesure, les deux graphes sont plus similaires si la taille du sous-graphe commun est grande, par conséquent la taille du plus petit super-graphe est petite. A la différences des mesures basées seulement sur le mcs ou le MCS, cette mesure permet de calculer la similarité entre deux graphes selon des niveaux de généralité et de spécificité. Le degré de similarité au niveau général est calculé par le plus petit super-graphe de deux graphes et le degré de similarité au niveau spécifique est calculé via le plus grand sous-graphe commun.

5.7.2.4 Distance d'édition

Une alternative de la mesure de similarité selon mcs est basée sur la distance d'édition entre deux graphes [60]. La distance d'édition entre deux graphes est définie par la séquence la plus courte des opérations d'édition qui transforment

un graphe g_1 en un graphe g_2 . Une opération d'édition désigne en général une insertion, une suppression ou une substitution. Cette mesure est défini par :

$$d(g_1, g_2) = \min \{C(\xi)\} \quad (5.15)$$

où ξ est une séquence d'opérations des transformation de g_1 en g_2 . Une séquence d'opérations d'édition ξ de g_1 à g_2 tel que $d(g_1, g_2) = C(\xi)$ est une correction d'erreur optimale d'appariement entre graphes (*optimal error-correcting graph matching*). Il a été montré qu'il existe une relation entre le plus petit super-graphe de deux graphes et leur distance d'édition.

5.7.3 Vers une mesure de distance sémantique entre graphes de profils

Nous présentons dans cette section une mesure de distance sémantique adaptée aux graphes issus de l'ontologie de l'ODP. Le but fondamental de cette mesure de similarité consiste à identifier un profil à long terme adéquat à une requête récurrente de l'utilisateur. Dans le but d'atteindre cet objectif, le problème revient à résoudre un problème de similitude entre graphes représentant la requête d'une part et chacun des profils à long terme d'autre part.

Nous proposons l'extension de la distance basée sur la combinaison du MCS et du mcs entre graphes de profils représentant respectivement la requête et le profil à long terme. Nous présentons dans cette section nos motivations concernant le choix de cette mesure ainsi que le principe d'adaptation de cette mesure pour le calcul de la distance sémantiques entre graphes de profils.

5.7.3.1 Motivations

Nous supposons que le choix de la distance basée sur le plus petit super-graphe (MCS) et le plus grand sous-graphe commun (mcs) est la plus appropriée dans le calcul de la similarité entre les graphes de la requête est du profil. En effet, le choix de cette mesure est basée sur les principes suivants :

- Il ne s'agit pas de tolérer les erreurs de transformations des graphes ou en d'autres termes de choisir la séquence optimale des opérations d'édition qui permet de transformer un graphe en un autre mais plutôt calculer une degré de couverture de concepts entre deux graphes de profils.
- Le plus petit super-graphe des deux graphes permet de mesurer la similarité entre les deux graphes à un niveau de généralité basé sur les niveaux hauts du référentiel commun, notamment l'ontologie de l'ODP,

- Le plus petit sous-graphe commun permet de mesurer la similarité entre les deux graphes à un niveau de spécificité basé sur les niveaux les plus bas de l'ontologie de l'ODP.
- Grâce au référentiel de l'ODP, la relation entre le sous-graphe commun et le plus petit super-graphe des deux graphes est toujours valide. En effet, plus le sous-graphe entre les deux graphes est grand, plus le super-graphe qui les unit est petit.

Toutefois, l'utilisation classique de la mesure de combinaison du MCS et du mcs n'est pas suffisante dans notre cas. En effet, cette mesure donne la même distance sémantique pour deux profils qui n'ont pas des concepts communs avec la requête et dont l'un dispose des liens de référence avec certains concepts de la requête et l'autre dispose seulement des liens hiérarchiques issus du super-graphe qui le relie à la requête. Cette limite est issue du calcul du sous-graphe selon un recouvrement exact basé sur le nombre de concepts communs entre deux graphes.

Nos intuitions concernant la distance sémantique entre graphes de profils est de considérer deux types de recouvrements :

1. recouvrement exact : désigne une similarité exacte entre un sous-ensemble des concepts de la requête et ceux du profil. Ce recouvrement est traduit par le nombre de concepts communs entre les deux graphes.
2. recouvrement approximatif : désigne une similarité approximative entre un sous-ensemble des concepts de la requête et les concepts du profil. Ce recouvrement est traduit par la présence des liens de référence entre deux graphes.

Selon ces deux notions de recouvrement, la mesure de distance doit produire des valeurs croissantes selon trois configurations possibles à distinguer :

1. *CR* : les deux graphes ont des concepts en commun et/ou présentent des liens de référence entre eux. Ce cas traduit un recouvrement exact des concepts de la requête par les concepts du profil et/ou un recouvrement approximatif via les liens de référence. Nous supposons qu'il existe un et un seul graphe qui intersecte avec le graphe d'une requête récurrente et/ou présente en plus des liens de référence avec lui.
2. *NCR* : les deux graphes n'ont pas des concepts en commun tout en disposant des liens de référence entre eux. Ce cas traite un recouvrement approximatif entre la requête et le profil.
3. *NCNR* : les deux graphes n'ont pas des concepts communs ni des liens de référence qui les relient.

Dans le but de répondre à ces objectifs nous définissons une extension sémantique du chacun des graphes contribuant ainsi à l'extension sémantique du plus grand sous-graphe commun. Cette extension sémantique est à la base de l'activation des concepts reliés des deux graphes via les liens de référence. Le calcul de la distance sémantique entre graphes est basé sur les étapes suivantes :

1. étendre chacun des graphes de la requête et du profil par des concepts activés via les liens de référence,
2. calculer le sous-graphe en considérant les concepts communs et les concepts activés ainsi que les concepts intermédiaires activés qui les relient via les liens hiérarchiques,
3. calculer classiquement le super-graphe des deux graphes en exploitant les liens hiérarchiques,
4. calculer la distance sémantique sur la base de la combinaison du mcs et MCS en introduisant un facteur d'atténuation de l'importance des concepts activés par rapport aux concepts communs originaux.

Nous présentons dans la suite la définition de l'extension sémantique des graphes, ainsi que le calcul de la mesure de distance sémantique entre graphes de profils exploitant le sous-graphe commun étendu.

5.7.3.2 Extension sémantique du plus grand sous-graphe commun mcs

Soient g_1 et g_2 deux graphes issus de l'ontologie de l'ODP représentant respectivement le graphe de la requête et le graphe d'un profil à long terme. L'ensemble des concepts du graphe g_2 connectés au graphe g_1 par des liens de référence forment ce qu'on appelle l'extension du graphe g_1 par des concepts activés de g_2 . Formellement nous définissons le graphe étendu g_1^* associé à g_1 par rapport à g_2 comme suit :

$$g_1^* = g_1 \cup \{c_i \in g_2 / \exists c_j \in g_1 \wedge e_{ij} \in S \cup R\} \quad (5.16)$$

e_{ij} est l'arc reliant le concept c_i au concept c_j et $S \cup R$ est l'ensemble des liens de référence ou symboliques de l'ontologie de l'ODP. Nous obtenons ainsi deux graphes étendus g_1^* et g_2^* sur lesquels nous appliquons la mesure de la distance sémantique par combinaison du MCS et mcs.

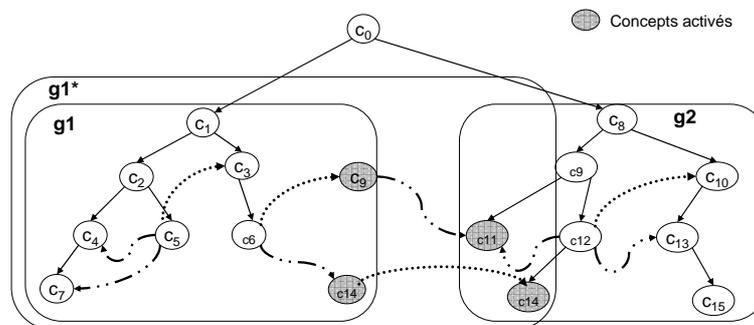


FIG. 5.6 – Extension sémantique du graphe par activation des liens de référence

5.7.3.3 La mesure de distance sémantique combinant *MCS* et *mcs*

Nous appliquons la mesure de distance sémantique entre les deux graphes augmentés g_1^* et g_2^* en utilisant la mesure basée sur la combinaison du plus grand sous-graphe commun *MCS* et le plus petit super-graphe commun *mcs*.

Formellement, la distance sémantique entre les deux graphes étendus g_1^* et g_2^* est calculée en fonction de la cardinalité du super-graphe *MCS*, du nombre de concepts communs mcs_{cc} et du nombre de concepts activés mcs_{ca} . Cette distance est donnée dans la formule suivante :

$$d_{MMCS}(g_1^*, g_2^*) = |MCS(g_1^*, g_2^*)| - (|mcs_{cc}(g_1^*, g_2^*)| - f_{ca} * |mcs_{ca}(g_1^*, g_2^*)|) \quad (5.17)$$

où $|mcs_{ca}(g_1^*, g_2^*)|$ contient les concepts activés via les liens de référence et les concepts intermédiaires qui les relient entre eux et aux concepts communs s'ils existent, f_{ca} est un facteur d'atténuation de l'importance des concepts activés dans le sous-graphe étendu introduit dans le but d'obtenir une similarité plus élevée entre deux graphes ayant des concepts communs que celle des deux graphes connectés seulement par des liens de référence. Ce facteur est défini comme suit :

$$f_{ca} = \frac{L_R}{1 + L_R}, \quad (5.18)$$

L_R désigne le nombre de liens de référence qui relient les concepts des deux graphes. La figure 5.7 illustre un exemple du plus grand sous-graphe commun entre deux graphes. Le sous-graphe reste vide dans le cas où les graphes sont déconnectés (fig.5.7 (a)). Il contiendra seulement des concepts activés et intermédiaires dans le cas où les deux graphes sont connectés seulement par des liens de référence (fig.5.7 (b)). Si les deux graphes intersectent, il contiendra des concepts communs et des concepts activés et/ou intermédiaires s'il existe des liens de référence connectant les deux graphes.

5.8 Exploitation combinée du profil utilisateur à court terme et à long terme dans la personnalisation du processus de recherche

L'hypothèse majeure de la personnalisation du processus de RI dans notre approche consiste à exploiter le profil utilisateur qui traite le même besoin en information que la requête. L'importance de cette hypothèse dans notre approche réside dans le fait que le profil utilisateur exploité, contient les concepts les plus adéquats à une recherche courante. Le système n'a pas à identifier les concepts du profil les plus similaires à chaque document résultat restitué par le système en réponse à une requête donnée.

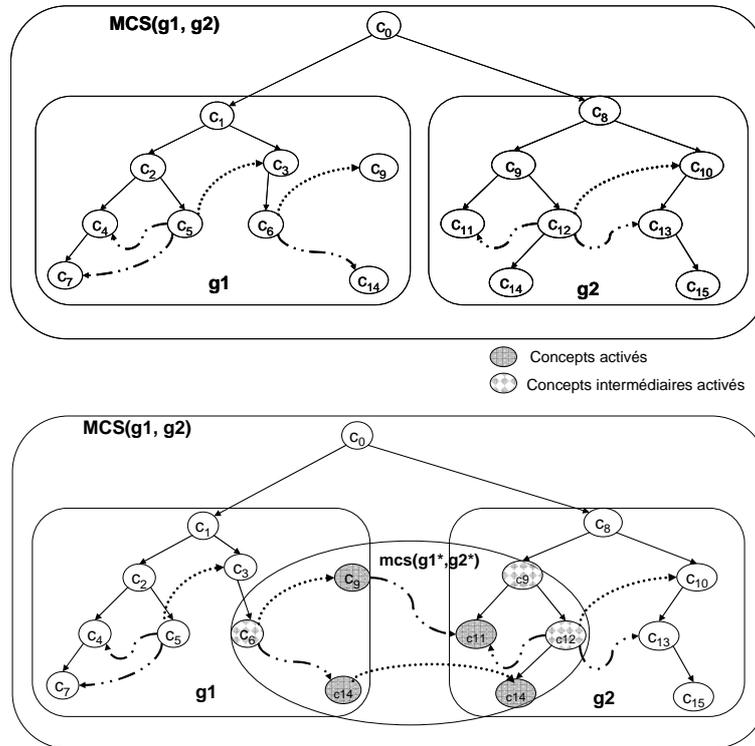


FIG. 5.7 – (a) les deux graphes n’ont pas des liens de références entre eux, (b) les deux graphes ont des liens de référence entre eux

Nous distinguons deux cas permettant d’exploiter le profil à long terme ou à court terme dans le processus de recherche en fonction de la requête émise :

1. Exploitation du profil à long terme : dans ce cas, la requête initie une session de recherche traitant un besoin en information récurrent. 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.
2. Exploitation du profil à court terme : dans ce cas, la requête est inscrite dans une session de recherche et est précédée par une séquence de requêtes traitant le même besoin en information. Le profil à court terme est ainsi construit sur la base de la séquence des requêtes précédentes.

L’exploitation du profil dans le processus de recherche consiste à ré-ordonner les résultats de recherche de la requête par combinaison du score original du document et de score personnalisé. Soit P_u^s , le profil utilisateur à exploiter dans le processus de recherche de la requête q^{s+1} , le système ré-ordonne la liste des résultats retournés pour cette requête en combinant pour chaque document d_k , son score initial, noté S_i avec son score personnalisé, noté S_p calculé comme suit :

$$S_f(d_k) = \gamma * S_i(q, d_k) + (1 - \gamma) * S_p(d_k, G_u^s) \quad (5.19)$$

$$0 < \gamma < 1$$

Si γ est égale à zéro, les résultats seront ordonnées selon le score personnalisé basé seulement sur leur similitude avec le profil utilisateur. Si γ est égale à 1, les résultats seront ordonnées selon leur score original issu du modèle de recherche classique du système. Si γ appartient à l'intervalle $[0, 1]$, les résultats seront ordonnés en tenant compte du score original du document et de son score personnalisé en accordant un degré d'importance à l'un au détriment de l'autre.

Le score personnalisé du document est calculé selon une mesure de similarité *cosine* entre le document d_k et les concepts du profil de l'utilisateur P_u^s .

$$S_p(d_k, G_u^s) = \frac{1}{h} \cdot \sum_{j=1..h} score(c_j) * \cos(\vec{d}_k, \vec{c}_j) \quad (5.20)$$

Où c_j représente un concept du profil, $score(c_j)$ est le poids du concept c_j dans le profil et h est le nombre de concepts du profil considérés dans le processus de personnalisation.

5.9 Conclusion

Nous avons présenté au cours de ce chapitre notre contribution portant sur la conception d'une approche de RI personnalisée basée sur la construction, l'évolution et l'exploitation des profils utilisateurs sémantiques à court et à long terme dans le processus de RI. Ce modèle présente les avantages suivants :

- L'avantage du modèle de représentation sémantique du profil selon un graphe de concepts est lié à sa cohérence et le niveau de spécificité de ses concepts, servant ainsi à répondre à des besoins en information spécifiques de l'utilisateur.
- La construction du profil utilisateur à court et à long terme permet respectivement de modéliser le contexte de recherche le plus adéquat à une tâche de recherche précise et répondre à des besoins en information récurrents de l'utilisateur traduisant ainsi des centres d'intérêts persistants de l'utilisateur.
- Finalement, la personnalisation du processus de recherche consiste à exploiter le profil utilisateur dans le réordonnement des résultats de recherche des requêtes traitant du même besoin en information.

L'évaluation de cette approche de personnalisation sera abordée selon deux cadres d'évaluation orientés-contexte présentés dans les deux derniers chapitre de ce mémoire.

Chapitre 6

Evaluation basée sur un cadre d'évaluation issu de TREC Adhoc : mise en œuvre et résultats

6.1 Introduction

Classiquement, l'évaluation des systèmes de recherche d'information suit le modèle de *Cranfield* [45] adopté généralement dans les campagnes d'évaluation (TREC, CLEF, INEX, etc.). Ce modèle permet de mener une étude comparative entre différents algorithmes de recherche sur la base des ressources communes et en utilisant les mêmes métriques d'évaluation (essentiellement basées sur le rappel-précision). Toutefois, ce paradigme d'évaluation n'est pas viable pour l'évaluation des SRI en présence du contexte. Le chapitre 3 présenté dans la première partie de cette thèse a permis de cerner l'évaluation des SRI en donnant un aperçu du paradigme d'évaluation classique de *Cranfield* et de ses limitations en présence du contexte, ainsi que l'émergence des approches d'évaluation orientées-contexte adaptées à l'évaluation des SRI personnalisés. La problématique de l'évaluation est d'autant plus particulière dans un SRI personnalisé basé-session où le profil est construit sur la base d'un ensemble de requêtes liées à un besoin en information.

Nous nous intéressons dans ce chapitre à présenter un cadre d'évaluation par simulation des contextes, appelé *TREC Adhoc*, préalablement proposé pour l'évaluation de l'efficacité de la recherche personnalisée basée sur l'utilisation d'un profil utilisateur, basé mots clés [180]. Nous étendons ce même cadre pour supporter un protocole d'évaluation des profils utilisateurs sémantiques,

représentés chacun par un ensemble de concepts issus d'une ontologie de domaines Web [11]. Nous présentons ensuite les résultats de l'évaluation de notre approche de personnalisation à court terme suivant ce cadre d'évaluation, plus précisément, nous évaluons la qualité des profils appris ainsi que l'efficacité de la recherche personnalisée. Finalement, nous évaluons la précision d'identification des profils à long terme adéquats à des requêtes traitant le même besoin en information et l'efficacité des ces profils dans la personnalisation du processus de RI.

Ce chapitre est organisé comme suit. La section 6.2 présente les motivations pour la proposition du cadre d'évaluation TREC Adhoc ainsi que ses composantes principales. La section 6.3 présente l'évaluation de la qualité des profils selon les deux modèles de représentation sémantique du profil, proposés dans notre approche. La section 7.4 présente l'évaluation de l'efficacité de la recherche personnalisée à court terme. La section 6.5 présente l'évaluation de la recherche personnalisée à long terme. La dernière section conclut le chapitre.

6.2 Le cadre d'évaluation issu de TREC *ad-hoc*

Ce cadre d'évaluation a été initialement défini pour l'évaluation de l'accès personnalisé guidé par le profil utilisateur, basé mots clés [180]. Il est basé sur l'augmentation d'une collection TREC par des profils utilisateur simulés dans des sessions de recherche hypothétiques. La collection utilisée est celle de la tâche Adhoc de TREC qui intègre des domaines d'intérêts définis chacun par un ensemble de requêtes traitant le sujet du domaine, simulant ainsi une session de recherche. Les interactions de l'utilisateur dans une session de recherche sont remplacées par les jugements de pertinence donnés par les assesseurs de TREC.

Nos objectifs dans ce cadre d'évaluation consistent à évaluer l'efficacité de la recherche personnalisée indépendamment du mécanisme de délimitation des sessions de recherche. Ce cadre peut être appliqué pour évaluer n'importe quel approche de personnalisation basée-session, et en utilisant toute autre collection en présence des composantes suivantes [176] :

1. une collection de test comprenant :
 - une collection de requêtes annotées d'un sujet de recherche ou d'un domaine d'intérêt,
 - une collection de documents à interroger,
 - les jugements de pertinence qui associent à chaque requête un ensemble de documents pertinents de la collection.
2. un algorithme de simulation du contexte de recherche. Un contexte de recherche est défini par une séquence hypothétique des requêtes annotées

d'un même domaine. Ce contexte permet de créer des centres d'intérêts associés aux domaines en exploitant les documents pertinents associés fournis par les assesseurs.

3. un protocole d'évaluation basé sur la validation croisée permettant de créer le profil utilisateur sur la base d'un ensemble des requêtes d'apprentissage et évaluer la performance de la RI personnalisée sur un ensemble des requêtes de test. Ceci permet d'éviter le biais de l'évaluation.
4. des métriques d'évaluation orientées rappel-précision

Nous présentons dans la suite de cette section la spécification de chacune de ces composantes ainsi que la mise en œuvre de ce cadre d'évaluation en utilisant particulièrement la collection TREC-Adhoc.

6.2.1 Collection de test

6.2.1.1 Requêtes

Nous avons utilisé les requêtes de la collection *TREC 1* numérotées de 51 à 100. Le choix de cette collection de requêtes est guidé par le fait qu'elles sont annotées d'un champ particulier noté "**Domain**" qui décrit un domaine d'intérêt traité par la requête. C'est à juste titre, cette métadonnée qui sera exploitée pour simuler des utilisateurs hypothétiques avec des centres d'intérêt issus de ces domaines. Ces requêtes couvrent six (6) domaines illustrés dans le tableau 6.1 et exploités pour nos différentes expérimentations. Le format typique d'une requête de cette collection est comme suit :

```
<top>
<head> Tipster Topic Description
<num> Number: 062
<dom> Domain: Military
<title> Topic: Military Coups D'etat
<desc> Description: Document will report a military
coup d'etat, either attempted or successful,
in any country.
<smry> Summary: Document will report a military
coup d'etat, either attempted or successful, in any country.
</top>
```

Domaines	Numéro des requêtes donné par TREC
Environment	59 77 78 83
Military	62 71 91 92
Law and Government	70 76 85 87
International Relations	64 67 69 79 100
US Economics	57 72 84
International Politics	61 74 80 93 99

TAB. 6.1 – Domaines de TREC choisies pour la simulation des profils utilisateurs

6.2.1.2 Collection de documents

La collection de test utilisée dans ce cadre est celle de la tâche ad-hoc définie dans la campagne d'évaluation *TREC 1*. Les documents de cette collection sont issus de différents articles de presse tels que *Associate Press (AP)*, *Wall street journal (WJS)*, *Financial times*. Les caractéristiques de la collection sont présentées dans le tableau 6.2.

Nombre de domaines	6
Nombre de documents	3576208
Nombre de requêtes	25
Nombre de termes distincts	589212
Longueur moyenne d'un document	53,64
Longueur moyenne d'une requête	3,5

TAB. 6.2 – Caractéristiques de la collection de test TREC 1

6.2.2 Les sessions de recherche

Une session de recherche est simulée par une séquence des requêtes annotées d'un même domaine de TREC. Le but de cette simulation est de définir un historique de recherche de l'utilisateur à court terme restreint à un nombre limité des requêtes traitant le même sujet d'intérêt. Cet historique permet de créer un profil utilisateur qui sera exploité dans la personnalisation du processus de recherche des requêtes de test. Il peut être créé sur la base de différentes combinaisons d'un sous ensemble de requêtes d'un même domaine.

6.2.3 Le profil utilisateur

Le profil utilisateur est un élément intégré dans la collection de test selon un algorithme de simulation qui le génère à partir des requêtes annotées du même domaine décrit comme suit :

1. pour chaque domaine k de la collection (noté Dom^k avec $k = (1..6)$), nous sélectionnons, parmi les n requêtes associées à ce domaine, un sous-ensemble de $n - 1$ requêtes qui constitue l'ensemble d'apprentissage d'un profil utilisateur,
2. à partir de cet ensemble d'apprentissage, un processus automatique se charge de récupérer, la liste des vecteurs associés aux documents pertinents de chaque requête,
3. partant de ces vecteurs documents, le processus de construction du profil utilisateur est déployé sur cet ensemble de requêtes. Un vecteur basé mots clés appelé vecteur de pertinence de la requête est construit puis projeté sur l'ontologie de l'ODP aboutissant à la construction du profil de la requête (cf. chap.5 sect. 5.6.1 p.157). Puis un processus de construction du profil utilisateur permet de le définir par combinaison des profils des requêtes d'apprentissage.

6.2.4 Protocole d'évaluation

Le protocole d'évaluation concerne l'évaluation de la qualité du profil et l'efficacité de RI personnalisée à court terme. La qualité des profils est évalué au niveau général, ou en d'autres termes par rapport aux domaines simulés de la collection et l'efficacité de la RI personnalisée est évaluée selon une stratégie basée sur la validation croisée intégrant ainsi deux phases : une phase d'apprentissage et une phase de test.

6.2.4.1 Protocole d'évaluation de la qualité du profil

L'évaluation de la qualité du profil est effectuée au niveau général en considérant chaque domaine de la collection de *TREC Adhoc* comme étant un centre d'intérêt à évaluer. La démarche d'évaluation consiste à créer les profils sémantiques associés aux domaines simulés de la collection et utiliser des utilisateurs réels (cinq thésards dans le domaine de la RI) qui jugent la qualité de ces profils selon le critère de pertinence suivant : "‘un concept est pertinent pour un domaine s'il est lié à la thématique d'une ou plusieurs requêtes du domaine’". Cette démarche est décrite comme suit :

- Pour chaque domaine de la collection de TREC, créer le profil par combinaison des profils de toutes les requêtes du domaine.
- En considérant les premiers X concepts du profil, cinq utilisateurs réels juge chaque concept par rapport au domaine associé et l’assigne l’un des deux labels, pertinent ou non-pertinent. Chaque utilisateur doit juger le concept non seulement en considérant le nom du domaine mais aussi les requêtes et leurs descriptions. Nous obtenons ainsi cinq annotations manuelles données par les cinq utilisateurs.
- Le jugement final d’un concept du profil est basé sur le vote majoritaire de tous les utilisateurs. Si le concept est jugé pertinent par la majorité des utilisateurs, il sera donc considéré pertinent et inversement.

Le tableau 6.3 montre un exemple typique d’un profil associé au domaine *Environment* de la collection de TREC Adhoc. Nous présentons dans cet exemple, le nom du domaine, les requêtes associées avec leur description et les premiers 10 concepts du profil à juger.

Domain	Environment
Requêtes	Weather Related Fatalities Poaching Greenpeace Measures to Protect the Atmosphere
Concepts	Top/Society/Issues/Environment/Climate Change Top/Science/Environment/Global Change/Carbon Management Top/Society/Issues/Environment/Climate Change/News Top/Science/Environment/Global Change Top/Science/Environment/Air Quality/Emissions Trading Top/Science/Environment/Agriculture/Livestock Methane Emissions Top/Society/Issues/Environment/Ozone Depletion Top/Society/Issues/Environment/News and Media Top/News/Weather/Hazards and Extremes Top/Society/Issues/Environment/Conservation and Endangered Species/Trade in Wildlife and Plants

TAB. 6.3 – Exemple du profil sémantique obtenu pour le domaine *Environment* à juger par des utilisateurs réels

6.2.4.2 Protocole d'évaluation de la RI personnalisée

La stratégie d'évaluation de l'efficacité de la recherche personnalisée consiste en un scénario qui se base sur la méthode de la validation croisée et ce, pour ne pas biaiser les résultats avec un seul jeu de test. Nous considérons ici que les sessions de recherche sont définies préalablement par l'ensemble de requêtes annotées de domaines de TREC. La validation croisée [130] ou la *k-fold cross validation* est une méthode d'évaluation qui consiste à faire varier l'ensemble des requêtes d'apprentissage servant de base à la construction d'un profil utilisateur, en l'alternant itérativement avec l'ensemble des requêtes à tester. Le profil appris à partir des requêtes d'apprentissage, et utilisé dans la personnalisation du processus de RI de la requête de test peut être vu comme un profil appris à court terme dans une session définie par les requêtes d'apprentissage et ensuite exploité dans la personnalisation de la requête de test considérée ainsi inscrite dans la session. La figure 6.1 illustre le protocole d'évaluation basé sur la stratégie de la validation croisée. Cette stratégie consiste en les étapes suivantes :

- Subdiviser l'ensemble des n requêtes du domaine en un sous-ensemble d'apprentissage de $n-1$ requêtes et en un sous-ensemble de test contenant la $n^{ième}$ requête à tester,
- Apprendre le profil utilisateur à partir des requêtes d'apprentissage soit P , en exploitant un sous-ensemble de documents pertinents par requête.
- Tester les requêtes de test avec P .

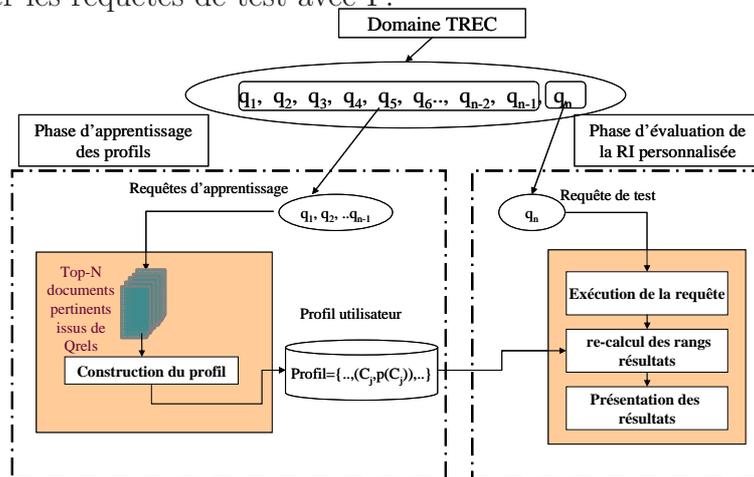


FIG. 6.1 – Protocole d'évaluation basée sur la stratégie de la cross-validation appliqué sur TREC adhoc

6.2.5 Métriques d'évaluation

Nous présentons deux classes de métriques d'évaluation dans ce cadre, les métriques d'évaluation de la qualité des profils sémantiques représentés par un

ensemble de concepts et ceux de l'évaluation de l'efficacité de recherche.

6.2.5.1 Métriques d'évaluation de la qualité du profil

Nous adoptons deux mesures pour l'évaluation de la qualité du profil :

- La précision aux premiers X concepts donné par la formule suivante :

$$P_p X = \frac{\# \text{nombre de concepts pertinents}}{X} \quad (6.1)$$

- Le rang moyen des concepts non-pertinents dans les premiers X concepts, donné par la formule suivante :

$$\text{AvgRank}_p X = \frac{\sum \text{rang des concepts nonpertinents}}{\# \text{nombre de concepts non pertinents}} \quad (6.2)$$

6.2.5.2 Métriques d'évaluation de l'efficacité de recherche

Dans ce cadre d'évaluation, nous avons utilisé les mesures classiques de la précision pour les X premiers documents restitués (PX) et la MAP (Mean Average Precision) calculées sur l'ensemble des requêtes de test. Ces deux métriques permettent d'évaluer les performances de recherche selon deux aspects :

- PX : est la proportion de documents pertinents dans les X premiers documents retrouvés. Elle permet d'exprimer la satisfaction de l'utilisateur vis-à-vis des X premiers résultats pertinents. Elle constitue ainsi une mesure importante pour l'évaluation de la haute précision. Dans notre cas, on retient les précisions pour les 10, 20, ..., 100 premiers documents notés respectivement P10, P20, ..., P100.
- MAP : est la précision moyenne pour l'ensemble des documents pertinents retournés. Elle exprime la capacité du modèle à sélectionner les documents pertinents en réponse à toutes les requêtes expérimentées.

6.3 Évaluation de la qualité du profil

Nous nous intéressons dans cette section à évaluer la qualité du profil représenté par la liste des premiers 10 concepts du graphe extrait de l'ontologie de l'ODP selon notre méthode de construction du profil (cf. chap.5 sect.5.5.1.3 p.145). Nous avons choisi d'évaluer la qualité du profil à 10 concepts puisque le nombre de concepts optimal utilisés dans le calcul du score personnalisé du document (cf. chap.5 sect.5.8 formule 5.20) est identifié expérimentalement à moins de 10 concepts (cf. section 6.4.3.1 p.190).

Pour chaque domaine de la collection, le profil associé est construit par combinaison des profils de ses requêtes en exploitant le nombre de documents pertinents optimal identifié expérimentalement à 25 permettant d’atteindre la meilleure personnalisation (cf. sect. 6.4.3.1 p.190).

La figure 6.2 montre l’évaluation de la qualité du profil en termes de la précision P_p10 et le rang moyen des concepts non-pertinents sur les premiers 10 concepts du profil $AvgRank_p10$. Les résultats montrent que nous obtenons plus de 50% de précision sur les Top-10 concepts du profil et un rang moyen minimal de 4. D’après ces résultats, nous pouvons confirmer l’exactitude des premiers concepts des profils exploités dans le processus de RI personnalisé. Cette exactitude est due à l’efficacité de la méthode de dérivation du profil et la précision de son modèle de représentation au moins sur les dix concepts les mieux pondérés.

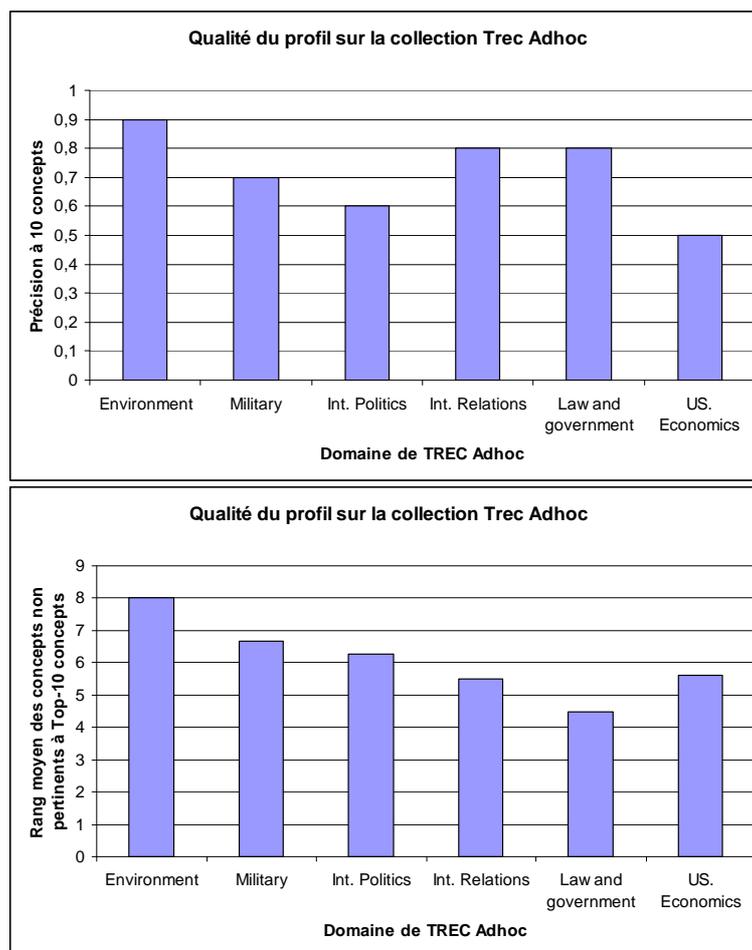


FIG. 6.2 – Evaluation de la qualité des profils graphes construits sur la collection TREC adhoc

6.4 Évaluation de l'efficacité de recherche la personnalisée à court terme

L'évaluation à ce niveau consiste à comparer la performance de recherche personnalisée intégrant le profil dans le processus de recherche à la performance de recherche classique suivant la stratégie de test du protocole d'évaluation présentée dans la section 6.2.4.2.

Nous présentons dans cette section le modèle de recherche classique qui représente la *baseline* de nos expérimentations et l'évaluation de l'efficacité de recherche personnalisée selon les deux modèles de représentation du profil, le profil comme une liste de concepts et le profil comme un graphe de concepts (cf. chap.5 sect.5.5.1.2 et 5.5.1.3 p.145).

6.4.1 Modèle de recherche classique basé sur BM25

Nous avons mené nos expérimentations sur le moteur de recherche "Mercury" [29]. Le modèle de recherche classique est basé sur la fonction d'appariement BM25 donnée dans la formule suivante :

$$w_{td} = tf_d \times \frac{\log\left(\frac{n-n_t+0.5}{n+0.5}\right)}{K_1 \times ((1-b) + b \times \frac{dl}{avgdl}) + tf} \quad (6.3)$$

où tf_d est la fréquence du terme t dans le document d , n est le nombre total des documents de la collection de test et n_t est le nombre de documents contenant le terme t , $K_1 = 2$ and $b = 0.75$.

6.4.2 Evaluation basée sur le profil comme une liste de concepts

Nous présentons dans cette section l'évaluation de l'efficacité de la recherche personnalisée en exploitant le profil comme une liste de concepts. Les objectifs de l'évaluation concerne l'étude de différents paramètres dépendant du processus de construction du profil utilisateur d'une part et de la méthode de personnalisation d'autre part. les objectifs de l'évaluation sont les suivants :

- Étudier l'impact du nombre de documents pertinents optimal utilisé dans la construction du profil de la requête (cf. chap.5 formule 5.6.1 p.157),
- Étudier l'impact du paramètre de combinaison du score original du document avec le score personnalisé (cf. chap.5 formule 5.19 p.172) sur la performance de la recherche personnalisée.

- Évaluer l’impact de la profondeur des concepts sur la performance de la recherche personnalisée en fonction du niveau des concepts du profil.
- Évaluer la performance de la recherche personnalisée comparativement à la recherche classique.

Dans toutes ces études, nous calculons du score personnalisé du document en utilisant les trois premiers concepts du profil ($h = 3$, formule 5.20, cf. chap.5 sect. 5.8 p.171). Notre choix concernant ce nombre se base sur différentes études [165] [123], qui ont montré que l’utilisation de trois concepts est suffisant pour la désambiguïsation de la recherche sur le Web. Ensuite, l’évolution du profil est effectuée par combinaison linéaire des profils des requêtes d’apprentissage en utilisant ($\beta = 0.2$, formule 5.11) fixé pour des buts expérimentaux.

6.4.2.1 Étude de l’impact du nombre de documents pertinents dans la construction du profil

Nous étudions dans cette expérimentation l’impact du nombre de documents pertinents servant à la construction du profil de la requête sur la performance de la recherche personnalisée. Cette étude est effectuée en appliquant le protocole d’évaluation basée sur la validation croisée (cf. sect. 6.2.4.2 p.180). Le modèle de la recherche personnalisée dans cette étude est basé seulement sur le score personnalisé du document.

Nous présentons dans la figure 6.3 le pourcentage du gain de performance de la recherche personnalisée à P5, P10 et la MAP en fonction du nombre de documents pertinents pris dans l’ensemble des valeurs $\{5, 10, 15, 20, 25, 30\}$. Nous constatons que l’efficacité de la recherche personnalisée est optimale en utilisant 20 documents pertinents pour les précisions P5, P10 et MAP, montrant que l’utilisation d’une plus grande quantité d’information pertinente permet d’obtenir la meilleure performance de personnalisation. Cette constatation traduit en partie la difficulté des requêtes de la collection, nécessitant ainsi assez bien de données d’apprentissage pour la construction des profils efficaces.

6.4.2.2 Étude de l’impact du paramètre de réordonnement des résultats

Le but de cette expérimentation est d’étudier l’impact du paramètre de réordonnement des résultats de recherche γ sur la performance de la recherche personnalisée. Ce paramètre sert à combiner le score original du document avec son score personnalisé.

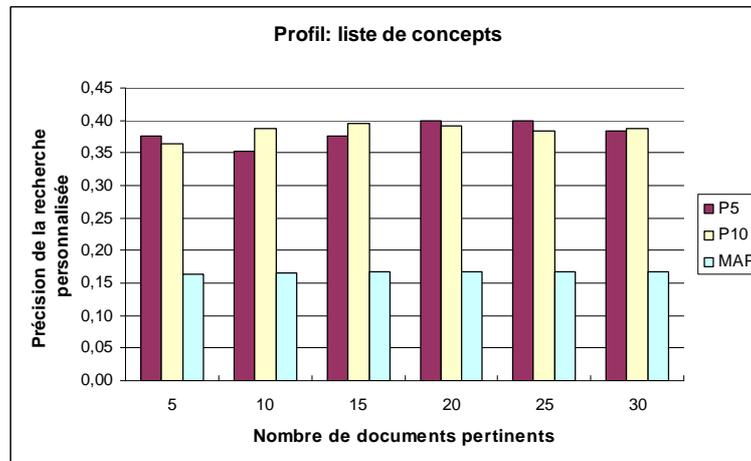


FIG. 6.3 – Variation de l’efficacité de la recherche personnalisée en fonction du nombre de documents pertinents pour le profil conceptuel ensembliste sur la collection TREC Adhoc

Nous avons fixé le nombre de documents pertinents utilisé pour la construction du profil de la requête d’un domaine à 20, la valeur optimale identifiée dans l’expérimentation précédente et avons calculé la précision de la recherche personnalisée à P10 et MAP pour chacune des valeurs de γ dans l’intervalle $[0 \ 1]$. Pour $\gamma = 0$, il s’agit de la recherche personnalisée basée seulement sur le score personnalisé du document. Pour $\gamma \in [0 \ 9]$, il s’agit de la recherche personnalisée basée sur la combinaison du score original du document et de son score personnalisé. Pour $\gamma = 0$, il s’agit de la recherche classique.

La variation de la précision de la recherche est présentée dans la figure 6.4. Nous constatons que la meilleure valeur de γ est identifiée à 0.2 pour la P10 et 0.3 pour la MAP. A partir des résultats présentés dans ces figures, nous constatons que :

- L’utilisation seule du score personnalisé du document ($\gamma = 0$) donne une précision faible à P10 et MAP. L’utilisation de cette valeur est biaisée dans le sens où les concepts du profil sont spécifiques à l’ensemble des requêtes d’apprentissage du même domaine et non pas à la requête en cours d’évaluation.
- Les courbes obtenues pour les précisions P10 et P20 sont décroissantes à partir de la valeur 0.2. Plus particulièrement, l’augmentation de l’importance du score personnalisé ($\gamma < 0.5$) permet d’obtenir une efficacité de la recherche personnalisée plus élevée à P10, P20 et P30, comparativement à l’efficacité de la recherche personnalisée basée sur l’attribution des poids élevés au score original du document.

La meilleure performance est obtenue pour une contribution faible du score original du document ($\gamma = 0.2$) en privilégiant la précision maximale obtenue à P10. Ceci montre que le score personnalisé du document contribue significativement dans l’augmentation de la performance de la recherche personnalisée

comparativement au score original du document. L'avantage de la combinaison du score original avec le score personnalisé du document permet de prendre en compte la pertinence des documents vis à vis la requête d'une part (via le score original) et leur pertinence vis à vis le profil de l'utilisateur d'autre part (via le score personnalisé).

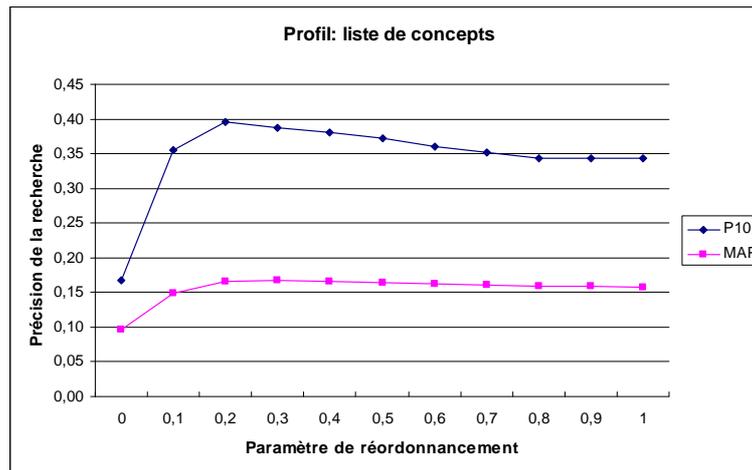


FIG. 6.4 – Variation de l'efficacité de la recherche personnalisée en fonction du paramètre de réordonnement des résultats pour le profil représenté par une liste de concepts sur la collection TREC Adhoc

6.4.2.3 Étude de l'impact de la profondeur du profil

Cette étude porte sur l'impact du niveau des concepts du profil sur la performance de la recherche personnalisée. Pour cela, nous calculons la variation de la performance de la recherche personnalisée obtenue pour le profil défini par une liste de concepts du niveau 3 de l'ontologie (cf. chap.5 sect.5.5.1.2 p.145), comparé à deux autres profils définis également par des listes de concepts pondérés par accumulation des poids des concepts issus des profils des requêtes d'un même domaine simulé. L'accumulation des poids des concepts pour un même domaine traduit l'importance des concepts récurrents sur l'ensemble des requêtes de ce domaine. Plus précisément ces deux profils sont définis comme suit :

- Le premier profil est représenté par tous les concepts d'un sous-arbre maximal, pondérée par accumulation des poids des concepts issus des profils des requêtes d'un même domaine simulé. Un sous-arbre est dit maximal si la somme des poids de ses concepts est la plus élevée parmi les autres sous-arbres.
- un profil représenté par les concepts de niveau 3 et 4 d'un sous-arbre maximal, pondérée par accumulation des poids des concepts issus des profils des requêtes d'un même domaine. Nous n'avons pas considéré les premiers

deux niveaux de l'ontologie puisqu'ils traduisent un niveau de généralité élevé ne permettant pas de répondre à la spécificité des requêtes traitant du même besoin en information, représenté par un domaine simulé de la collection.

La figure 6.5 présente le pourcentage du gain de performance de la recherche personnalisée comparativement à la recherche classique en fonction de la variation des niveaux des concepts du profil. D'après les résultats obtenus nous constatons que :

- l'utilisation du profil comme une liste de concepts de niveau 3 de l'ontologie donne la meilleure performance à P5, P10 et MAP.
- l'utilisation des concepts de différents niveaux du sous-arbre de niveau 3 de l'ontologie donne une performance plus significative que l'utilisation seule des concepts du niveau 3 et 4 du sous-arbre et presque similaire à l'utilisation du profil comme une liste de concepts de niveau 3.

Les résultats obtenus confirment l'efficacité du profil représenté par une liste de concepts de niveau 3 de l'ontologie. Ceci montre que les concepts du profil, étant du niveau 3 de l'ontologie, sont pertinents à l'ensemble des requêtes d'un même domaine. Ces résultats montrent également l'importance de l'utilisation des concepts récurrents de différents niveaux issus d'une même branche de l'ontologie (le sous-arbre maximal).

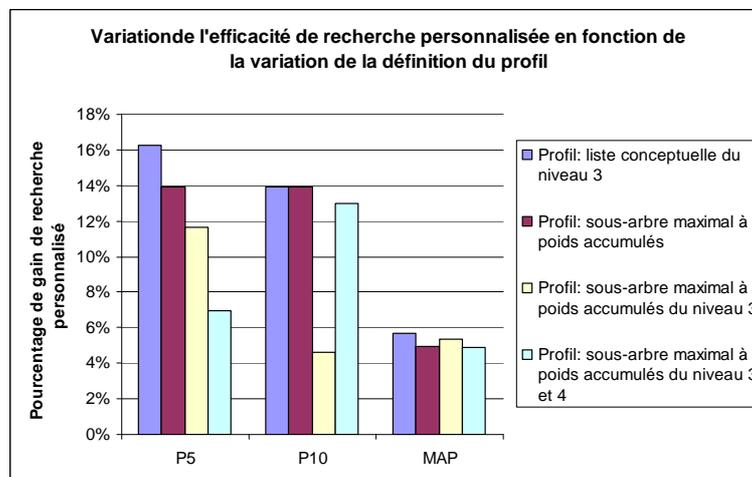


FIG. 6.5 – Variation de l'efficacité de la recherche personnalisée en fonction de différentes représentations du profil sur la collection TREC Adhoc

6.4.2.4 Évaluation de l'efficacité de la recherche personnalisée

Nous présentons dans cette section l'évaluation de l'efficacité de la recherche personnalisée comparativement à la recherche classique en utilisant les valeurs optimales identifiées dans les expérimentations précédentes. Plus particulièrement, nous utilisons 20 documents pertinents pour la construction du

profil d'une requête d'un domaine simulé, $\beta = 0.2$ pour l'évolution du profil entre des requêtes du domaine, $h = 3$ pour le calcul du score personnalisé du document et $\gamma = 0.2$ pour la combinaison du score original du document avec son score personnalisé.

La figure 6.6 montre le pourcentage de gain de la recherche personnalisée comparativement à la recherche classique à P5, P10 et MAP. Nous constatons un gain de performance important pour chacune des précisions calculées. Le meilleur gain est obtenu à P5 ($> 16\%$).

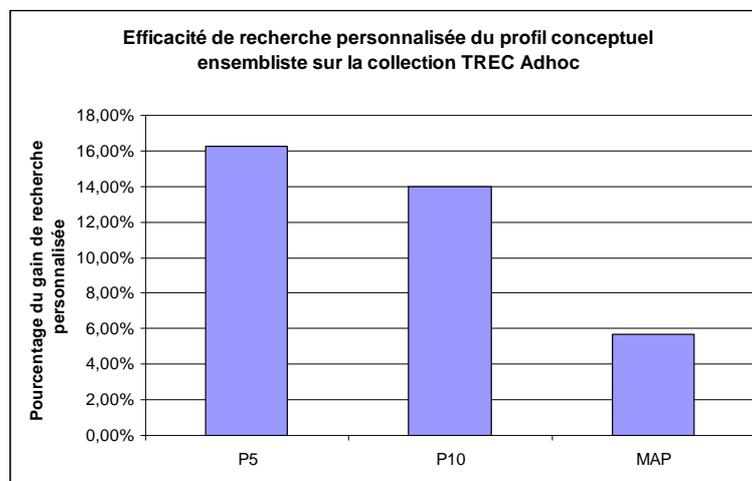


FIG. 6.6 – Gain de performance de la recherche personnalisée comparativement à la recherche classique pour le profil conceptuel ensembliste sur la collection TREC Adhoc

Dans le but d'élucider le gain de performance par domaine, nous l'avons calculé pour chaque domaine simulé de la collection à P10 et MAP. La figure 6.7 présente les résultats obtenus. Nous constatons que le gain de performance à P10 est significatif ($> 5\%$) pour quatre domaines parmi les six domaines simulés. Le domaine *Military* ne révèle aucun gain à P10 et le domaine *Law and government* révèle un gain négatif. En termes de la MAP, le gain est significatif pour l'ensemble des derniers quatre domaines sur l'axe des X.

La variabilité des taux de gains entre les domaines est dépendant de deux facteurs :

1. Le degré de corrélation des requêtes d'un même domaine : ce facteur est dépendant de la spécificité des sujets de recherche traités par les requêtes d'un même domaine dans la collection de test et de leur corrélation conceptuelle dans l'ontologie de l'ODP. L'étude effectuée sur cette collection dans [19] a montré la spécificité des requêtes appartenant à un même domaine traduisant ainsi des domaines larges dans le sens où chaque domaine regroupe divers centres d'intérêts spécifiques de l'utilisateur. Même

si la précision des profils appris sur toutes les requêtes d'un domaine est élevée, la spécificité d'une requête par rapport aux concepts du profil appris sur l'ensemble des requêtes restantes peut affecter l'efficacité de la personnalisation. A titre d'exemple, nous avons constaté que la requête *Weather related fatalities* définie dans le domaine *Environment*, n'a aucun concept en commun avec l'ensemble des concepts associés au domaine *Environment*.

2. La variabilité de la taille des domaines : ce facteur a un impact sur la qualité des profils appris en exploitant un nombre de requêtes d'apprentissage variable par domaine.
3. Le niveau fixe des concepts du profil : ce facteur a un impact sur l'efficacité du profil pour des requêtes spécifiques d'un domaine. Le profil représenté au niveau 3 de l'ontologie est efficace pour des requêtes traitant du besoin en information général et il l'est moins pour des requêtes nécessitant des niveaux de spécificité élevés.

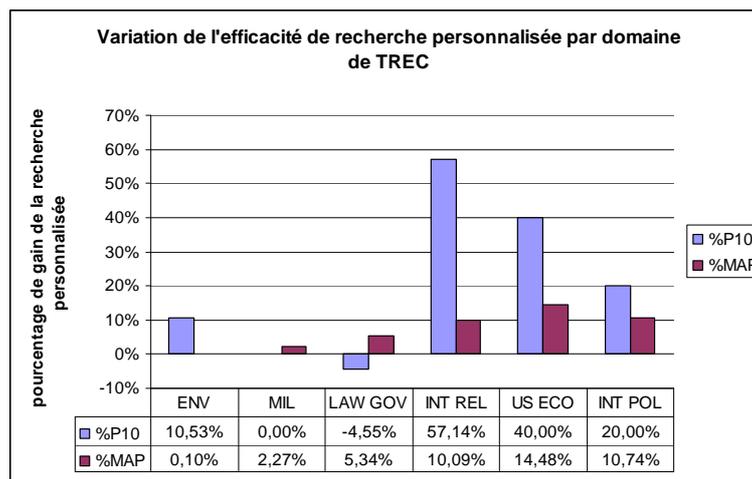


FIG. 6.7 – Gain de performance calculé par domaine de la recherche personnalisé comparativement à la recherche classique pour le profil conceptuel ensemble sur la collection TREC Adhoc

6.4.3 Évaluation basée sur le profil représenté par un graphe de concepts

Nous présentons dans cette section l'évaluation de l'efficacité de la recherche personnalisée en exploitant le profil représenté par un graphe de concepts sémantiquement liés (cf. chap.5 sect.5.5.1.3 p.145). Les objectifs de l'évaluation à ce niveau sont les suivants :

- Étudier l'impact de deux variables corrélées qui sont le nombre de documents pertinents optimal utilisé dans la construction du profil de la requête (cf. chap.5 formule 5.6.1 p.157) et le nombre de concepts utilisés

dans le calcul du score personnalisé d'un document (cf. chap.5 formule 5.20 p.173),

- Étudier l'impact du paramètre de réordonnement des résultats utilisé dans la combinaison du score original du document avec son score personnalisé (cf. chap.5 formule 5.19 p.172).
- Évaluer la performance de la recherche personnalisée comparativement à la recherche classique.

Dans ces expérimentations, nous appliquons la phase de test du protocole d'évaluation TREC adhoc sur l'ensemble des requêtes des six domaines simulés (cf. sect. 6.2.4.2).

6.4.3.1 Étude de l'impact du nombre de documents pertinents et du nombre de concepts du profil

Dans cette expérimentation, la recherche personnalisée est basée seulement sur le score personnalisé du document. L'évaluation à ce niveau est basée sur le principe suivant :

- Créer des profils associés à l'ensemble des requêtes des 6 domaines simulés de la collection en exploitant n documents pertinents dans la construction du profil des requêtes avec $n \in \{5, 10, 15, 20, 25, 30\}$.
- pour chaque ensemble de profils créés, calculer la précision de la recherche personnalisée en fonction du nombre de concepts considérés dans le calcul du score personnalisé du document dans l'ensemble $\{3, 5, 7, 10, 20\}$.

Nous présentons dans les figures 6.8 et 6.9 les résultats de la précision de la recherche personnalisée respectivement à P10 et MAP en fonction de ces deux variables. Nous avons calculé la MAP dans le but d'admettre les valeurs optimales identifiées à P10.

D'après les résultats obtenus, nous constatons que :

- La meilleure précision à P10 est obtenue en utilisant 25 documents pertinents dans la construction du profil de la requête et 7 concepts dans le calcul du score personnalisé.
- La meilleure valeur de la MAP est obtenue en utilisant 5 documents pertinents dans la construction du profil de la requête et 5 ou 7 concepts dans le calcul du score personnalisé d'un document.

D'après ces résultats, le nombre optimal de documents pertinents utilisés dans la construction du profil, est fixé à 25 et le nombre optimal de concepts considérés dans le calcul du score personnalisé du document est identifié à 7 tout en privilégiant les paramètres permettant d'obtenir une précision P10 élevée.

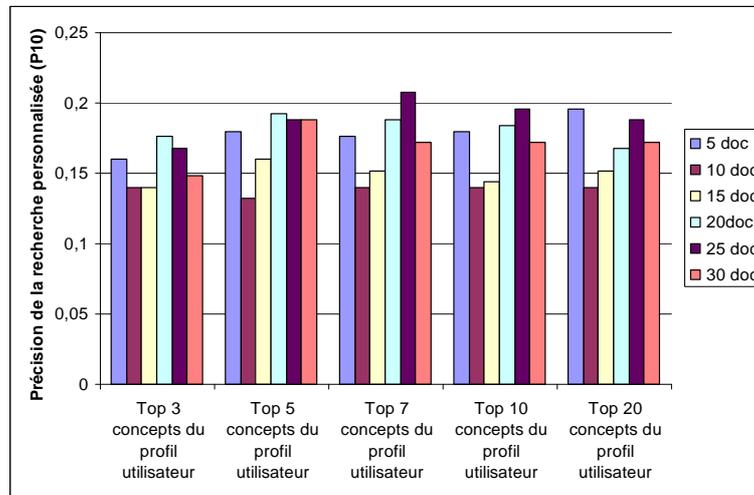


FIG. 6.8 – Variation de la recherche personnalisée à P10 en fonction des paramètres de construction du profil graphe sur la collection TREC Adhoc

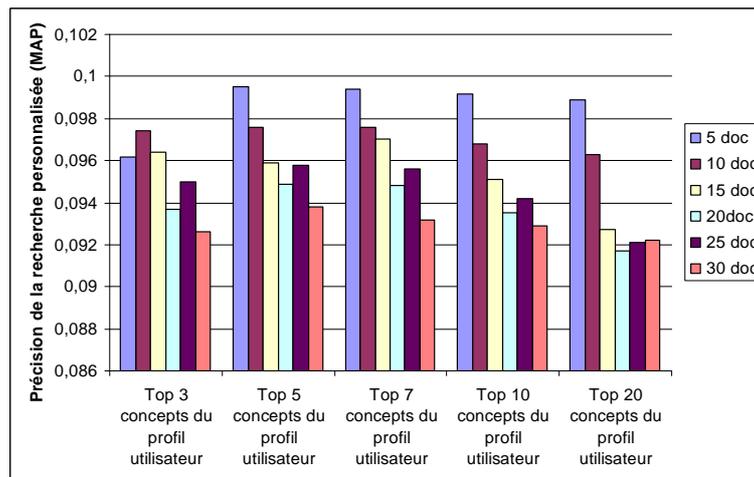


FIG. 6.9 – Variation de la précision MAP de la recherche personnalisée en fonction des paramètres de construction du profil graphe sur la collection TREC Adhoc

6.4.3.2 Étude de l'impact du paramètre de réordonnement des résultats

Dans cette étude, nous avons calculé la précision de la recherche personnalisée en variant le paramètre de réordonnement $\gamma \in [0, 1]$ sur l'efficacité de la recherche personnalisée. Nous avons utilisé les valeurs optimales identifiées précédemment pour le nombre de documents pertinents (25) et du nombre de concepts considérés dans le calcul du score personnalisé du document ($h = 7$). La figure 6.10 montre la variation de l'efficacité de la recherche personnalisée en fonction de γ .

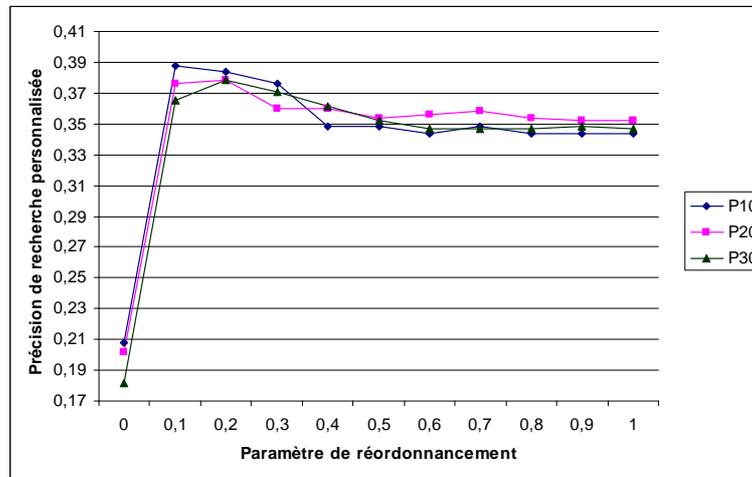


FIG. 6.10 – Variation de l'efficacité la recherche personnalisée en fonction du paramètre de réordonnement des résultats pour le profil graphe sur la collection TREC Adhoc

Nous constatons que la meilleure performance est obtenue pour une contribution faible du score original du document ($\gamma = 0.1$). Cette valeur est plus petite que la valeur optimale identifiée à $\gamma = 0.2$ pour le profil représenté par une liste de concepts. Ceci montre effectivement que le score personnalisé calculé selon le profil représenté par un graphe de concepts contribue grandement à l'augmentation du score de pertinence du document que le score original.

6.4.3.3 Évaluation de l'efficacité de la recherche personnalisée

Dans cette expérimentation nous comparons la précision obtenue par la recherche personnalisée comparativement à la recherche classique en exploitant les meilleures valeurs identifiées dans les expérimentations précédentes.

1. le nombre de documents pertinents utilisés dans la construction du profil des requêtes est égal à 25.
2. le nombre de concepts considérés dans le calcul du score personnalisé $h = 7$,
3. Le modèle de RI personnalisée est basé sur le réordonnement des résultats de recherche de la requête avec $\gamma = 0.1$.

Les résultats obtenus sont présentés dans la figure 6.11 en termes de la précision et du rappel calculés à différents points (5, 10, ..., 100 premiers documents restitués). Nous constatons que la précision de la recherche personnalisée dépasse celle de la recherche classique, plus particulièrement sur les premiers documents retrouvés.

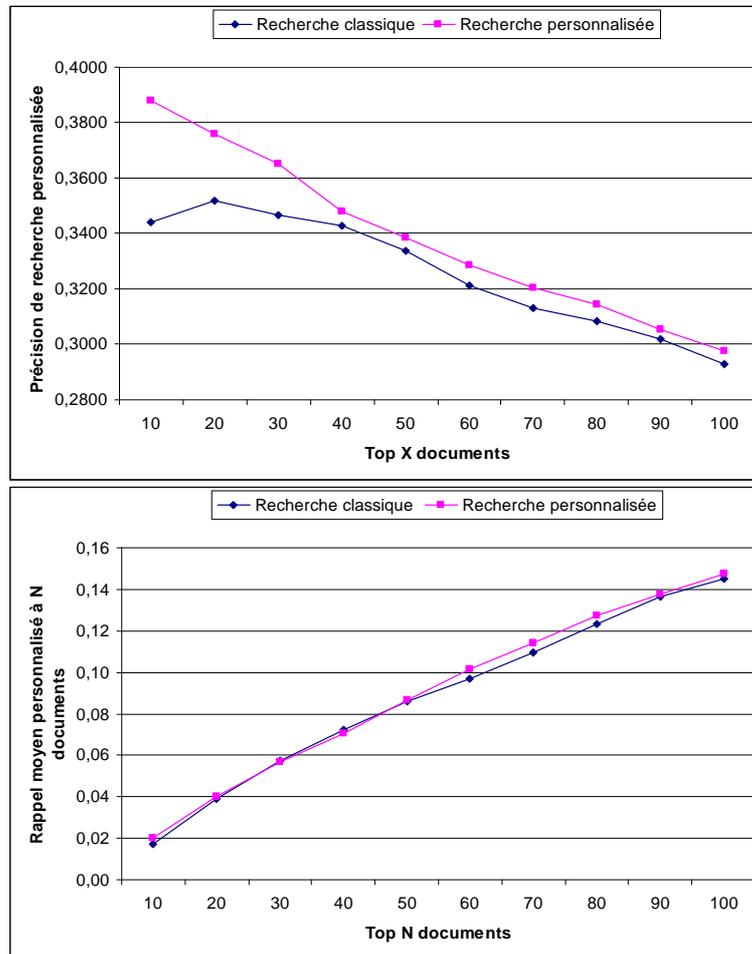


FIG. 6.11 – Évaluation de la performance du modèle de RI personnalisée en termes de Top-n précision moyenne et Top-n rappel moyen en utilisant le profil graphe sur TREC adhoc

Dans le but d'évaluer l'efficacité de notre approche de RI personnalisé dans un cadre d'évaluation plus réaliste, nous avons utilisé 5 documents pertinents dans la construction du profil par requête et les mêmes valeurs optimales des autres paramètres de personnalisation ($h = 7, \gamma = 0.1$). La figure 6.12 montre le gain de performance de la recherche personnalisée comparativement à la recherche classique en termes de la précision moyenne et du rappel moyen en utilisant respectivement 5 documents pertinents ou 25 documents pertinents. Nous pouvons constater que le gain de performance en utilisant 25 documents pertinents est toujours positif et important à P10 (12.79%), P20 (6.8%) et P30 (5.3%) avec un taux d'amélioration $> 5\%$. Il est plus élevé que celui obtenu en utilisant 5 documents pertinents sur la majorité des points de précisions calculés. Cependant, la précision à P10 en utilisant 5 documents pertinents dépasse la recherche classique à 7% et elle est variable pour le reste des précisions calculées. Les pourcentages d'amélioration de la MAP sont estimés à 2.6% et 6% respectivement en utilisant 25 documents pertinents et 5 documents pertinents.

Ceci montre effectivement qu'en utilisant un nombre faible de documents pertinents, l'amélioration de la précision P10 et de la MAP reste significatif.

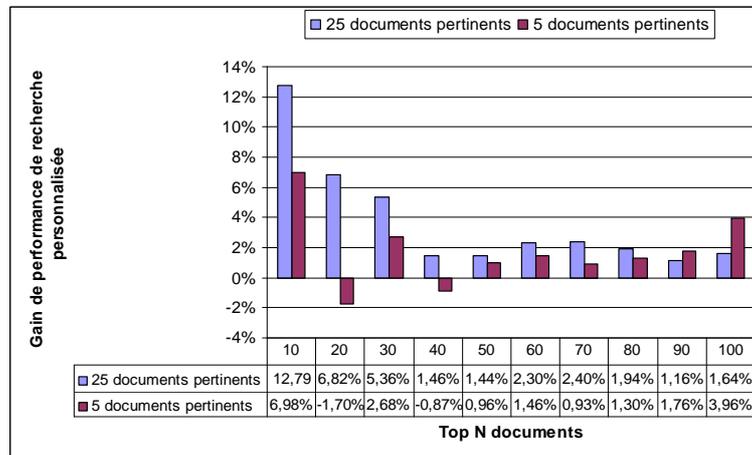


FIG. 6.12 – Evaluation du gain de performance de la recherche personnalisée en termes de Top-n précision moyenne et Top-n rappel moyen en utilisant en utilisant 25 documents pertinents sur TREC Adhoc

Dans le but de mieux comprendre la variation du gain sur l'ensemble des requêtes de test, nous l'avons calculé à P10 documents pour chaque domaine simulé tout en utilisant séparément 5 ou 25 documents pertinents. Le figure 6.13 présente les résultats obtenus.

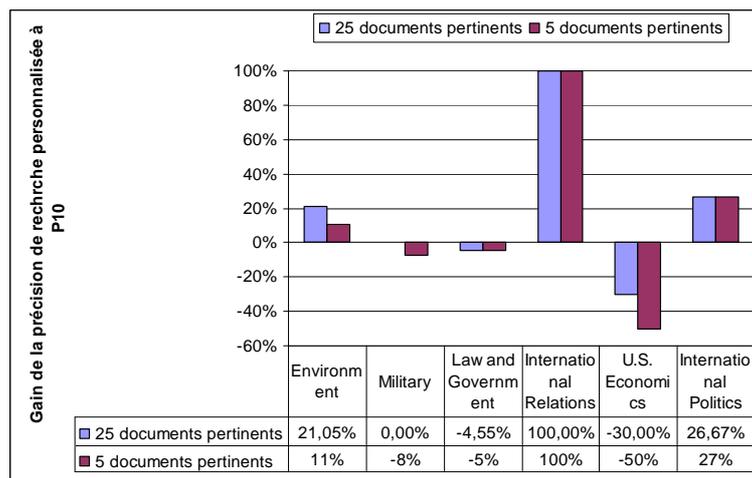


FIG. 6.13 – Evaluation de gain de performance de la recherche personnalisée en termes de la précision P10 et la MAP par domaine de TREC Adhoc

Nous remarquons que les gains sont variables cependant en fonction du domaine d'intérêt simulé. En utilisant 25 documents pertinents, les gains de précision sont significatifs à P10 pour l'ensemble des domaines International

Relations (100%), Environment (21.5%) et International Politics (26.67%). Des gains négatifs sont obtenus pour deux domaines simulés, notamment pour *law and Government*, et pour *US economics*. En utilisant 5 documents pertinents, les gains de précision sont également intéressants à P10 pour l'ensemble des domaines International Relations (100%), Environment (11%) et International Politics (27%). Ces gains sont négativement plus grands en utilisant 5 documents pertinents. Notons que le domaine *Military* ne révèle aucun gain en utilisant 25 documents pertinents et a un gain négatif en utilisant 5 documents pertinents. Cette comparaison montre bien que la qualité des profils est d'autant plus précise que le nombre de documents pertinents exploités dans la construction des profils est grand. Ceci est due effectivement à la difficulté des requêtes de la collection et la nécessité d'une quantité de données d'apprentissage assez importante pour la construction des profils efficaces.

Les facteurs impliqués ayant un impact sur la variation du gain de performance entre les domaines sont déjà cités dans l'analyse de résultats pour le profil représenté par une liste de concepts (cf. sect. 6.4.2.4) à l'exception de l'impact du niveau fixe des concepts du profil. En effet le profil utilisé dans cette expérimentation est représenté par un graphe de concepts disposant des niveaux de spécificités des concepts associés et des liens de référence liant ses différents concepts. Comparativement à l'efficacité de la recherche personnalisée obtenue en utilisant le profil représenté par une liste de concepts, le gain de performance obtenu selon le profil représenté par un graphe de concepts est plus important pour certains domaines et moins important pour certains autres. La précision P10 est plus élevée pour trois domaines, *Environment*, *International Relations*, et *international politics*. Elle est moins importante pour le domaine *US Economics* et stable pour les domaines *Military* et *Law and Government*. Bien que la qualité des profils représentés selon un graphe de concepts contribue plus efficacement à l'amélioration de la précision de recherche, nous pouvons constater les points suivants :

- L'efficacité de la recherche personnalisée selon le profil représenté par un graphe de concepts est d'autant plus élevée que les requêtes d'un même domaine sont corrélées.
- Le degré de spécificité de la requête par rapport au domaine a un impact majeur sur l'efficacité de la recherche personnalisée.

6.5 Évaluation de l'efficacité de la recherche la personnalisée à long terme

Nous nous intéressons dans cette section à l'évaluation de la précision d'identification d'un profil à long terme adéquat à une requête, représentés chacun par un graphe de concepts ainsi que l'efficacité des profils identifiés dans la

personnalisation du processus de RI. Le modèle de recherche adopté dans cette évaluation est analogue à celui utilisé dans l'évaluation de l'efficacité de la RI personnalisée à court terme.

6.5.1 Évaluation de la précision d'identification des profils à long terme

L'évaluation de la précision d'identification des profils à long terme revient à évaluer la précision de la mesure de distance sémantique basée-graphe (cf. 5 sect. 5.7.3 p. 168). Le principe de l'évaluation dans ce but est basée sur le calcul de la distance sémantique entre le graphe de la requête et chacun des profils des six domaines simulés de la collection. La précision de la mesure est ensuite calculée par comparaison des profils des domaines identifiés automatiquement comme étant similaires aux requêtes testées, aux annotations prédéfinies des requêtes par les domaines de la collection.

Plus précisément, la démarche de l'évaluation est décrite comme suit :

- pour chaque domaine d'intérêt prédéfini Dom_j , créer le profil G_{Dom_j} à partir des toutes ses requêtes associées. Nous avons choisi de créer le profil à long terme en utilisant les requêtes numérotées de 51 à 150 de la collection dans le but d'utiliser assez de données d'apprentissage simulant ainsi un profil à long terme construit sur tout l'historique de recherche de l'utilisateur.
- pour chaque requête q_k d'un domaine Dom_i :
 1. créer le profil $G_{q_k Dom_i}$ à partir des requêtes restantes du domaine en suivant la stratégie de la validation croisée.
 2. calculer la distance sémantique entre le graphe de la requête G_{q_k} et le profil de son domaine $G_{q_k Dom_i}$, et sa distance avec chacun des profils des domaines G_{Dom_j} .
 3. identifier le profil le plus similaire à la requête en minimisant les distances sémantiques calculées comme suit :

$$G^* = \operatorname{argmin}_{j \in \{1, \dots, 5\} \wedge i \neq j} \left\{ d(G_{q_k}, G_{q_k Dom_i}), d(G_{q_k}, G_{q_k Dom_j}) \right\}$$

La précision de la mesure est ensuite calculée par la fraction des requêtes correctement affectées à leur domaine sur le nombre total des requêtes de tous les domaines simulés. Cette précision peut être calculée par domaine en divisant le nombre de requêtes correctement affectées au domaine en question sur le nombre total des requêtes du domaine.

Suivant cette démarche d'évaluation, nous avons obtenu une précision globale de 36.5% sur l'ensemble des 52 requêtes appartenant aux 6 domaines simulés. Cette précision faible est due à deux facteurs :

- L'impact du degré de corrélation conceptuelle du sujet de la requête au sujet traité par l'ensemble des requêtes restantes du domaine en question.
- Le nombre de requêtes servant à la construction des profils à long terme est variable entre les domaines.
- L'annotation uni-domaine de la requête dans la collection ne permet pas de prendre en compte sa corrélation conceptuelle aux sujets traités par les autres domaines simulés. En effet, nous avons constaté que certaines requêtes sont en corrélation thématique avec plusieurs domaines simulés, par conséquent, leur classification dans le domaine d'origine et dans l'un des domaines corrélés doit être correct. Cette constatation a été tirée à partir du calcul du taux de corrélation conceptuelle calculé entre les profils des six domaines simulés (cf. Fig. 6.14). Pour chaque domaine, nous avons calculé le pourcentage des concepts communs avec chacun des autres domaines simulés et avons constaté ainsi que les requêtes des domaines sont corrélés à des taux différents.

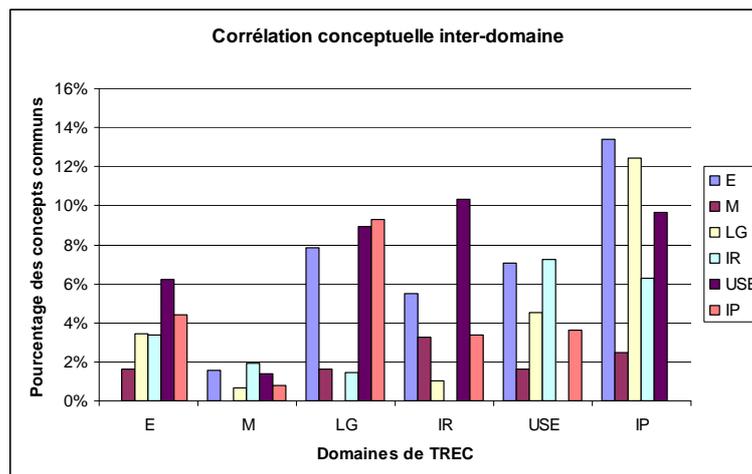


FIG. 6.14 – Taux de la corrélation conceptuelle entre profils de domaines

Dans le but de pallier le problème de l'annotation uni-domaine des requêtes, nous avons ré-annoté manuellement les requêtes des six domaines simulés de la collection en attribuant à chaque requête en plus du domaine d'origine, un ou plusieurs domaines reliés. Le critère de cette annotation multi-domaines est basé sur la corrélation de la thématique traitée par la requête à un ou plusieurs domaines simulés. Selon cette annotation, nous avons recalculé la précision d'identification des profils à long terme et avons obtenu une précision de 61.5%. La figure 6.15 présente la précision de classification des requêtes calculées par domaine, selon l'annotation d'origine des requêtes (uni-domaine) et l'annotation multi-domaines des requêtes. Nous constatons que la précision est plus élevée pour l'annotation multi-domaines que pour l'annotation uni-domaine des requêtes. Toutefois, la précision globale de classification est moyennement limitée sur quelques domaines (*International relations, US. Economics, International Politics*). Ceci montre effectivement la difficulté des requêtes des do-

maines même avec l'annotation multi-domaines. En effet, celle-ci est effectuée au niveau thématique des requêtes indépendamment de la corrélation conceptuelle de son domaine d'origine avec les autres domaines simulés dans l'ontologie.

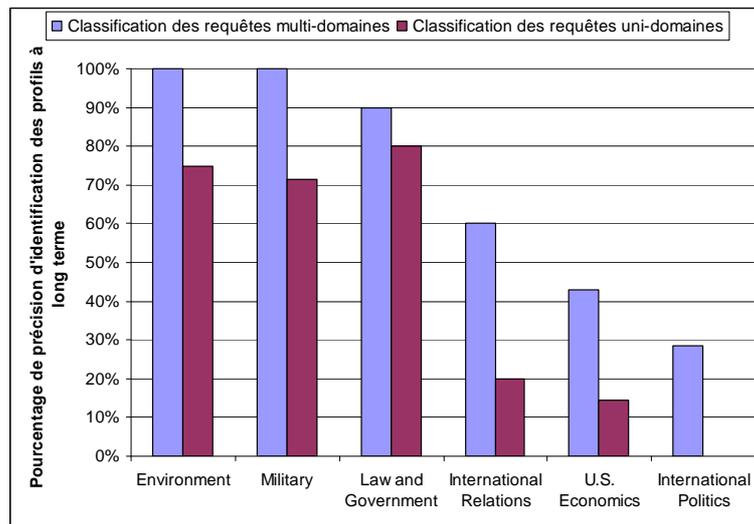


FIG. 6.15 – Évaluation de la précision d'identification des profils à long terme adéquats aux requêtes issus des domaine de TREC Adhoc

6.5.2 Évaluation de l'efficacité de la recherche personnalisée à long terme

Même si la précision de la classification des requêtes aux domaines associées n'atteint pas une performance élevée, nous nous intéressons à évaluer l'efficacité la personnalisation du processus de RI des requêtes en exploitant pour chaque requête, le profil identifié comme étant le plus similaire.

Le processus de personnalisation appliqué à ce stade consiste à réordonner les résultats de recherche de la requête selon la formule 5.19 (chap. 5,172) tout en exploitant les concepts les plus similaires à la requête. En effet, il s'agit d'instancier les concepts du profil à long terme en fonction de leur similarité à la requête en cours d'évaluation. Dans ce but, nous recalculons le score de chaque concept du profil selon la mesure de similarité cosinus avec la requête comme suit :

$$Score(c_i) = \cos(\vec{c}_i, \vec{q}_t)$$

Le score personnalisé d'un document est ensuite calculé en exploitant les dix premiers concepts les plus similaires à la requête. Ce nombre est identifié précédemment comme étant la valeur optimale permettant d'atteindre la meilleure personnalisation. La figure 6.16 présente la précision de la RI personnalisée

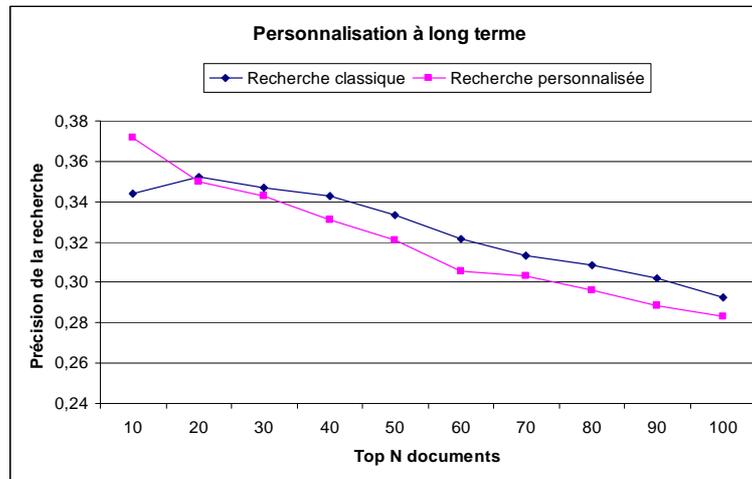


FIG. 6.16 – Évaluation de l'efficacité de la RI personnalisée à long terme sur TREC Adhoc

comparativement à la RI classique. Nous constatons que la précision de la RI personnalisée à P10 est plus élevée que celle de la RI classique et est moins importante à partir de P30. L'amélioration seule de la précision à P10 montre bien l'efficacité des profils identifiés est sur les premiers 10 documents présentés à l'utilisateur. Toutefois, la différence des degrés de corrélation des requêtes aux profils identifiés à long terme, peut réduire la précision de recherche. En effet, le calcul du score personnalisé d'un document pertinent en fonction de sa similarité avec tous les dix premiers concepts du profil est susceptible de réduire son score de pertinence dans l'ensemble de documents retournés.

La figure 6.17 présente la précision de la RI personnalisée comparativement à la RI classique pour chaque domaine simulé. Nous remarquons que le gain de précision de la recherche est variable entre les domaines. Le gain négatif obtenu pour le domaine *US. Economics* impacte le gain de la MAP sur l'ensemble des requêtes de test, estimée à -5% . Comparativement à la personnalisation à court terme, les facteurs additionnels qui sont à l'origine de la différence de ces gains dépendent de l'augmentation du nombre de requêtes de chaque domaine servant à la création de son profil d'une part et à la différence de corrélation des requêtes aux profils identifiés à long terme. Plus le nombre de requêtes utilisées dans la construction du profil est grand, plus le profil porte des concepts généraux représentant l'ensemble des toutes les requêtes du domaine et qui seront ainsi moyennement efficaces dans l'amélioration de la précision des requêtes.

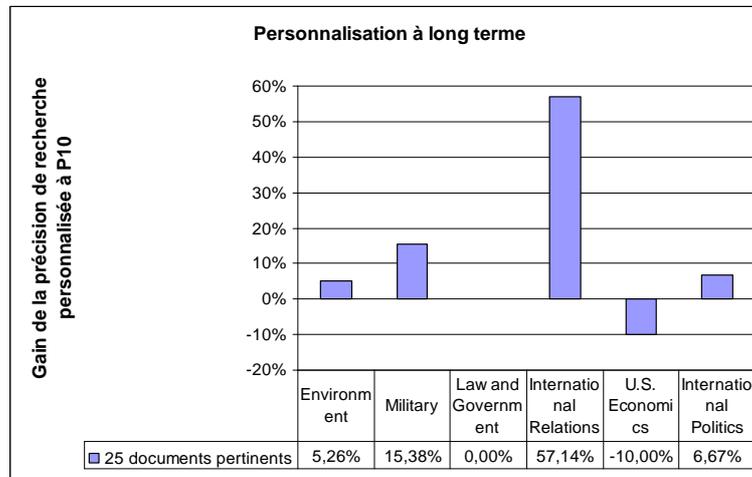


FIG. 6.17 – Évaluation de l’efficacité de la RI personnalisée à long terme par domaine sur TREC Adhoc

6.6 Conclusion

Nous avons présenté au cours de ce chapitre l’évaluation de notre approche de personnalisation selon les deux modèles de représentation du profil (liste de concepts et graphe de concepts) sur le cadre d’évaluation TREC Adhoc. Ce cadre d’évaluation a permis d’évaluer l’efficacité de la recherche personnalisée en simulant à la fois un centre d’intérêt générique de l’utilisateur selon la méthode de la cross-validation et indépendamment du mécanisme de la délimitation des sessions de recherche. Nous avons ensuite évalué l’efficacité de la recherche personnalisée en instanciant le profil le plus adéquat à une requête donnée parmi un ensemble de profils représentant divers centres d’intérêts.

Les résultats expérimentaux montrent bien l’efficacité des différentes composantes de notre approche, notamment :

- une précision significative de la qualité des profils construits sur la base des domaines prédéfinis dans la collection TREC Adhoc.
- une précision moyenne d’identification des profils à long terme selon une mesure de distance sémantique basée-graphe liée à la difficulté des requêtes de la collection.
- un gain de performance de la recherche personnalisée variable entre les domaines. cette variabilité dépend principalement du degré de spécificité des requêtes par rapport au domaine et du degré de corrélation des requêtes d’un même domaine.

L’importance du cadre d’évaluation TREC Adhoc consiste à estimer la performance de la recherche personnalisée dans des sessions de recherche prédéfinis. Toutefois, les résultats de l’évaluation de l’efficacité de la recherche person-

nalisee selon le cadre d'evaluation issu de *TREC Adhoc* sont relativement dependantes de la collection de requetes annotees. Le degre de correlation entre les requetes d'un meme domaine est defini au niveau general dans le sens ou les requetes appartiennent a un meme domaine d'interet general. Dans le but de d'evaluer l'apport de l'approche de personnalisation en utilisant une sequence de requetes, etant des reformulations au niveau specifique d'un meme besoin en information, nous definissons dans le chapitre suivant un nouveau cadre d'evaluation permettant de simuler une session via la generation differentes formulations possibles d'une meme requete et evaluer l'impact du mecanisme de delimitation des sessions de recherche sur la performance de la recherche personnalisee. Plus precisement, notre objectif consiste a introduire un scenario d'evaluation qui integre le basculement des centres d'interets entre les sessions de recherche et en absence des annotations predefiniees des requetes a des domaines d'interets.

Chapitre 7

Evaluation basée sur un cadre d'évaluation issu de TREC HARD 2003 : propositions, mise en œuvre et résultats

7.1 Introduction

Notre objectif dans ce chapitre est de mesurer la performance de notre approche de personnalisation en présence du basculement des centres d'intérêts entre les sessions de recherche et où le profil utilisateur est construit sur la base d'un ensemble de requêtes non préalablement annotées de domaines d'intérêts. Dans ce but, nous proposons un nouveau cadre d'évaluation [11], basé sur la réutilisation des ressources de la tâche HARD de TREC 2003. Ce cadre est basé sur les points principaux suivants :

1. L'augmentation de la collection de test par des profils utilisateurs simulés par les sujets des requêtes principales de la collection,
2. La simulation d'une session de recherche par un ensemble de sous-requêtes générées automatiquement pour une même requête, considérées ainsi des reformulations différentes du même sujet de recherche,
3. une stratégie d'évaluation permettant d'évaluer l'efficacité de la recherche personnalisée sur une séquence de sessions simulées et où le profil utilisateur est construit sur un ensemble des requêtes groupées dans une même session selon un seuil prédéterminé.

Nos évaluations selon ce cadre portent principalement sur les objectifs suivants :

1. L'évaluation de la qualité des profils construits selon un graphe de concepts,

2. L'évaluation de l'efficacité de la recherche personnalisée sur des séquences de sessions de recherche simulées. Nous nous intéressons en particulier à l'évaluation de la délimitation des sessions de recherche dans une première phase et la performance de la recherche personnalisée comparativement à la recherche classique dans une deuxième phase.
3. L'évaluation comparative de notre approche à une approche de personnalisation proposée dans le domaine [67].

Ce chapitre est organisé comme suit. La section 7.2 présente le cadre d'évaluation issu de TREC HARD ainsi que les composantes principales de ce cadre. La section 7.3 présente l'évaluation de la qualité des profils représentés selon des graphes de concepts. La section 7.4 présente l'évaluation de l'efficacité de la recherche personnalisée en utilisant un seuil prédéfini de délimitation des sessions de recherche. La section 7.5 présente une évaluation comparative de notre approche avec une approche dans le domaine en utilisant les requêtes difficiles de la collection TREC HARD 2003. La dernière section conclut le chapitre.

7.2 Le cadre d'évaluation issu de TREC HARD 2003 : un cadre dédié à une recherche personnalisée orientée-session

Notre objectif est de mettre en place un scénario d'évaluation permettant d'évaluer l'efficacité de recherche personnalisée dans un contexte de basculement des centres d'intérêts entre les sessions de recherche et en absence des corrélations prédéfinies entre les requêtes.

Les motivations principales conduisant à la proposition du cadre d'évaluation issu de TREC HARD sont résumées comme suit :

- Les requêtes utilisées dans le cadre d'évaluation issu de TREC Adhoc ne sont pas ambiguës ni difficiles et sont annotées clairement par des domaines d'intérêts.
- L'évaluation de l'efficacité de la recherche personnalisée est effectuée indépendamment du mécanisme de délimitation des sessions de recherche. Elle est effectuée selon le protocole de la validation croisée qui permet d'estimer les performances de recherche personnalisée pour des situations de recherche traitant d'un unique sujet d'intérêt.

Dans le but d'évaluer l'efficacité de recherche personnalisée compte tenu du basculement dans le sujet des requêtes (scénario plus proche de la réalité), nous proposons un nouveau cadre d'évaluation [11] basé sur l'utilisation des requêtes difficiles non clairement annotées par un domaine d'intérêt, l'augmen-

tation de la collection de test par des profils utilisateurs simulés, et l'évaluation de l'efficacité de la recherche personnalisée tout en appliquant la délimitation des sessions de recherche. Ce cadre a été spécifié et mis en œuvre sur la base de la collection HARD de TREC (comportant des requêtes particulièrement difficiles), cependant le protocole d'évaluation proposé est généralisable pour évaluer toute approche de RI personnalisée basée-session et en utilisant toute autre collection de test en présence des composantes suivantes :

1. une collection de test comprenant des requêtes, des documents et des jugements de pertinence.
2. Un algorithme de simulation des contextes de recherche de l'utilisateur et de construction des profils utilisateur, où un contexte de recherche est défini par une séquence hypothétique de sessions de recherche simulées selon le principe suivant :
 - Générer des sous-requêtes comme étant différentes reformulations possibles d'une même requête, dite principale, définissant ainsi une session de recherche simulée. Le profil utilisateur construit dans une session désigne le sujet de recherche traité par la requête principale.
 - Aligner les sous-requêtes liées à une même requête aléatoirement sur la séquence,
 - Construire le profil utilisateur dans une session de recherche simulée en exploitant un sous-ensemble de documents pertinents associés à la requête principale.
3. Un protocole d'évaluation permettant d'évaluer l'efficacité de recherche personnalisée à travers une séquence de sessions de recherche simulées traitant de sujets différents. Ce protocole consiste en une phase de test qui permet d'utiliser un seuil prédéfini de délimitation des sessions pour grouper les sous-requêtes corrélées dans une même session lors de l'évaluation de l'efficacité de la recherche personnalisée.

7.2.1 Collection de test

7.2.1.1 Collection de requêtes

Nous avons utilisé les requêtes de la collection *HARD TREC 2003*, qu'on nomme *requêtes principales*. Le choix de cette collection a pour but d'évaluer l'efficacité de la personnalisation sur des requêtes difficiles. Le format général des requêtes (*topics*) de cette collection est le suivant :

```
<num> HARD-033 </num>
<title> Animal Protection </title>
<descr>
```

What have countries or groups of people done to protect animals in recent years?

</descr>

<narr>

All new laws passed in 1999 to protect animals, and all efforts made by the government or by people to protect animals are on topic. All personal reasons behind the protection of animals (like elections), that contributed to set new animal's protection's laws are off-topic. But reasons behind a law that a government put to protect animals are on topic.

</narr>

<searchterms>

people, animal protection, new laws, legislation, SPCA

</searchterms>

<hard> item=purpose, value=Answer Answer: How did humans help protect animals in 1999?</hard>

<hard> item=genre, value=Any</hard>

<hard> item=granularity, value=Sentence</hard>

<hard> item=familiarity, value=Unknown</hard>

Bien que le format général d'une requête de la collection dispose des métadonnées liés au contexte de recherche de l'utilisateur, nous ne les avons pas exploitées puisque le contexte de recherche dans notre modèle désigne un profil de l'utilisateur évolutif qui décrit ses divers centres d'intérêts personnels.

7.2.1.2 Collection de sous-requêtes

Comme nous ne disposons d'aucune information permettant l'élaboration de sessions de recherche, nous tentons de la construire de la manière suivante. Nous définissons la notion des sous-requêtes d'une requête principale et nous considérons que les sous requêtes sont vues comme des reformulations différentes du même besoin en information représenté par le sujet de la requête principale simulant ainsi une session de recherche.

Le principe de génération des sous-requêtes d'une même requête est basée sur la sélection et la subdivision d'un ensemble de documents pertinents de la requête principale, noté *source pertinence*, en n sous-ensembles de documents de taille égale. Chaque sous-ensemble sert de base à la définition d'une sous-requête et représente l'ensemble de documents pertinents utilisés dans la construction du profil de la sous-requête (cf. chap. 5, sect. 5.6.1). Nous générons deux ensembles de sous-requêtes en fonction de la définition de la *source pertinence* de la requête principale.

- Source *pertinence-qrels* : les documents pertinents sont les premiers N documents listés dans le fichier des jugements de pertinence et associés à la requête,
- Source *pertinence-qrels-Top50* : les documents pertinents sont les premiers N documents qui appartiennent aux cinquante (50) premiers documents retrouvés par le système en réponse à la requête principale.

L'objectif des deux méthodes de définition de la *source pertinence* pour chaque requête est d'étudier l'impact de l'utilisation de documents pertinents effectivement retrouvés par le système pour une requête donnée versus l'utilisation des documents pertinents non nécessairement retrouvés par le système sur la qualité du profil d'une part et l'efficacité de recherche personnalisée d'autre part. Nous abordons dans cette section le principe général de la génération des sous-requêtes à partir de la source *pertinence-qrels* et celle générée à partir de la source *pertinence-qrels-Top50*.

7.2.1.2.1 Principe de génération des sous-requêtes et de simulation des sessions de recherche La génération d'un ensemble de sous-requêtes d'une même requête principale est à la base de la définition des sessions de recherche simulées. Une session de recherche est simulée par alignement séquentiel des sous-requêtes associées à une même requête principale.

Le principe de génération des sous-requêtes d'une requête principale est illustré dans la figure 7.1, il est décrit comme suit :

1. Extraire la *source pertinence* pour une requête principale q en construisant l'ensemble des N vecteurs documents pertinents associés, soit dp_q .
2. Subdiviser la *source pertinence* en p sous-sources, notés sp_i , $sp_i \subset dp_q$,
3. Pour chaque sous-source *pertinence* sp_i , créer un vecteur centroïde selon la formule : $c_i(t) = \frac{1}{|sp_i|} \sum_{d \in sp_i} w_{td}$, w_{td} est le poids du terme t dans le document d calculé selon la fonction de pondération classique : $w_{td} = f_{td} * \log(N/N_t)$, où f_{td} est la fréquence du terme t dans le document d , N_t est le nombre de documents de la collection contenant le terme t et N est le nombre total de documents de la collection
4. Extraire de chaque centroïde la sous-requête représentée par les k termes les mieux pondérés,

Dans le but de valider le processus de génération des sous-requêtes, nous devons montrer que les sous-requêtes générées traitent le même sujet que la requête principale d'une part, et que ces sous-requêtes ne sont pas exactement similaires dans le sens où elles représentent différentes formulations possibles d'une même requête. Dans ce but, nous proposons deux mesures permettant de traduire ces deux critères :

- **Taux de recouvrement de pertinence** : ce taux est calculé pour

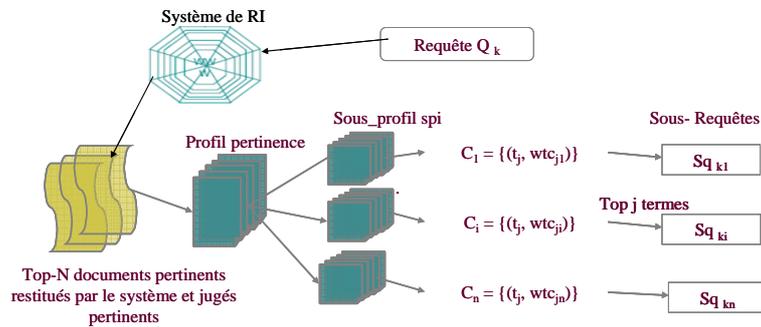


FIG. 7.1 – Principe de génération des sous-requêtes d’une même requête de la collection TREC HARD 2003 selon la *source pertinence-qrels-Top50*

chaque sous-requête relativement à la requête principale. Il est calculé par estimation du pourcentage de documents pertinents communs retournés par ces deux types de requêtes. Cette estimation implique que la sous-requête traite le même sujet que la requête principale vu qu’elle retourne des documents pertinents communs avec celle-ci.

- **Taux de non recouvrement de documents** : ce taux est calculé entre les sous-requêtes par estimation du pourcentage de documents différents sur les 20 ou 50 premiers documents retournés par le système pour chacune des sous-requêtes. Cette estimation implique que les sous-requêtes diffèrent de certains termes puisqu’elles ne renvoient pas exactement la même liste de résultats.

Dans nos expérimentations, Nous avons généré trois sous-requêtes pour chaque requête principale dans le but de définir une session de recherche de longueur 3. Le choix de ce nombre a un double objectif :

1. Simuler des sessions de recherches de longueur moyenne proche d’un contexte de recherche réel. Selon différentes études [89, 77], analysant des fichiers log des moteurs de recherche, plus que 79% des sessions sont en effet composées en moyenne d’au moins trois requêtes.
2. évaluer notre approche de personnalisation en utilisant un profil à court terme construit sur la base d’un historique de recherche court composé d’un nombre faible de requêtes.

7.2.1.2.2 Collection de sous-requêtes générées de la source *pertinence-qrels*

Le nombre total des requêtes principales selon la source *pertinence-qrels* est de 30 requêtes. Ce nombre est obtenu en éliminant :

- les requêtes qui produisent une valeur de précision moyenne (MAP) nulle vu que notre modèle consiste à réordonner les résultats de recherche et non pas à personnaliser le modèle de sélection de l’information.
- les requêtes qui ont un nombre de documents pertinents < 30 , et ce dans le but de créer des sous-sources de pertinence de taille égale.

Pour chaque requête, nous avons généré 3 sous-requêtes par requête en considérant une source *pertinence-qrels* contenant 30 documents pertinents. Nous obtenons ainsi un ensemble de 90 sous-requêtes en total. La taille de la *source pertinence* est fixée selon deux critères :

1. l'utilisation d'un nombre moyen de documents pour la construction du profil associé à chaque sous-requête (soit 10 documents par sous-requête).
2. ce nombre est dépendant du nombre moyen de documents pertinents par requête, considéré faible dans la collection de test. Ces requêtes sont effectivement difficiles et caractérisées par une difficulté car leur valeur de précision faible.

Nous présentons dans le tableau 7.1 un exemple de sous-requêtes générées à partir de la requête principale numérotée 77 dans la collection de requêtes de TREC HARD 2003 en utilisant la source *pertinence-qrels*. Les documents traitant des maladies hors US sont considérés non pertinents.

Topic HARD-77	Insect-borne illnesses
sous-requête 1	encephalitis, lyme, state
sous-requête 2	encephalitis, mosquito, spray
sous-requête 3	state, encephalitis, nile
Termes donnés par TREC	insects, Lyme Disease, ticks, West Nile virus, mosquitos

TAB. 7.1 – Un exemple de sous-requêtes générées de la source *pertinence-qrels*

La figure 7.2 présente les taux de recouvrement de pertinence de chaque sous-requête relativement à la requête principale. La figure 7.3 présente le taux de non-recouvrement moyen entre les sous-requêtes en considérant la source *pertinence-qrels*. D'après les résultats obtenus nous pouvons tirer deux conclusions :

1. La figure 7.2 montre bien que les sous-requêtes permettent de retourner autant, sinon plus de documents pertinents que la requête principale, ce qui traduit bien que les sous-requêtes traitent du sujet de la requête principale.
2. La figure 7.3 montre bien, qu'avec un taux de non-recouvrement de plus de 40%, les sous-requêtes ne contiennent pas les mêmes termes tout en traitant le même sujet, ce qui va dans le sens de la complétude du sujet traité par la requête principale.

7.2.1.2.3 Collection de sous-requêtes générées de la source *pertinence-qrels-Top50* Nous avons généré la collection des sous-requêtes à partir de 19 requêtes principales de la collection en considérant la taille de la source *pertinence-qrels-Top50* au plus égale à 9 documents pertinents par requête. Les requêtes sont obtenues par élimination des requêtes principales selon les mêmes critères cités dans la génération de la collection des sous-requêtes précédente,

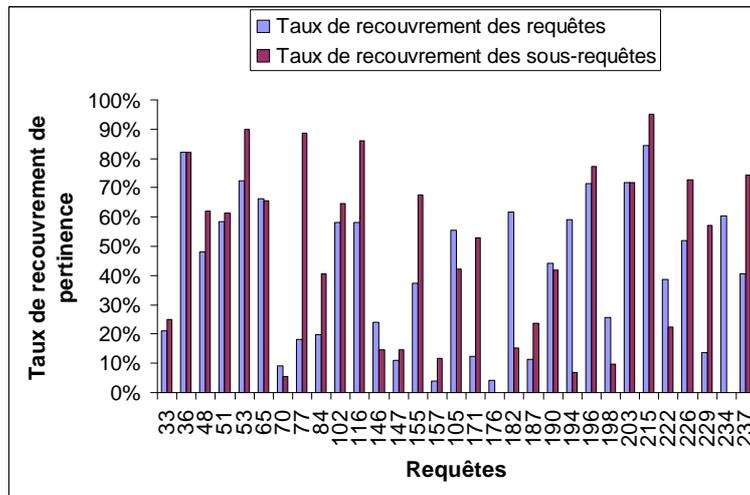


FIG. 7.2 – Taux de recouvrement de pertinence des requêtes et des sous-requêtes en termes du nombre de documents pertinents

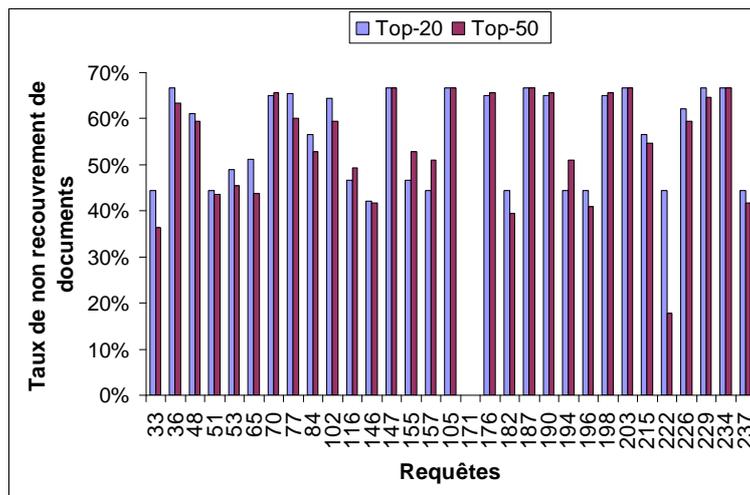


FIG. 7.3 – Taux de non-recouvrement moyen des sous-requêtes en termes du nombre de documents différents sur les Top-20 et Top-50 documents

et par élimination des requêtes qui possèdent moins que 3 documents pertinents sur les Top 50, et ce dans le but de créer des profils des sous-requêtes en exploitant le même nombre de documents pertinents. Nous avons obtenu ainsi un total de 57 sous-requêtes.

Nous présentons dans la figure 7.4 le taux de recouvrement de pertinence et le taux de non-recouvrement des documents pour le sous-requêtes générées à partir de la source *pertinence-qrrels-Top50*. Nous constatons également que les sous-requêtes traitent le même sujet que la requête principale et que la majorité des sous-requêtes ne renvoient pas exactement la même liste de résultats sur les premiers 50 documents retournés par le système. Dans le but d'élucider la

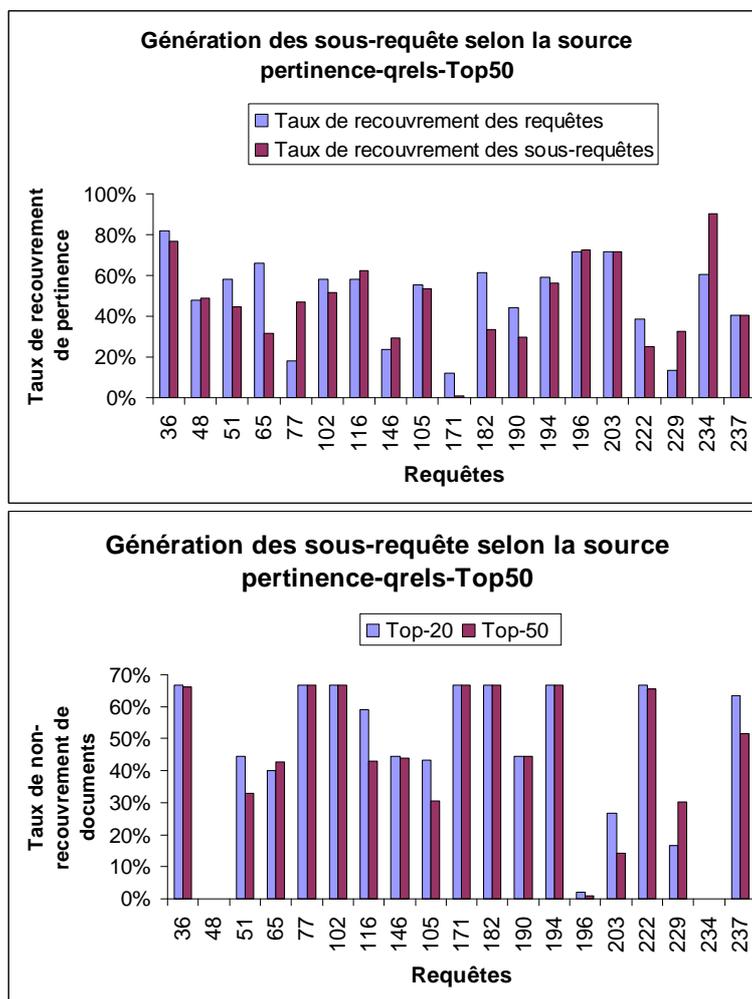


FIG. 7.4 – Taux de recouvrement et de pertinence moyen et de non-recouvrement de documents des sous-requêtes générées à partir de la source *pertinence-qrels-Top50*

différence des deux collections des sous-requêtes, la figure 7.5 présente le taux de recouvrement de pertinence calculé respectivement pour les requêtes principales communes, les sous-requêtes générées à partir de la source *pertinence-qrels* et les sous-requêtes générées à partir de la source *pertinence-qrels-Top50*. Cette comparaison est effectuée dans le but de comparer l'impact de la source de pertinence utilisée sur la sélection des termes des sous-requêtes.

Nous constatons que le taux de recouvrement de pertinence des sous-requêtes générées à partir de la source *pertinence-qrels* est plus élevée que celui des sous-requêtes générées à partir de la source *pertinence-qrels-Top50*, et il est presque similaire à celui obtenu pour les requêtes principales. Ceci est du au fait que le nombre de documents pertinents dans la source *pertinence-qrels* est plus grand que dans la source *pertinence-qrels-Top50*, par conséquent les termes extraits représentant les sous-requêtes associées permettent de cibler la recherche vers

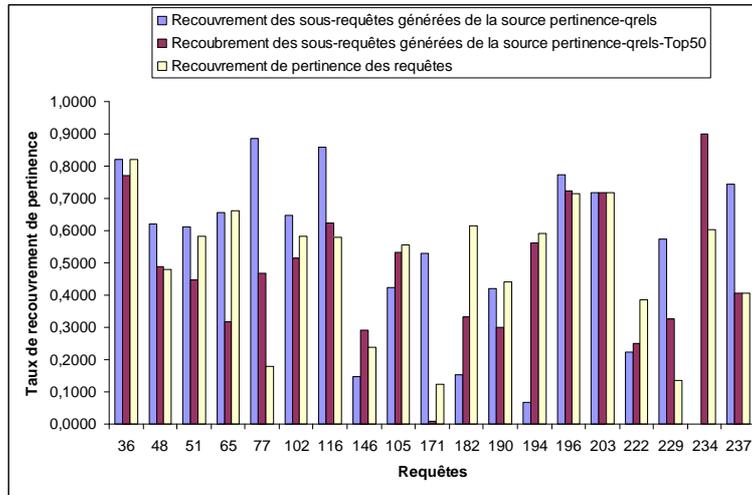


FIG. 7.5 – Comparaison du taux de recouvrement de pertinence moyen des requêtes, des sous-requêtes générées à partir de la source *pertinence-qrels* et des sous-requêtes générées à partir de la source *pertinence-qrels-Top50* retournés par le système

les documents pertinents plus que les termes des sous-requêtes générées à partir de la source *pertinence-qrels-Top50*. En outre, les sous-requêtes générées de cette dernière source recouvrent généralement la même quantité d'information pertinente que les requêtes principales, par conséquent elles ont un niveau de rappel presque similaire à celles-ci.

En conclusion, nous avons généré deux types des sous-requêtes utilisées dans l'évaluation de l'efficacité de la recherche personnalisée.

1. Le premier type est caractérisé par un niveau de rappel plus élevé que la requête principale.
2. le deuxième type est caractérisé par un niveau de rappel similaire aux requêtes principales.

7.2.1.3 Collection de documents

Le corpus HARD comprend des documents contenant des textes issus du *NewsWire 1999*, *AQUAINT corpus* et *U.S. government*. Le tableau 7.2.1.3 résume les caractéristiques de la collection HARD de TREC 2003. Le tableau 7.2.1.3 décrit les caractéristiques de cette collection.

TAB. 7.2 – Caractéristiques de la collection de test TREC HARD 2003

Nombre de documents	1033461
Nombre de requêtes	30
Nombre de termes distincts	592373
Longueur moyenne d'un document	219.894952

7.2.1.4 Les jugements de pertinence

Les jugements de pertinence des requêtes principales sont ceux qui sont définis dans la tâche HARD de TREC 2003. Les sous-requêtes associées à une requête principale partagent avec celle-ci les jugements de pertinence donnés dans la collection. En effet, la sous-requête n'est qu'une reformulation du même besoin en information traité par la requête principale, par conséquent les critères de pertinence des documents retournés par le moteur de recherche en réponse à la sous-requête sont les mêmes que ceux de la requête principale.

7.2.2 Les sessions de recherche

Nous définissons une session de recherche par une séquence de requêtes traitant d'un même sujet de recherche. Dans notre protocole, une session de recherche est simulée par l'alignement de l'ensemble des sous-requêtes associées à une même requête principale.

Nous définissons deux types de séquences de sessions de recherche ; une séquence d'apprentissage et une séquence de test. chaque séquence est créée à partir d'un sous-ensemble de requêtes principales d'apprentissage ou de test par alignement aléatoire des sous-requêtes associées.

1. La *séquence d'apprentissage* est dédiée à l'identification du seuil optimal de délimitation des sessions de recherche.
2. La *séquence de test* permet de créer les profils utilisateurs à partir des sous-requêtes groupées dans une même session selon un seuil identifié sur la séquence d'apprentissage, puis de les exploiter pour personnaliser le processus de RI de toute sous-requête corrélée traitée sur la séquence de test.

7.2.3 Le profil utilisateur

Le principe de construction du profil utilisateur est analogue à celui décrit dans la cadre d'évaluation TREC *ad hoc*. Dans ce cadre précisément,

- la notion de domaine, clairement identifié dans le cas de la collection *TREC ad-hoc* est remplacée par la notion de sujet de la requête principale, non connu *a priori*
- les requêtes associées aux domaines, sont remplacées par les sous-requêtes associées à la requête principale,
- le profil de l'utilisateur est construit tout au long d'une séquence de sous-requêtes groupées dans une même session. Cette méthode de construction du profil est adoptée dans le but d'évaluer l'efficacité du système tout en appliquant la délimitation des sessions de recherche.

7.2.4 Stratégie d'évaluation

Cette section présente la stratégie d'évaluation de la qualité des profils et celle de l'efficacité de la recherche personnalisée.

7.2.4.1 Stratégie d'évaluation de la qualité du profil

L'évaluation de la qualité du profil revient à mesurer la précision de la représentation sémantique du profil compte tenu des sujets de recherche de l'utilisateur définis par les requêtes principales de la collection.

La démarche d'évaluation est similaire à celle proposée dans le cadre d'évaluation TREC adhoc (cf. chap.6 sect.6.2.4.1 p.178). Dans ce cadre précisément,

1. la notion de domaine, clairement identifié dans le cas de la collection TREC ad-hoc est remplacée par la notion de sujet de la requête principale,
2. les requêtes associées aux domaines, sont remplacées par les sous-requêtes associées à la requête principale en cours de traitement.
3. Le profil à évaluer est construit par combinaison de graphes des profils des sous-requêtes associées.

7.2.4.2 Stratégie d'évaluation de l'efficacité de la recherche personnalisée

Cette phase est basée sur l'évaluation de notre approche de personnalisation le long d'une séquence de sessions issue de l'ensemble de requêtes de test traitant de sujets différents. Cette phase est illustrée dans la figure 7.6 et est décrite selon les étapes suivantes :

- Définir la séquence des sous-requêtes de test par alignement des sous-requêtes associées aux requêtes de test principales. L'ordre des requêtes

est défini par leur numérotation donnée par TREC HARD 2003.

- Tout au long de la séquence des sous-requêtes de test, la méthodologie d'évaluation consiste à :
 1. Créer le profil de la sous-requête en utilisant la sous-source des documents pertinents issus de qrels (cf chap.5, sect.5.6.1 p.157),
 2. Initialiser le profil de l'utilisateur par le profil de la première sous-requête d'une session simulée,
 3. Pour chaque sous-requête suivante, appliquer le mécanisme de délimitation des sessions de recherche en utilisant le seuil optimal σ^* identifié dans la phase d'apprentissage (cf. chap.5 sect.5.5.2 p.147).
 4. si la sous-requête appartient à la session définie par la requête principale, le profil de l'utilisateur est utilisé dans le réordonnancement des résultats de cette sous-requête (cf. chap.5 sect.5.8 p.171). Ceci définit la recherche personnalisée d'une sous-requête servant de base à l'évaluation de la performance globale du système.
 5. le profil de l'utilisateur est ensuite mis à jour par combinaison avec le profil de cette nouvelle sous-requête (cf. chap.5 sect.5.6.3 p.162).

Sur la base de l'ensemble des sous-requêtes de la séquence de test, nous comparons la recherche classique à la recherche personnalisée selon des métriques d'évaluation appropriées.

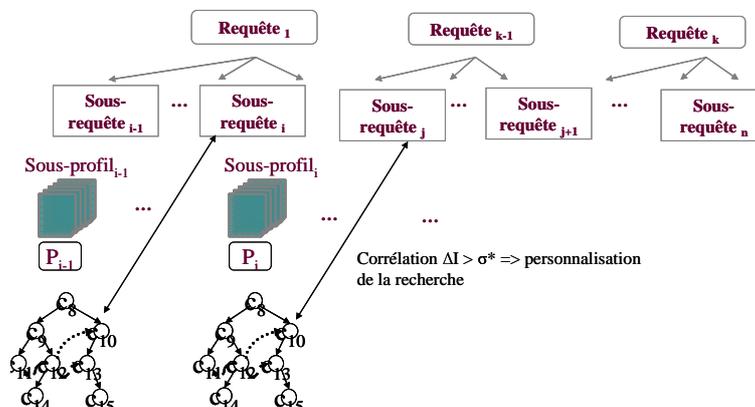


FIG. 7.6 – Phase de test du protocole d'évaluation appliqué sur TREC HARD 2003

7.2.5 Métriques d'évaluation

De même que pour le cadre d'évaluation *TREC Adhoc*, nous utilisons les métriques précédemment proposées pour l'évaluation de la qualité du profil sémantique (P_p10 et $AvgRank_p10$) et des métriques classiques dérivées du rappel-précision pour l'évaluation de l'efficacité de recherche personnalisée. Plus précisément, nous calculons des valeurs de précisions moyenne à n documents

pertinents, $n \in \{10, 20, \dots, 100\}$, sur l'ensemble des requêtes ou des sous-requêtes de test.

Notons que les documents pertinents ayant servis à la création des profils utilisateurs, notamment la source *pertinence* de toutes les requêtes ne sont pas considérés pour l'évaluation des performances. Ceci permet en effet de ne pas biaiser les résultats dans le sens des documents pertinents déjà considérés dans la construction des profils.

7.3 Évaluation de la qualité du profil

Nous évaluons la qualité du profil représenté par un graphe de concepts (cf. chap.5 sect.5.5.1.3 p.145) en appliquant le protocole d'évaluation défini dans la section 7.2.4.2. Pour chaque requête principale, nous évaluons le profil construit sur la base des deux sources de pertinence définies précédemment, la *pertinence-qrels* et le *pertinence-qrels-Top50*. L'évaluation de la qualité des profils selon ces deux types de source de pertinence permet d'étudier l'impact de l'utilisation d'une quantité d'information pertinente importante (*source pertinence-qrels*) non nécessairement retrouvés par le système en réponse à une requête principale donnée dans la construction du profil versus l'impact de l'utilisation d'une plus petite quantité d'information pertinente et retrouvée par le système en réponse à la requête principale (*source pertinence-qrels-Top50*).

7.3.1 Évaluation des profils générés de la source *pertinence-qrels*

Nous présentons dans la figure 7.7 la précision P_p10 et le rang moyen des concepts non-pertinents $AvgRank_p10$ des profils construits à partir de la source *pertinence-qrels* les 15 requêtes de test principales. D'après cette figure, nous constatons ce qui suit :

1. la majorité des profils construits présentent des précisions de plus de 50% à P_p10 ainsi qu'un rang moyen des concepts non-pertinents supérieure à 5.
2. Un seul profil est mal construit correspondant à la requête 84 et dont tous les 10 premiers concepts sont non-pertinents produisant ainsi une précision P_p10 nulle. Le problème est lié à l'ambiguïté de la requête d'une part et la méthode de construction des profils des sous-requêtes qui risque d'activer plus de concepts non pertinents via les liens sémantiques et par conséquent de sélectionner le graphe de concepts non pertinents.

- Les profils correspondants aux requêtes 65 et 77 ne disposent aucun concept non-pertinent sur les premiers 10 concepts produisant ainsi une précision maximale $P_p10 = 1$ et un rang moyen des concepts non pertinents nul $AvgRank_p = 0$.

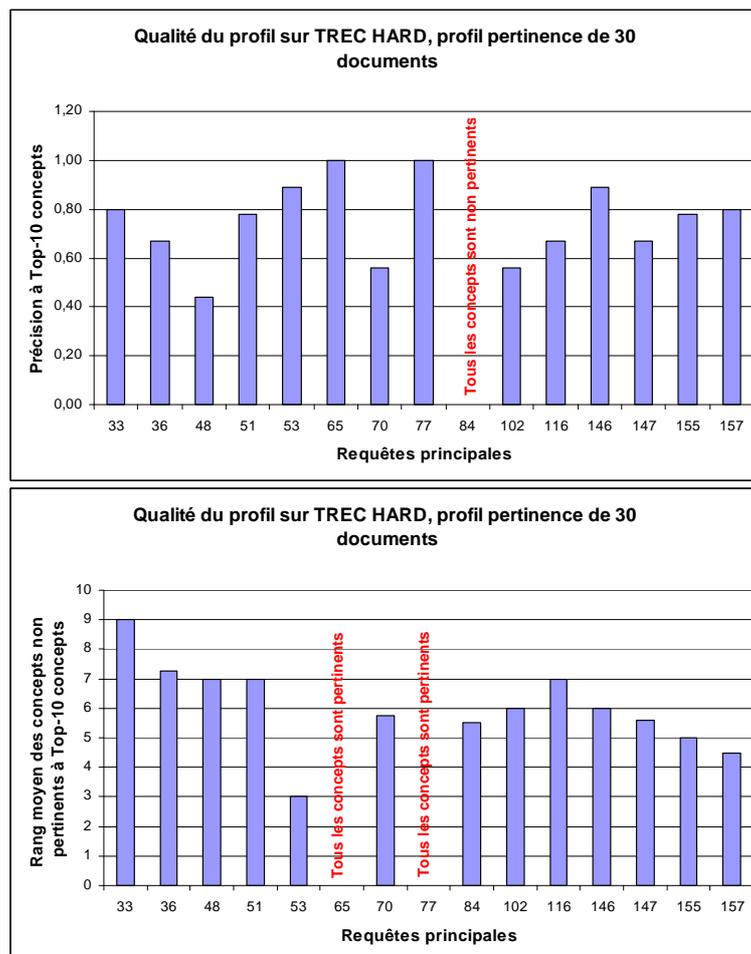


FIG. 7.7 – Qualité des profils construits pour les requêtes principales de la collection TREC HARD en utilisant la source *pertinence-qrrels*

La différence des qualités des profils construits dépend de la méthode d’activation des concepts du profil et de la perception de pertinence des utilisateurs qui ont jugé la qualité des profils.

- notre méthode de construction du profil est basée en partie sur l’activation des liens de référence permettant d’augmenter la représentation du profil par des concepts reliés. Les concepts activés peuvent être loins thématiquement de la requête principale et jugés ainsi non-pertinents.
- La différence de la perception de pertinence entre les utilisateurs a également un impact direct sur l’évaluation de la qualité des profils. Lors de l’évaluation du profil, l’attribution finale du jugement de pertinence au concept est basée sur la technique de vote majoritaire. Si les utilisateurs se divisent

presque également sur le jugement de pertinence d'un concept avec une différence d'un seul jugement négatif, il est probable que le concept sera mal jugé.

7.3.2 Évaluation des profils générés de la source *pertinence-qrels-Top50*

L'évaluation de la qualité des profils à ce niveau consiste à évaluer la précision des profils construits pour les 10 requêtes principales en exploitant la source *pertinence-qrelsTop50* pour chaque requête. Nous présentons dans la figure 7.8 la précision P_{p10} et le rang moyen des concepts non-pertinents $AvgRank_p10$ pour les 10 requêtes principales.

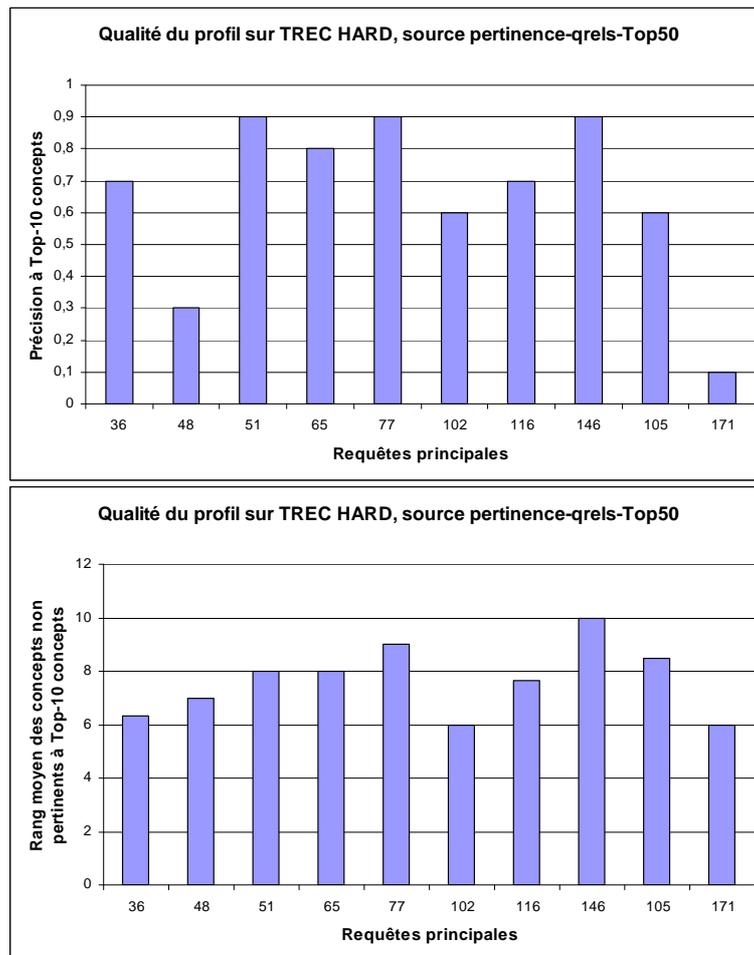


FIG. 7.8 – Qualité des profils construits pour les requêtes principales de la collection TREC HARD en utilisant le profil pertinence de 10 documents

D'après ces résultats, nous constatons le suivant :

1. La majorité des profils construits présentent des précisions de plus de 60% à P_p10 ainsi qu'un rang moyen des concepts non-pertinents supérieure à 7.
2. Aucun profil n'est complètement mal créé sur tous les premiers 10 concepts.

La différence de la précision entre les profils dépend des mêmes facteurs cités dans l'évaluation précédente. En comparant les profils construits de la source *pertinence-qrels* à ceux générées à partir de la source *pertinence-qrels-Top50* pour l'ensemble des requêtes communes entre les deux collections, nous constatons que la précision des profils construits selon la source *pertinence-qrels* est plus importante pour les requêtes 48, 65, 77. Elle est égale ou légèrement moins importante pour le reste des requêtes communes. Ceci confirme effectivement l'impact du nombre de documents pertinents pour certaines requêtes d'une part et l'importance des documents pertinents figurant dans les premiers rangs de la liste de documents retournés par le SRI dans la construction des profils précis même avec un nombre faible de documents pertinents. La différence de l'efficacité d'une source de pertinence par rapport à une autre est dépendante du degré de difficulté des requêtes qui, par l'utilisation d'un plus grand nombre de documents pertinents, produisent des profils plus précis.

7.4 Évaluation de l'efficacité de la recherche personnalisée

L'évaluation de l'efficacité de la recherche personnalisée est effectuée en appliquant le protocole d'évaluation présenté dans la section 7.2.4.2 en utilisant séparément les ensembles des requêtes et des sous-requêtes définies dans la section 7.2.1.2 et le profil représenté par un graphe de concepts (cf. chap.5 sect.5.5.1.3 p.145). Plus précisément, nos objectifs d'évaluation à ce niveau consiste à :

- Évaluer le mécanisme de délimitation des sessions de recherche sur la collection des sous-requêtes générées à partir de la source *pertinence-qrels*,
- Étudier les paramètres de personnalisation, plus particulièrement il s'agit d'étudier :
 1. l'impact du nombre de concepts utilisé dans le calcul du score personnalisé du document (cf. chap.5 formule 5.20 p.173) sur la performance de la recherche personnalisée.
 2. Étudier l'impact du paramètre de combinaison du score original du document avec le score personnalisé (cf. chap.5 formule 5.19 p.172) sur la performance de la recherche personnalisée.
- Évaluer l'effet de la désambiguïsation par activation des liens de référence dans la construction du profil de la requête servant ultérieurement à la construction du profil (cf. chap.5 sect. 5.6.2.2 p.159).

- Évaluer la précision de la recherche personnalisée (basée sur l'utilisation de la requête et du profil associé) comparativement à la précision de recherche classique (basée sur l'utilisation seule de la requête) sur l'ensemble des requêtes principales et des sous-requêtes générées de la source *pertinence-qrels* ou de la *pertinence-qrels-Top50*.

7.4.1 Modèle de recherche classique basé sur BM25

Nous avons mené également nos expérimentations dans ce cadre sur le moteur de recherche "Mercurie" [29]. Le modèle de la recherche classique dans notre système est basé sur la modèle d'appariement de la fonction BM25 défini comme suit :

$$w_{td} = tf_d \times \frac{\log\left(\frac{N-n_t+0.5}{n+0.5}\right)}{K_1 \times ((1-b) + b \times \frac{dl}{avgdl}) + tf} \quad (7.1)$$

où tf_d est la fréquence du terme t dans le document d , N est le nombre total des documents de la collection et n_t est le nombre de documents contenant le terme t , $K_1 = 2$ et $b = 0.75$.

7.4.2 Évaluation du mécanisme de délimitation des sessions de recherche

Cette évaluation est une étape préliminaire qui consiste principalement à déterminer le seuil de corrélation requête-profil optimal à partir d'une séquence des sessions d'apprentissage et évaluer selon ce seuil la précision de délimitation des sessions de recherche. Dans ce but, nous avons créé une séquence aléatoire des 45 sous-requêtes d'apprentissage générées à partir de l'ensemble des 15 requêtes principales de test, tout en exploitant la source *pertinence-qrels*. Notre choix concernant cet ensemble de requêtes est basé sur le fait qu'elles permettent de créer une séquence des sous-requêtes contenant un nombre suffisant des frontières des sessions de recherche servant ainsi à une meilleure identification du seuil de délimitation des sessions de recherche. Selon cette séquence, nous avons 14 frontières de sessions à détecter (TBQ=14) et chaque session contient deux *sous-requêtes* qui doivent être identifiés comme corrélées ce qui permet d'avoir en total 30 *sous-requêtes* corrélées (TRQ=30).

Le principe d'identification du seuil optimal dans cette phase est illustrée dans la figure 7.9 et est décrit selon les étapes suivantes :

- Définir une séquence aléatoire des sessions d'apprentissage par alignement séquentiel des sous-requêtes associées à une même requête principale.

- Tout au long de la séquence de sous-requêtes d'apprentissage définie, calculer la valeur de corrélation (ΔI) entre une sous-requête traitée de la séquence et le profil utilisateur créé sur l'ensemble des sous-requêtes précédentes et issues d'une même requête (cf. chap.5 sect.5.5.2 p.147).
- Pour chaque valeur de corrélation obtenue, calculer la précision de détection des requêtes corrélées P_{intra} et celle de délimitation de sessions de recherche P_{inter} selon les formules suivantes :

$$P_{intra}(\sigma) = \frac{|CQ|}{|TCQ|}, P_{inter}(\sigma) = \frac{|FQ|}{|TFQ|} \quad (7.2)$$

où $|CQ|$ est le nombre de sous-requêtes correctement classifiées comme corrélées, $|TCQ|$ est le nombre total de sous-requêtes qui doivent être identifiées comme corrélées sur la séquence, $|FQ|$ est le nombre de sous-requêtes indiquant des nouvelles sessions et $|TFQ|$ est le nombre total des frontières des sessions de la séquence.

Le seuil de corrélation optimal σ^* est ensuite identifié en maximisant le produit des précisions $P_{intra}(\sigma)$ et $P_{inter}(\sigma)$ comme suit :

$$\sigma^* = \operatorname{argmax}_{\sigma}(P_{intra}(\sigma) * P_{inter}(\sigma)) \quad (7.3)$$

Nous répétons cette procédure selon la méthode de la validation croisée à N itérations pour la création des séquences d'apprentissage à partir de N requêtes principales. Le seuil optimal est finalement calculé comme étant la moyenne des seuils calculés sur l'ensemble des séquences d'apprentissage. Ce seuil est exploité dans la phase de test dans le but de grouper les sous-requêtes de test dans une même session.

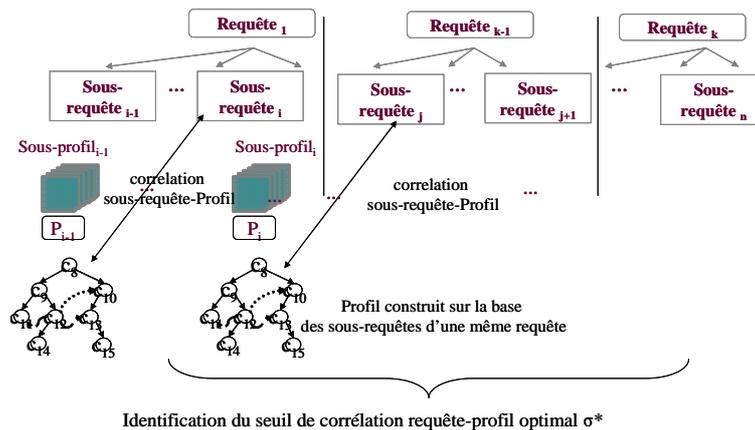


FIG. 7.9 – Phase d'apprentissage du protocole d'évaluation appliqué sur TREC HARD 2003

7.4.2.1 Analyse de corrélation requête-profil

Nous effectuons l'analyse de corrélation requête-profil sur une séquence critique des sessions simulées, qui peut être l'une des séquences créées selon le

principe de validation croisée. Le principe de l’alignement des sessions simulées est fondé sur la corrélation thématique maximale entre sessions successives, ayant pour objectif d’évaluer la précision de délimitation des sessions dans un contexte de basculement entre les sujets de recherche éventuellement difficile à identifier. La procédure d’alignement de la séquence d’apprentissage est décrite comme suit :

1. initialiser le séquençement critique $S =$,
2. choisir aléatoirement une requête principale $q_i \in Q$ puis aligner ses sous-requêtes q_i^1, q_i^2, q_i^3 dans le séquençement S , $S = S \cup \{q_i^1, q_i^2, q_i^3\}$,
3. calculer la corrélation thématique de la requête q_i à chaque requête $q_j \in Q \wedge j \neq i$, soit $Corr(q_i, q_j)$. Cette corrélation est calculée suivant la mesure de délimitation des sessions de recherche entre la requête q_i et le profil créé sur l’ensemble des sous-requêtes générées à partir de la requête q_j .
4. retenir $q_j^* = \text{argmax}(Corr(q_i, q_j))$,
5. aligner les sous-requêtes associées à q_j^* dans S , $S = S \cup \{q_j^{*1}, q_j^{*2}, q_j^{*3}\}$,
6. $Q = Q - q_j^*$, $q_i = q_j^*$,
7. si $Q \neq \emptyset$ aller à 3.

Nous avons créé une séquence critique pour chacune des mesures de corrélation conceptuelle *Kendall* et *Webjaccard* et calculons les valeurs de corrélation requête-profil tout au long de la séquence critique associée. Nous avons appliqué également cette procédure sur la mesure de similarité vectorielle *cosinus* dans le but d’évaluer l’impact de l’utilisation du poids des concepts sur la précision de la délimitation des sessions.

Nous présentons dans la figure 7.10 les valeurs de corrélation requête-profil calculées selon chacune des mesures de *Kendall*, *WebJaccard* et *cosinus* sur la séquence critique associée (composée de 45 sous-requêtes d’apprentissage). Notons que nous avons numéroté une sous-requête par son numéro dans l’ensemble des sous-requêtes générées, notamment $\{1, 2, 3\}$. Une délimitation correcte entre les sessions de recherche simulées est marqué par une ligne verticale. Les valeurs de corrélation selon les mesures de *Kendall*, *WebJaccard* et *Cosine* sont obtenues respectivement dans les intervalles des valeurs $[-0.6 + 0.01]$, $[0.01 \ 0.3]$ et $[0.1 \ 0.9]$.

L’analyse des courbes de corrélation permet de distinguer les valeurs de corrélation obtenues entre des requêtes d’une même session de recherche (corrélation intra-session) et celles obtenues au niveau du basculement entre les sessions de recherche (corrélation inter-session). Généralement la chute de la courbe de corrélation implique une diminution de la corrélation avec le profil et un basculement possible dans les sujets de recherche entre les sessions simulés.

– Corrélation intra-session :

1. *WebJaccard* ou *Cosinus* : selon ces mesures, les valeurs de corrélation intra-session doivent être positives. Pourtant certaines valeurs de

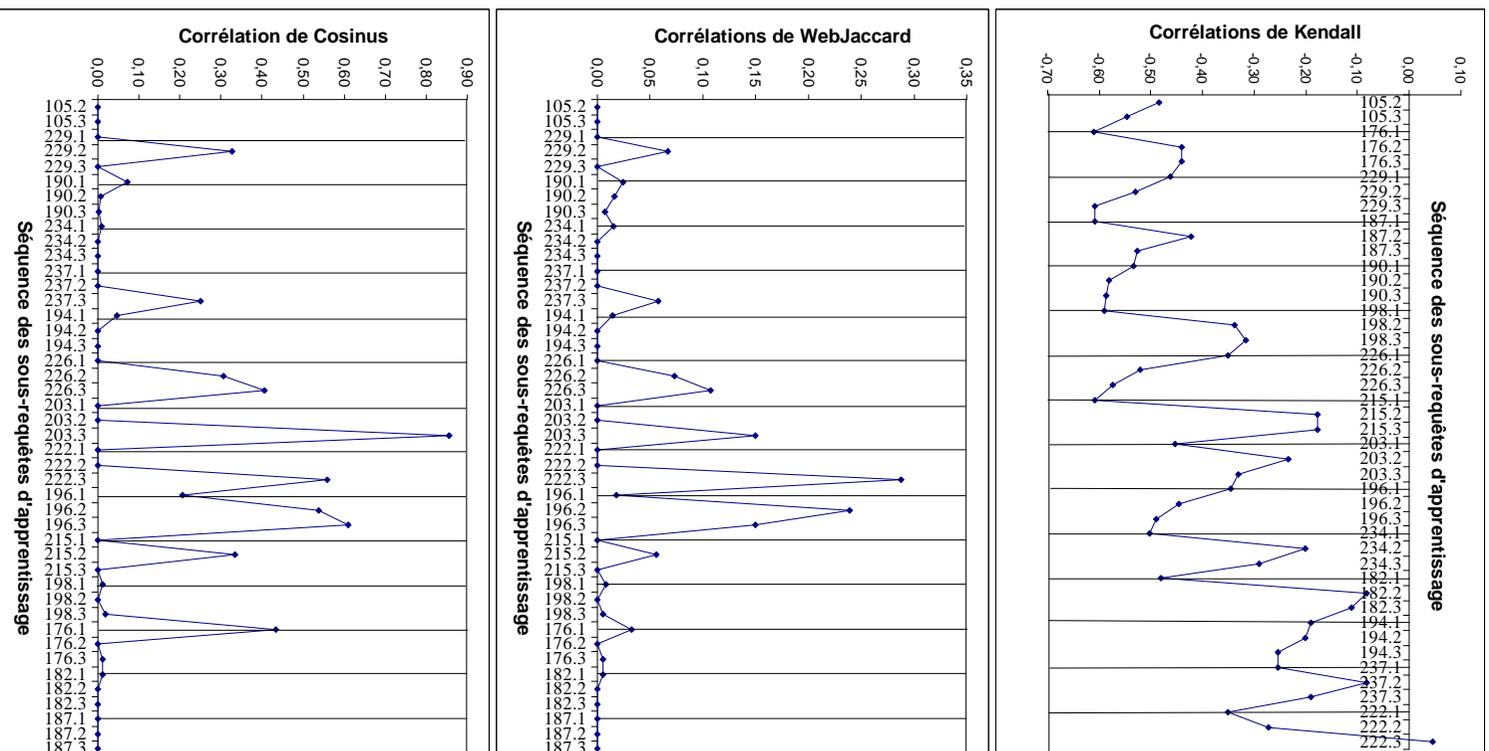


FIG. 7.10 – Analyse de corrélations requête-profil selon les mesures Kendall, WebJaccard et cosinus calculées le long de la séquence des sous-requêtes d'apprentissage

corrélations intra-session sont nulles concernant les requêtes 105, 234, 182, 187. Ceci est dû au fait que certaines sous-requêtes associées à une même requête principale ne disposent pas nécessairement des concepts communs dans l'ontologie de l'ODP.

2. *Kendall* : Selon la mesure de *kendall*, la majorité des valeurs de corrélation sont négatives. Parmi ces valeurs il s'agit des corrélations intra-session correspondant à des petites valeurs de corrélations négatives. En effet, les concepts associés à une requête donnée sont partiellement représentés dans l'ensemble des concepts associés au profil de l'utilisateur. Par conséquent, des valeurs négatives dans une même session simulée ne signifie pas que la requête et le profil sont négativement corrélés (par défaut, une valeur négative de la mesure de *Kendall* entre deux distributions signifient que celles-ci sont négativement corrélées), mais correspond à une valeur de corrélation minimale entre la requête et le profil de l'utilisateur due à la proximité des rangs des concepts communs. Ceci permet de calculer des valeurs de corrélation basées sur le changement des rangs des concepts communs et non pas sur leur nombre (*WebJaccard*) ou leur poids (*cosinus*), par conséquent considérer des corrélations sémantiques possibles entre deux ensembles de concepts même s'ils ne disposent pas des concepts communs.

– Corrélation inter-session :

1. *Cosinus* et *WebJaccard* : selon ces mesures, quand la valeur de corrélation est nulle, il s'agit d'une nouvelle session de recherche. En effet ces mesures sont basées sur la présence des concepts communs entre requêtes successives d'une même session.
2. Selon la mesure de *Kendall*, quand la valeur de corrélation est largement négative à partir d'un certain seuil, il s'agit d'une nouvelle session de recherche. Des grandes valeurs négatives sont calculées sur les frontières des sessions simulées en partant de l'hypothèse qu'il n'existe pas des concepts communs entre une sous-requête donnée et le profil créé sur la session précédente. Ceci n'implique pas les concepts de la requête et du profil sont dans l'ordre inverse (considéré par défaut selon la mesure de *Kendall*). En effet, il est difficile d'avoir le même ensemble de concepts qui figurent à la fois dans la requête et le profil.

7.4.2.2 Évaluation de la précision de la délimitation des sessions de recherche

L'objectif de cette étape est d'évaluer le mécanisme de délimitation des sessions de recherche et de valider que la mesure de *Kendall* utilisée dans ce

mécanisme, permet d'obtenir une précision plus élevée. Cette étape fait partie de la phase d'apprentissage du protocole d'évaluation qui permet d'identifier également le seuil optimal de délimitation des sessions de recherche (cf. chap5 sect.5.5.2 p.147).

Selon les intervalles des valeurs de corrélation obtenues précédemment pour chacune des mesures, nous calculons pour chaque valeur de corrélation σ deux précisions, la précision de détection des requêtes corrélées $P_{intra}(\sigma)$, et la précision de délimitation des sessions de recherche $P_{inter}(\sigma)$. Le seuil qui permet de maximiser ces deux précisions est retenu comme le seuil optimal (σ^*) de ce mécanisme et caractérise la performance de la délimitation des sessions de recherche.

La figure 7.11 montre Les précisions $P_{intra}(\sigma)$ et $P_{inter}(\sigma)$ et le produit des deux précisions ($P_{intra}(\sigma) * P_{inter}(\sigma)$) pour chaque valeur de σ appartenant à l'intervalle des valeurs $[-0.6 + 0.01]$, $[0.01 0.3]$ et $[0.1 0.9]$ respectivement pour la mesure de *Kendall*, *Webjaccard* et *Cosinus*.

les valeurs des précisions de délimitation des sessions de recherche selon les mesures *Kendall*, *Webjaccard* et *cosinus*. sont présentées sur l'axe des Y

Le seuil optimal obtenu pour la mesure de *Kendall* est de $\sigma^* = -0.34$ permettant d'atteindre une précision de classification maximale par rapport à *WebJaccard* et *Cosinus*.

- La précision globale obtenue par *Kendall* est de 45.71% avec des précisions P_{intra} de 53.33% et P_{inter} de 85.71%.
- La précision globale obtenue par *WebJaccard* est de 30% avec une précision P_{intra} de 30% et P_{inter} de 100%.
- La précision globale obtenue par *Cosinus* est de 27.85% avec une précision P_{intra} de 30% et P_{inter} de 92.85%.

Les résultats montrent que notre mesure de délimitation des sessions de recherche (*Kendall*) nous permet effectivement d'identifier des jalons des sessions avec des taux de précision significatifs. Ceci montre l'importance du changement des rangs des concepts d'une requête à une autre pour détecter une nouvelle session comparativement aux mesures basées sur l'existence des concepts communs entre deux requêtes d'une même session. En effet, cette mesure calcule une valeur de corrélation requête-profil plus élevée quand les concepts dominants du profil sont représentés dans les premiers rangs du vecteur conceptuel de la requête, et une valeur de corrélation plus faible si ces concepts n'occupent pas les premiers rangs. Les mesures *WebJaccard* et *Cosinus* présentent des limites dans certaines situations.

1. La mesure de *WebJaccard* est basée sur le nombre de concepts communs entre deux ensembles de concepts en ignorant leur rangs. Ceci implique que quelque soit l'ordre des concepts communs entre la requête et le

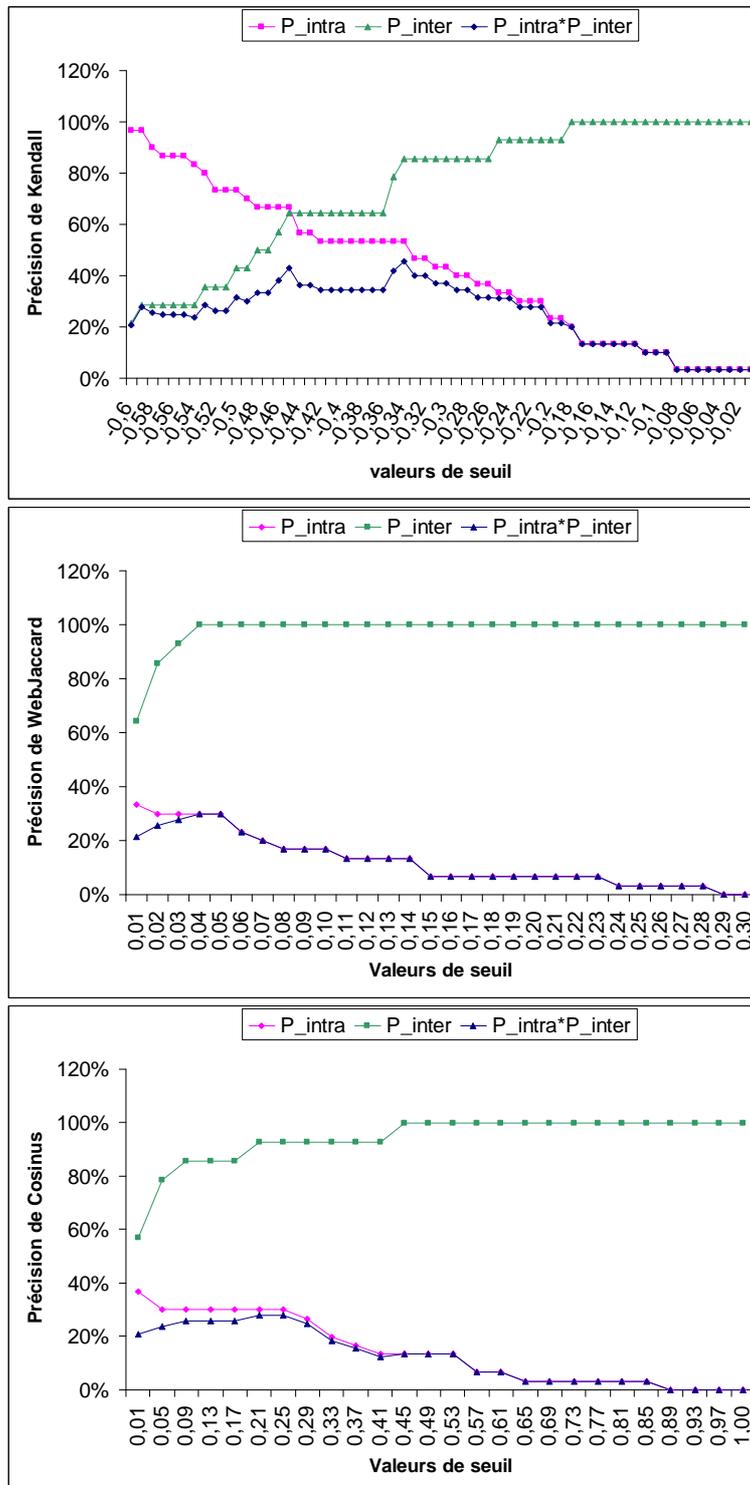


FIG. 7.11 – Précision de classification des sessions selon les mesures de *Kendall* et *WebJaccard* et *Cosinus* en fonction de la variation du seuil de corrélation

profil, *WebJaccard* donne la même valeur, ce qui ne traduit pas le fait que l'utilisateur augmente son intérêt pour des concepts spécifiques au

cours de la session de recherche.

2. La mesure de *Cosine* prend en compte les poids des concepts dans les deux vecteurs représentant la requête et le profil et donne des valeurs différentes pour deux vecteurs pondérés différemment et ayant le même ordre des concepts. Ceci implique que si l'utilisateur a diminué son intérêt dans certains concepts tout en conservant le même ordre des concepts du profil d'une requête à une autre, la mesure de *Cosine* détecte un changement du sujet de recherche sans qu'il y ait changement de l'importance des concepts entre une requête et un profil.

D'après les résultats obtenus, nous utilisons le seuil optimal identifié selon la mesure de *Kendall* dans l'évaluation de l'efficacité de recherche personnalisée sur les sous-requêtes générées à partir du *profil pertinence-qrels* dans la phase de test du protocole d'évaluation dans le cadre TREC HARD.

7.4.3 Étude des paramètres de personnalisation

Nous identifions dans cette section les valeurs optimaux des paramètres de personnalisation permettant d'obtenir la meilleure personnalisation. Ces paramètres concernent :

1. Le nombre de concepts considérés dans le calcul du score personnalisé.
2. Le paramètre de combinaison linéaire du score personnalisé et du score original des résultats de recherche.

Pour chacun des paramètres, l'identification de la valeur optimale est effectuée sur la collection de requêtes de test principales données par HARD TREC 2003. Pour chaque requête principale de test, le profil est construit par combinaison des profils de ses sous-requêtes générées à partir de la source *pertinence-qrels* ou de la source *pertinence-qrels-Top50*. Ceci permet de comparer plus tard l'efficacité de recherche personnalisé en exploitant chacun de ces deux types de profils sur l'ensemble de requêtes et de sous-requêtes de test. Nous n'avons pas mené cette étude sur la séquence des sous-requêtes dans le but d'éliminer l'impact du seuil de délimitation des sessions de recherche appliqué sur les séquences des sessions simulées sur l'analyse des résultats d'une part et d'identifier les valeurs optimales des paramètres de personnalisation en utilisant des requêtes difficiles d'autre part.

7.4.3.1 Étude de l'impact du nombre de concepts utilisés dans le score personnalisé

Le but de cette expérimentation est d'identifier la valeur optimale du nombre de concepts utilisés dans le calcul du score personnalisé du document sur l'ensemble des requêtes principales de test (cf. formule 5.20 chap. 5, p. 173). Le

modèle de recherche personnalisée à ce niveau est basé seulement sur le score personnalisé du document.

Les figures 7.12 et 7.13 montrent les taux de précision de la recherche personnalisée à P10, P20 et P30 ainsi que la MAP en exploitant respectivement pour chaque requête de test la source *pertinence-qrels* et la source *pertinence-qrels-Top50*.

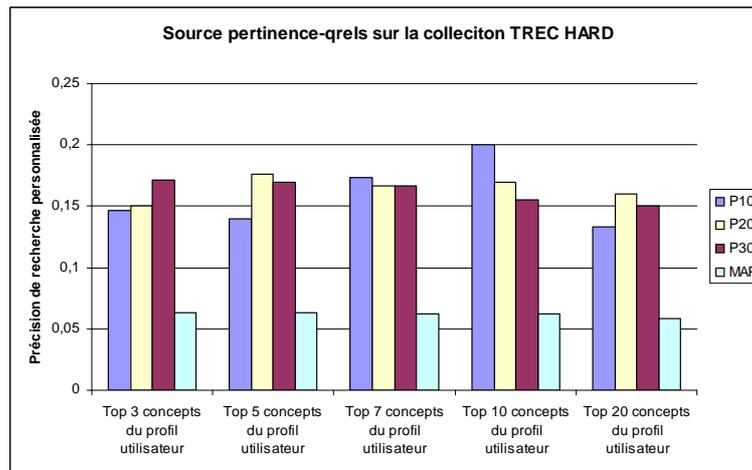


FIG. 7.12 – Efficacité de la recherche personnalisée en fonction du nombre de concepts considérés dans le calcul du score personnalisé en utilisant la source *pertinence-qrels* sur la collection TREC HARD

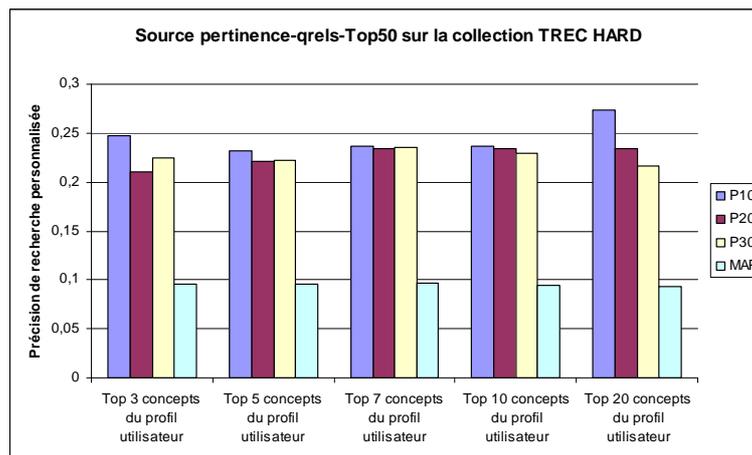


FIG. 7.13 – Efficacité de la recherche personnalisée en fonction du nombre de concepts considérés dans le calcul du score personnalisé en utilisant la source *pertinence-qrels-Top50* sur la collection TREC HARD

D’après les résultats présentés dans ces figures, nous constatons que :

1. La meilleure valeur de la précision à P10, P20 et P30 ainsi que la meilleure valeur de la MAP est obtenue en utilisant 10 concepts du profil dans le cal-

cul du score personnalisé du document en exploitant la source *pertinence-qrels*.

2. La meilleure valeur de la précision à P10 est obtenue en utilisant 20 concepts du profil dans le calcul du score personnalisé du document en exploitant la source *pertinence-qrels-Top50*.

Cette différence est dûe principalement à l'impact de l'utilisation des documents retrouvés par le système issus de la source *pertinence-qrels-Top50* sur la précision du profil. En effet, l'utilisation de ces documents permet de construire un profil plus précis que celui construit en utilisant la source *pertinence-qrels* à P_p20 et qui atteint la meilleure performance.

Nous choisissons de retenir 10 concepts comme étant la valeur optimale du nombre de concepts utilisés dans le calcul du score personnalisé du document pour les raisons suivantes :

1. L'utilisation de 10 concepts dans le calcul du score personnalisé du document donne une meilleure précision P20 et P30 et une précision MAP quasi similaire comparativement à l'utilisation de 20 concepts.
2. L'utilisation de 10 concepts permet de réduire le temps d'exécution par rapport à l'utilisation de 20 concepts.

7.4.3.2 Étude de l'impact du paramètre de réordonnement des résultats

Le but de cette expérimentation est d'identifier la valeur optimale du paramètre de réordonnement des résultats γ (cf. chap.5 formule 5.19 p.172).

Nous calculons l'efficacité de la recherche personnalisée en fonction de la variation du paramètre γ dans l'intervalle [0 1]. Le modèle de la recherche personnalisée est basée sur la combinaison du score original du document et de son score personnalisé en fixant le nombre de concepts utilisé dans le calcul du score personnalisé à la valeur optimale identifiée dans l'expérimentation précédente ($h = 10$).

Les figures 7.14 et 7.15 montrent les valeurs de la précision de la recherche personnalisée obtenues respectivement en exploitant séparément la source de *pertinence-qrels* et la source de *pertinence-qrels-Top50*.

D'après les courbes de résultats obtenus, nous constatons ce qui suit :

1. la valeur nulle du paramètre de réordonnement ($\gamma = 0$) correspond à la recherche personnalisée basée sur l'ordonnement des résultats selon le score personnalisé du document. Cette valeur permet d'obtenir la meilleure valeur de la précision P10 pour les deux sources de pertinence

étudiées. Plus la valeur de γ augmente ($\gamma > 0$), les valeurs de précisions P10, P20 et P30 baissent généralement ce qui montre l'importance du score personnalisé dans l'amélioration de performance de la recherche personnalisée.

2. En utilisant la source *pertinence-qrels*, nous constatons une stabilité des courbes de précision à partir de $\gamma = 0.1$. La précision P10 est maximale pour $\gamma = 0$ et les précisions P20 et P30 sont maximales pour $\gamma = 0.1$. Nous privilégions la valeur optimale de γ produisant la meilleure précision à P10, soit $\gamma = 0$.
3. En utilisant la source *pertinence-qrels-top50*, nous constatons une stabilité des courbes de précision en fonction de γ à P10, P20 et P30. La valeur optimale de γ est fixée à $\gamma = 0$.

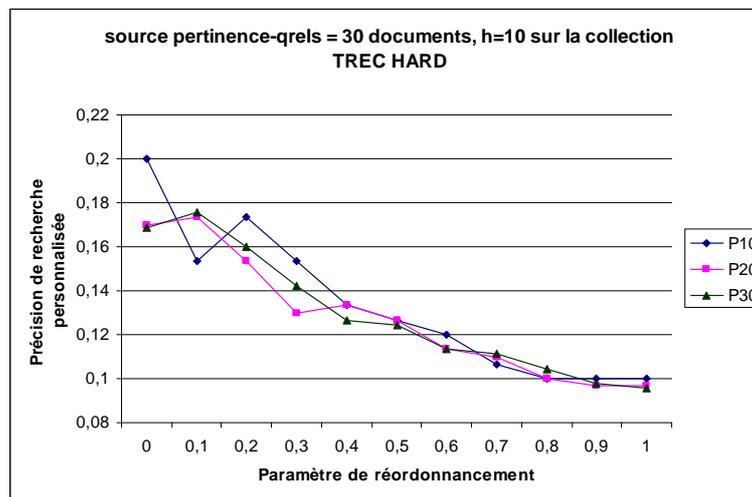


FIG. 7.14 – Efficacité de la recherche personnalisée en fonction du paramètre de réordonnement en utilisant la source pertinence-qrels documents sur la collection TREC HARD

Nous rappelons que la valeur optimale du paramètre de réordonnement des résultats dans la collection *TREC Adhoc* est $\gamma = 0.1$ (cf. chap.5 sect.6.4.3.2 p.191). Nous constatons que les valeurs optimales de γ sont faibles sur les deux collections privilégiant ainsi le score personnalisé. Toutefois, la différence des valeurs optimales identifiées entre les deux collections est principalement lié à la spécificité du profil appris par rapport à la requête de test. Plus précisément :

1. Dans le cadre de la collection TREC Adhoc, il est utile d'utiliser le score original des documents restitués en réponse à une requête d'un domaine puisque le profil associé est construit sur la base des requêtes restantes du même domaine, considéré ainsi moyennement général par rapport au sujet de la requête en cours d'évaluation. La combinaison des deux scores du document permet de privilégier ceux qui sont spécifiques à la requête via leur score original et qui répondent en même temps aux centres d'intérêts du profil via leur score personnalisé.

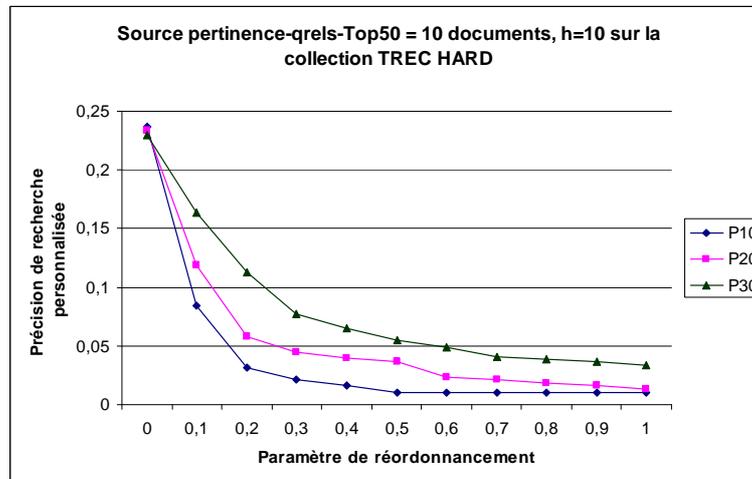


FIG. 7.15 – Efficacité de la recherche personnalisée en fonction du paramètre de réordonnement en utilisant la source pertinence-qrels-Top50 sur la collection TREC HARD

2. Dans le cadre de la collection *TREC HARD*, le score personnalisé du document répond directement à la spécificité de la requête puisqu'il est construit à partir des documents pertinents associées à la requête principale elle-même.

7.4.4 Evaluation de l'effet de la propagation du score sur la recherche personnalisée

L'objectif de cette expérimentation est d'évaluer l'effet de la construction du profil de la requête représenté par un graphe de concepts via l'expansion des concepts associés par activation des liens de référence et la sélection du graphe des concepts les plus pertinents (cf. chap.5 sect. 5.6.2.2 p.159).

Dans cette expérimentation, nous avons utilisé l'ensemble des 15 requêtes de test principales issus de la source *pertinence-qrels*. Nous n'avons pas utilisé les séquences de sous-requêtes dans le but d'évaluer l'effet de la désambiguïsation du profil sur un ensemble de requêtes effectivement difficiles d'une part et d'exclure l'impact du seuil de délimitation des sessions de recherche sur l'analyse des résultats d'autre part.

Dans ce but, nous comparons la performance de la recherche personnalisée en exploitant respectivement le profil construit sans appliquer la propagation de scores, noté *profil non propagé*, le profil construit avec l'application de la propagation des scores, noté *profil propagé* et le profil construit selon notre méthode, noté *profil propagé avec extraction du graphe* construit en appliquant

la propagation des scores et l'extraction du graphe des concepts les plus pertinents. Les méthodes de dérivation de ces trois types de profils sont décrits comme suit :

1. Le *profil non propagé* est représenté par un ensemble des concepts initialement pondérés sans appliquer la propagation des scores. Le profil associé à chaque requête principale de test est construit par combinaison des *profils non propagés* des sous-requêtes.
2. Le *profil propagé* est représenté par l'ensemble de concepts initialement pondérés et augmentés par des concepts activés par propagation des scores de ces concepts sur les liens de référence. Le profil associé à chaque requête principale de test est construit par combinaison des *profils propagés* de ses sous-requêtes sans extraire le graphe maximal ayant la somme des poids des concepts la plus élevée.
3. Le profil construit selon notre approche, nommé *profil propagé avec extraction de graphe*, est représenté par l'ensemble de concepts initialement pondérés et augmentés par des concepts activés par propagation des scores suivi par l'extraction du graphe des concepts les plus pertinents. Le profil associé à chaque requête principale de test est construit par combinaison des profils propagés de ses sous-requêtes et identifié finalement par extraction du graphe maximal ayant la somme des poids des concepts la plus élevée.

La figure 7.16 montre les valeurs de la précision moyenne et du rappel moyen obtenues pour la recherche personnalisée en utilisant les trois types de profils. Nous constatons une précision moyenne élevée pour la personnalisation avec le profil construit selon notre approche, suivi par celle du *profil propagé* comparativement au *profil non propagé* sur la majorité des précisions à n documents. Ces résultats montrent l'impact positif de la méthode de désambiguïsation du profil. Plus précisément,

1. L'expansion de la représentation du profil avec des nouveaux concepts sémantiquement liés dans le *profil propagé* permet d'améliorer aussi bien la précision que le rappel de la recherche personnalisée comparativement à la non propagation de scores dans le *profil non propagé*. Plus précisément, elle permet d'améliorer de manière stupéfiante la précision à P10.
2. L'extraction du graphe le plus pertinent à la requête dans le *profil propagé avec extraction du graphe* assure une efficacité de la recherche personnalisée beaucoup plus significative au niveau de la précision à P10 ainsi qu'au niveau du rappel comparativement aux deux autres profils.

En comparant les trois types de profils, nous remarquons que le taux d'amélioration est significatif sur les premiers 10 documents présentés à l'utilisateur en utilisant les profils propagés. Au delà de la précision à P20, l'amélioration de la précision est variable jusqu'à P80. Elle tend finalement à ramener les valeurs de précision obtenues par les différents profils au même niveau. En termes du rappel, il est plus élevé en utilisant notre profil sur la majorité des points de rappel pour devenir moins important à R90 et R100.

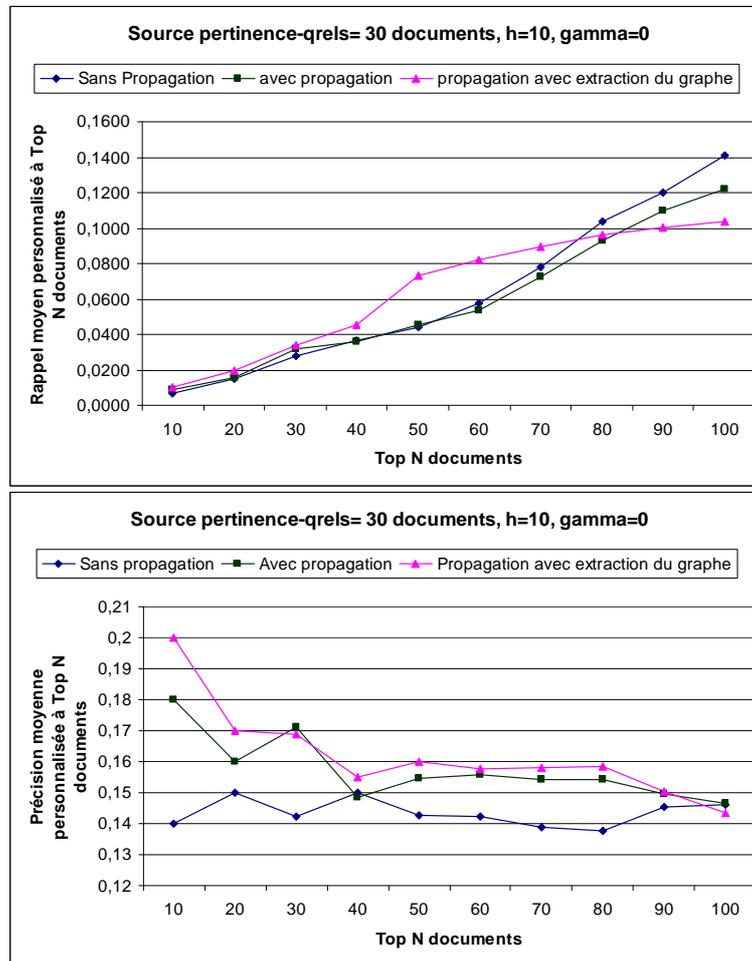


FIG. 7.16 – Validation du processus de propagation des scores lors de la construction du profil sur la collection TREC HARD

En s'intéressant à la précision au premiers documents retournés par le système, nous constatons que la propagation des scores dans le profil a une tendance de ramener plus de quantité d'information pertinente en haut de la liste retournée par le système par rapport à la non propagation de scores. En plus l'extraction du graphe pertinent représentant le profil de l'utilisateur permet d'assurer une qualité du profil significativement plus précise et efficace dans la personnalisation de la recherche.

7.4.5 Évaluation de l'efficacité de la personnalisation

L'évaluation de l'efficacité de la recherche personnalisée consiste à comparer la précision obtenue par la recherche personnalisée à la précision obtenue par la recherche classique. Cette évaluation est effectuée sur l'ensemble des sous-requêtes générées à partir de la source *pertinence-qrels* ou de la source

pertinence-qrels-Top50 ainsi que sur les requêtes principales de la collection. L'utilisation des requêtes principales permet d'évaluer la performance de la recherche personnalisée en utilisant des requêtes difficiles. En outre, l'évaluation de la performance de la RI personnalisée en comparant deux les sources de pertinence permet d'évaluer :

- l'impact du volume des documents pertinents exploités dans la construction du profil de l'utilisateur fixé respectivement dans la source *pertinence-qrels* et *pertinence-qrels-Top50* à 30 et au plus 9 documents pertinents,
- l'importance de l'utilisation des documents pertinents classés dans les premiers rangs de la liste des résultats retournés en réponse à la requête.

Le protocole d'évaluation selon le type des requêtes utilisées est le suivant :

1. Évaluation sur la séquence des sous-requêtes de test : le scénario d'évaluation consiste à appliquer la phase de test dans le cadre d'évaluation TREC HARD (sect. 7.2.4.2 p.213). Pour chaque séquence des sous-requêtes de test générées respectivement à partir de la source *pertinence-qrels* ou de la source *pertinence-qrels-Top50*, un seuil de délimitation des sessions optimal est identifié comme étant la moyenne des seuils calculés sur l'ensemble des séquences créés itérativement par validation croisée.
2. Évaluation sur l'ensemble de requêtes de test principales : pour chaque requête principale de test, le profil utilisé dans le processus de personnalisation est construit sur l'ensemble des sous-requêtes en exploitant la source *pertinence-qrels* ou la source *pertinence-qrels-Top50*.

7.4.5.1 Évaluation sur la séquence des sous-requêtes de test

L'évaluation de l'efficacité de la recherche personnalisée est effectuée sur la base de deux séquences de sous-requêtes de test générées respectivement à partir de la source de *pertinence-qrels* et de la source *pertinence-qrels-Top50*.

1. Utilisation de la source *pertinence-qrels* : selon cette source, la séquence des sous-requêtes de test est composée de 45 sous-requêtes alignés en exploitant les 15 requêtes principales de test. Le seuil moyen de délimitation des sessions de recherche obtenu selon la méthode de la validation croisée à 15 itérations est estimé à $\sigma^* = -0.35$. Selon ce seuil, nous avons obtenu une précision de groupage des requêtes corrélées de 56.66% et une précision de détection des frontières des sessions de recherche de 71% sur la séquence des sous-requêtes de test.

Nous présentons dans la figure 7.17 la précision moyenne et le rappel moyen sur les n premiers documents, $n \in \{10, 20, \dots, 100\}$, présentés à l'utilisateur obtenus par la recherche personnalisée en appliquant le seuil moyen sur la séquence des sous-requêtes de test. Nous avons calculé également le pourcentage de gain obtenu pour chacune des précisions et rappel calculées.

Les résultats montrent un gain de performance significatif aussi bien selon la mesure du rappel que de la précision. Plus particulièrement, la meilleure amélioration est obtenue à $P20$ (22%) et $R20$ (12%). La chute du taux du rappel moyen à 10 documents est du effectivement au réordonnement des résultats de certains sous-requêtes avec des profils non précis. En effet, du à la précision de détection des frontières des sessions décelée à 71.42%, l'ensemble des 28.58% des sous-requêtes restantes sont personnalisées avec des profils construits à partir des sous-requêtes appartenant à une ou plusieurs requêtes principales traitant des besoins en information différents. Cependant, nous avons obtenu une amélioration de la MAP de 13% pour l'ensemble des sous-requêtes de la séquence. Nous pouvons conclure que la qualité du profil sémantique est efficace dans le sens où le réordonnement des résultats de recherche en utilisant seulement les premiers concepts du profil permettent d'obtenir une performance de recherche personnalisée plus élevée que la recherche classique.

2. Utilisation de la source *pertinence-qrels-Top50* : selon cette source, la séquence des sous-requêtes de test est composée de 30 sous-requêtes alignés en exploitant les 10 requêtes principales de test. Le seuil moyen de délimitation des sessions de recherche obtenu selon la méthode de la validation croisée à 10 itérations est de $\sigma^* = -0.18$. Selon ce seuil, nous avons obtenu une précision de groupage des requêtes corrélées de 35% et une précision de détection des frontières des sessions de recherche de 77% sur la séquence des sous-requêtes de test. Nous constatons ainsi un nombre faible des sous-requêtes, soit 35%, effectivement personnalisées. Ceci est du effectivement à la valeur négative faible du seuil moyen identifié selon la méthode de la validation croisée comparativement à la valeur du seuil identifié en utilisant la source *pertinence-qrels*, soit $\sigma^* = -0.35$. La différence des seuils est principalement liée à la significativité des termes des sous-requêtes générées en exploitant la source *pertinence-qrels-Top50*. En effet, le taux de recouvrement de pertinence de ces sous-requêtes comparativement aux sous-requêtes générées en exploitant la source *pertinence-qrels* est relativement faible. Par conséquent, la sélection des termes des sous-requêtes est biaisée dans le sens où le calcul de la corrélation conceptuelle entre sous-requêtes successives issues des requêtes différentes ne révèle pas un début d'une nouvelle session. Dans le but d'évaluer l'efficacité de la recherche personnalisée en exploitant un seuil négatif non biaisé, nous avons calculé le seuil minimal obtenu sur les séquences des sous-requêtes d'apprentissage générées selon la méthode de la validation croisée à 10 itérations. Le seuil minimal est ainsi estimé à -0.39 et appliqué sur la séquence des sous-requêtes de test. Selon ce seuil, nous avons obtenu une précision de groupage des requêtes corrélées de 55% et une précision de détection des frontières des sessions de recherche de 66.66% sur la séquence des sous-requêtes de test.

La figure 7.18 présente la précision moyenne et le rappel moyen sur les n premiers documents présentés à l'utilisateur en utilisant respectivement

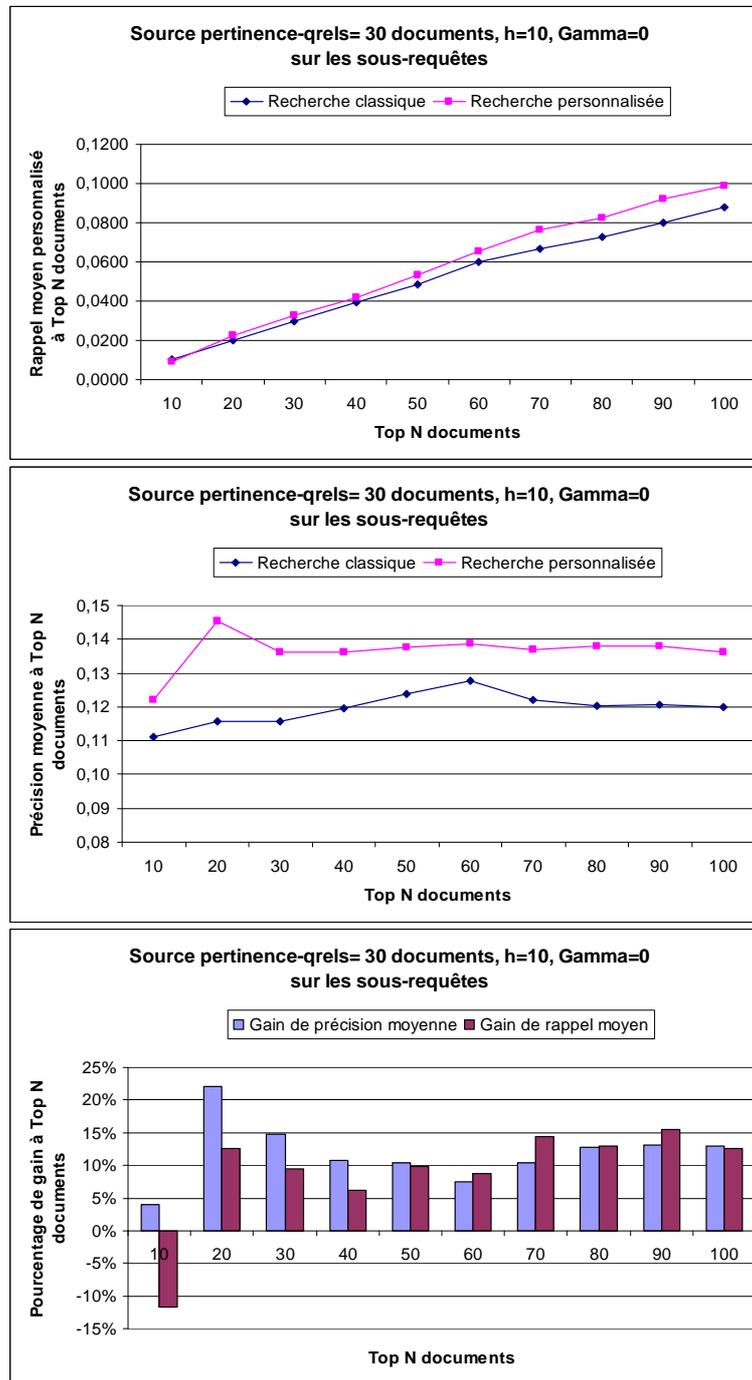


FIG. 7.17 – Comparaison de la recherche personnalisé à la recherche classique en termes de la précision moyenne et rappel moyen sur la séquence des sous-requêtes de test

le seuil moyen ($\sigma^* = -0.18$) et le seuil minimal ($\sigma^* = -0.39$). Les gains de précision et de rappel correspondants sont également présentés dans la figure 7.19. Les résultats montrent que la précision à P10 de la recherche personnalisée dépasse la recherche classique pour les deux valeurs de seuils

à plus de 100%. L'amélioration de la précision est également significative à P20 pour devenir négatif sur le reste des précisions calculées. Notons que l'amélioration de la MAP est estimée à 8.27% et 4.32% en utilisant respectivement le seuil -0.39 et -0.08 .

Nous remarquons que l'amélioration de la précision et du rappel est d'autant plus importante en utilisant le seuil minimal que le seuil moyen. La différence de performance obtenue entre les deux seuils utilisés est principalement liée à la précision de groupage des sous-requêtes dans une même session et par conséquent à la qualité des profils utilisés dans la personnalisation des sous-requêtes ainsi qu'au nombre de sous-requêtes effectivement personnalisées.

- le seuil minimal de délimitation des sessions ($\sigma^* = -0.18$), permet de détecter plus de frontières des sessions que des requêtes corrélées. Dans ce cas, le profil utilisé dans la personnalisation est appris sur un nombre faible de sous-requêtes précédentes, par conséquent, sa précision risque donc d'être plus faible s'il est appris sur des sous-requêtes issues d'une requête principale différente traitant un besoin en information différent. En outre, le nombre faible des sous-requêtes mal personnalisées réduit grandement le gain de la précision si certaines sous-requêtes sont jugées corrélées à des profils non effectivement adéquats selon le seuil moyen.
- le seuil moyen de délimitation des sessions ($\sigma^* = -0.39$) permet de détecter plus de requêtes corrélées que des frontières des sessions. Dans ce cas, même si le profil est appris sur des sous-requêtes appartenant à des requêtes principales différentes, il reste en partie pertinent à une sous-requête de test jugée corrélée au profil selon ce seuil.

En comparant le pourcentage de gain de la précision et de rappel obtenus par la recherche personnalisée entre les deux sous-collections des sous-requêtes, nous constatons qu'il est plus important à P10 et R10 en exploitant la source *pertinence-qrels-Top50* et il est moins important sur le reste des gains calculés. Ceci montre que même avec un nombre faible de documents pertinents, si ces documents sont retrouvés par le système en réponse à la requête (source *pertinence-qrels-Top50*), ils servent à la construction des profils efficaces. La chute des gains sur les points des précisions restantes est due à l'impact du seuil de délimitation des sessions de recherche d'une part et de la taille de la source pertinence utilisée dans la construction des profils entre les deux collections des sous-requêtes d'autre part. En exploitant la source *pertinence-qrels-Top50*, le seuil minimal ($\sigma^* = -0.39$) permet de grouper plus de sous-requêtes non corrélées et le seuil moyen ($\sigma^* = -0.08$) permet de détecter plus des sessions simulées produisant ainsi des profils moins efficaces comparativement aux profils construits en exploitant la source *pertinence-qrels* avec le $\sigma^* = -0.35$.

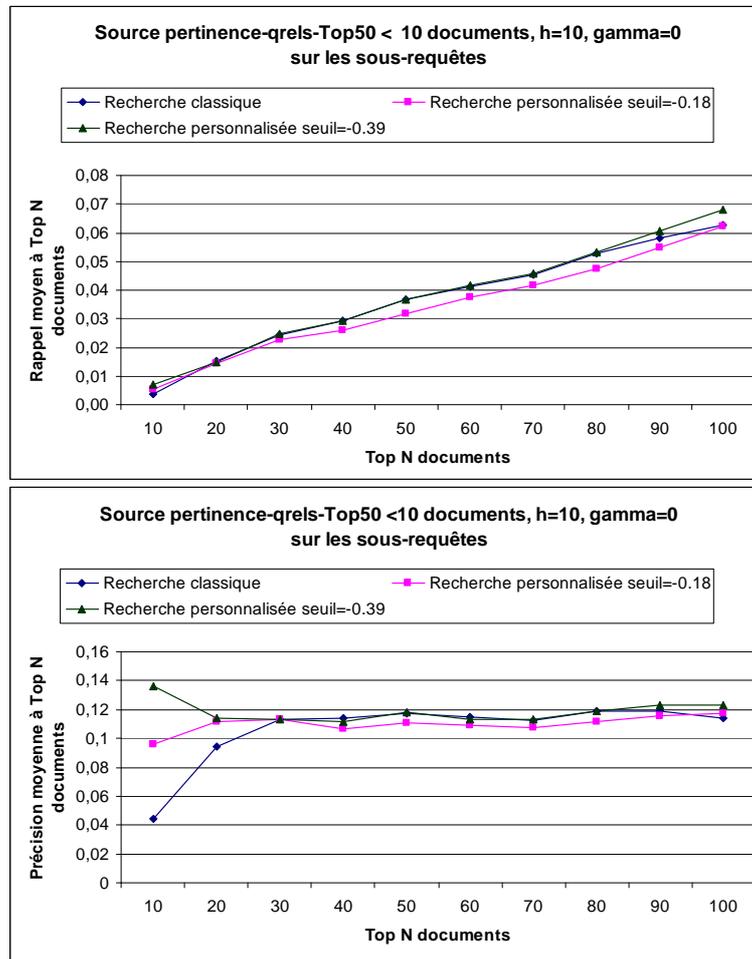


FIG. 7.18 – Comparaison de la recherche personnalisé à la recherche classique en termes de la précision moyenne et rappel moyen sur la séquence des sous-requêtes de test

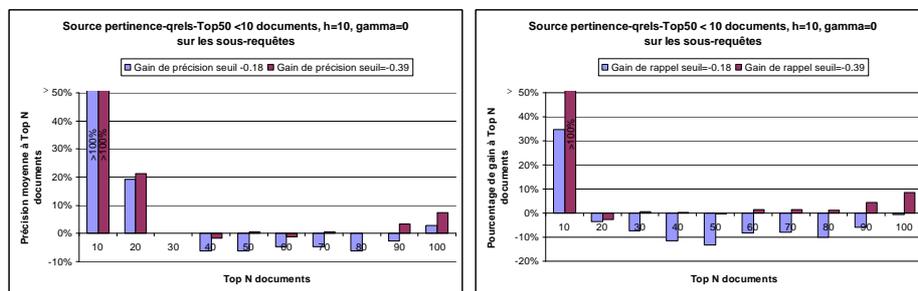


FIG. 7.19 – Comparaison de la recherche personnalisé à la recherche classique en termes de la précision moyenne et rappel moyen sur la séquence des sous-requêtes de test

7.4.5.2 Évaluation sur l'ensemble des requêtes principales de test

L'évaluation de l'efficacité de la recherche personnalisée à ce niveau est effectuée en utilisant les deux ensembles des requêtes de test principales et où le profil est construit en exploitant respectivement la source *pertinence-qrels* et la source *pertinence-qrels-Top50*.

1. Utilisation de la source *pertinence-qrels* : la figure 7.20 présente les valeurs des précisions et des rappels obtenues par la recherche personnalisée comparativement à la recherche classique sur l'ensemble des 15 requêtes principales de test. Nous constatons un gain de performance significatif pour toutes les précisions et rappels calculés. Le meilleur gain est obtenu à P10 et R10 dépassant ainsi 100%. D'après ces résultats, nous pouvons confirmer que notre méthode de personnalisation est efficace sur les premiers rangs des documents retournés par le système et même sur des requêtes difficiles, notamment les requêtes de la collection TREC HARD 2003.
2. Utilisation de la source *pertinence-qrels-Top50* : la figure 7.21 présente les valeurs des précisions et des rappels obtenues par la recherche personnalisée comparativement à la recherche classique sur l'ensemble des 10 requêtes principales de test. Nous constatons également que notre modèle de recherche personnalisée améliore grandement la performance de recherche sur toutes les précisions moyennes et rappels moyens calculés. Le gain de performance est significativement positif, plus particulièrement sur les premiers 10 documents présentés à l'utilisateur.

En l'absence de l'impact du seuil de délimitation des sessions de recherche et de l'incertitude des sous-requêtes de test, la comparaison entre les gains de précision et de rappel obtenus en exploitant respectivement la source *pertinence-qrels* et la source *pertinence-qrels-Top50* montre à nouveau l'efficacité de la recherche personnalisée en exploitant les documents effectivement retrouvés par le système en réponse à une requête et dans des situations idéales de délimitation des sessions de recherche (profil appris sur des sous-requêtes appartenant à une même requête principale). L'évaluation de l'efficacité de notre modèle de RI personnalisée selon la source de *pertinence-qrels-Top50* peut être considérée comme une situation de recherche simulant un contexte de recherche réel où les documents *cliqués* ou visualisés par l'utilisateur sont remplacés par les documents pertinents retrouvés par le système en réponse à la requête.

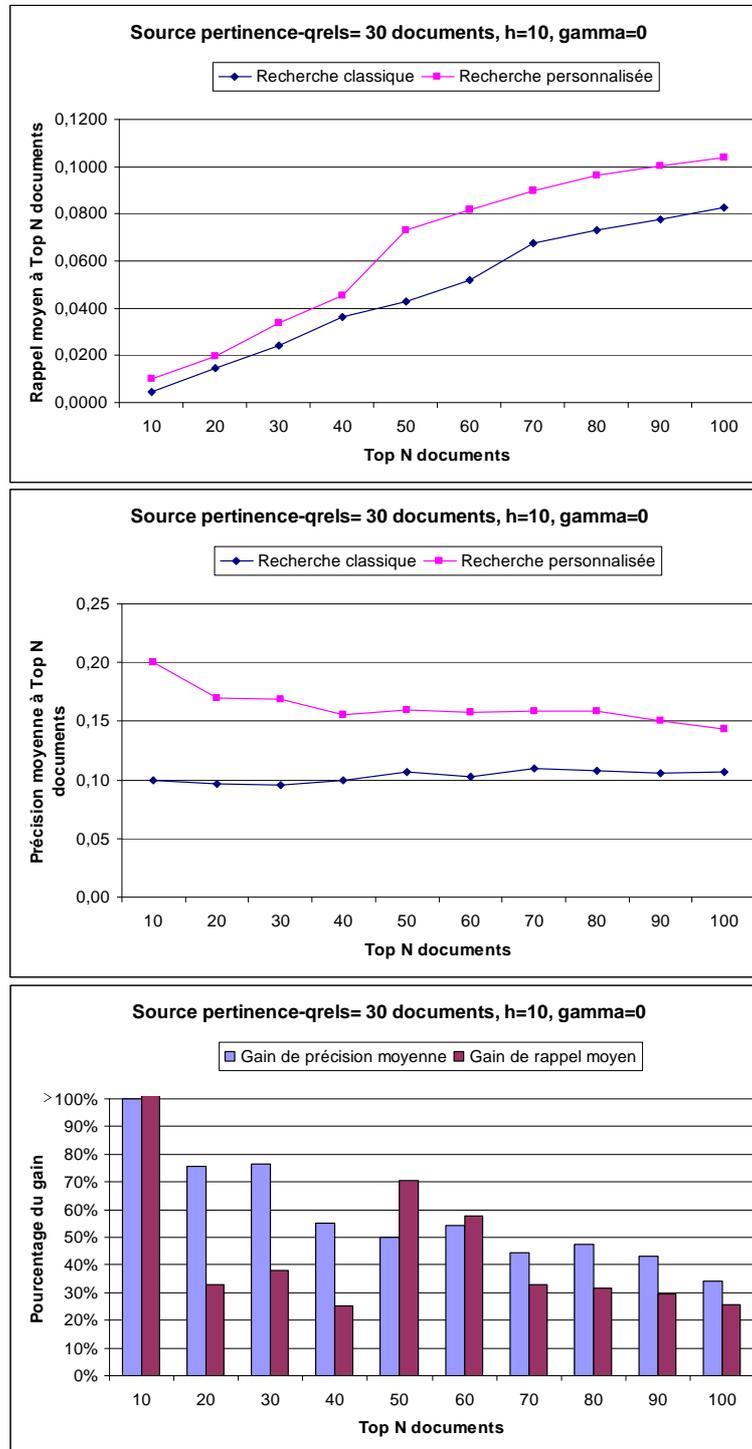


FIG. 7.20 – Comparaison de la recherche personnalisé à la recherche classique en termes de la précision moyenne et rappel moyen sur la séquence des requêtes principales de test

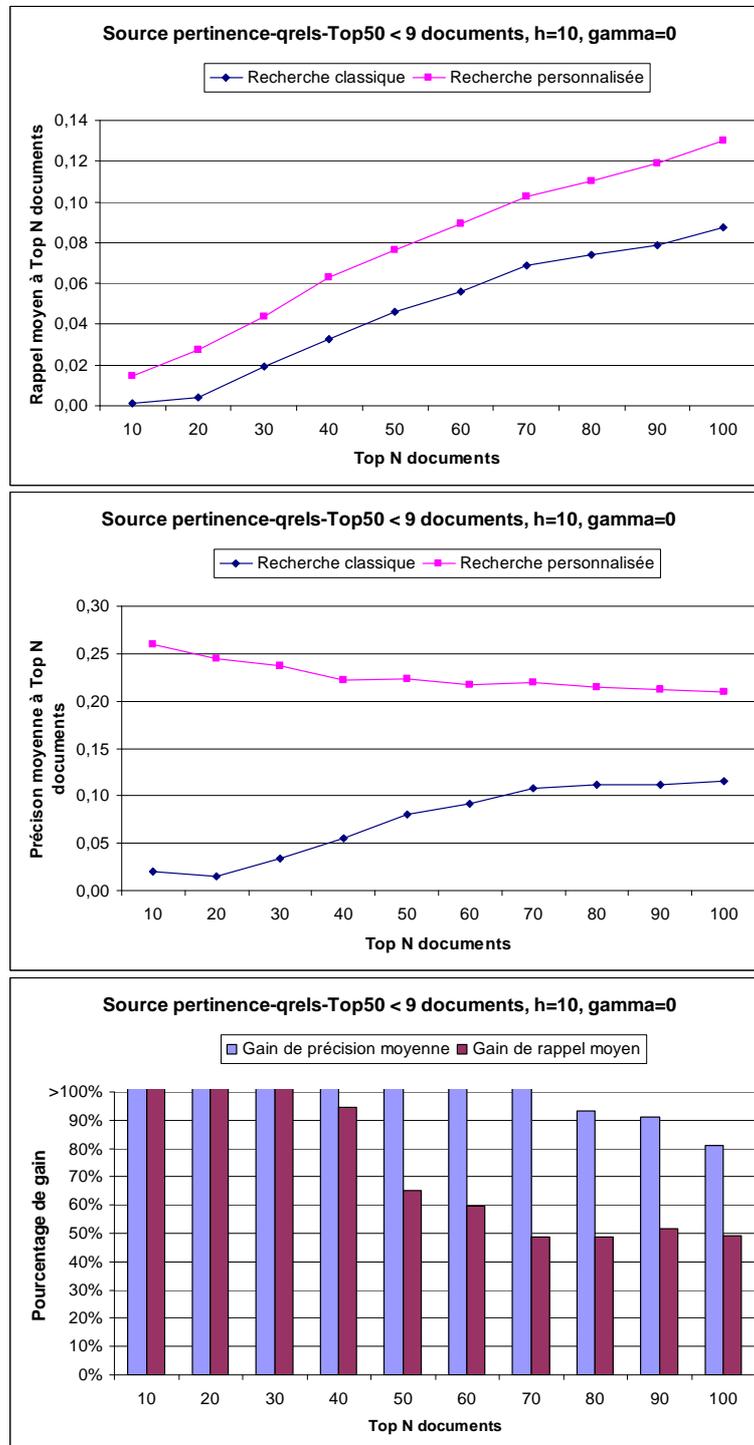


FIG. 7.21 – Comparaison de la recherche personnalisé à la recherche classique en termes de la précision moyenne et rappel moyen sur les requêtes principales de test

7.5 Une évaluation comparative avec l'approche de *Gauch*

L'objectif de cette évaluation est de comparer l'efficacité de notre approche à une approche proposée dans le domaine décrite dans [67] tout en utilisant des requêtes difficiles, notamment les requêtes principales de la collection TREC HARD 2003. Plus précisément, nous comparons l'efficacité de la recherche classique obtenue en utilisant seulement la requête principale à l'efficacité de notre approche de personnalisation, et celle obtenue par la méthode proposée dans [67], notée approche de *Gauch03*.

Les raisons pour lesquelles nous avons choisi de comparer notre approche à celle de *Gauch03* sont les suivantes :

1. le profil utilisateur est représenté sémantiquement par un ensemble de concepts issus de l'ontologie de domaine Web .
2. la personnalisation du processus de RI est basée sur un modèle de réordonnement des résultats de recherche permettant de combiner le score original du document et son score personnalisé calculé en fonction de son degré de similarité avec les concepts du profil.

7.5.1 Aperçu de l'approche de (Gauch et al. 03)

Dans l'approche de *Gauch03* [67], le profil de l'utilisateur décrit les centres d'intérêts de l'utilisateur à court terme dans un intervalle de temps précis. Il est représenté sémantiquement par un ensemble de concepts issus d'une ontologie de domaine Web construite à partir de plusieurs taxonomies de concepts, notamment, l'hierarchie de concepts de Yahoo, Lycos, Magellan et l'ODP.

La construction et l'évolution du profil de l'utilisateur est à la base de l'exploitation des documents visualisés ou *cliqués* par l'utilisateur au cours de ses sessions de recherche. En effet, la dérivation du profil utilisateur est décrite par les étapes suivantes :

1. Collecter les documents/pages web visités ou *cliqués* par l'utilisateur durant un intervalle de temps précis. Ces pages Web sont considérées les sources d'évidence de base qui permettent de dégager les centres d'intérêts de l'utilisateur.
2. Classifier les pages Web collectées dans les concepts appropriés de l'ontologie de référence utilisée. En effet, les pages Web et les concepts de l'ontologie sont représentés dans l'espace vectoriel, chacun par un vecteur de termes pondérés selon le schéma classique de pondération $tf * idf$. Le

principe de la classification des pages web dans les concepts de l'ontologie est le suivant :

- calculer la similarité cosinus entre le document et chaque concept représenté dans l'ontologie.
- gérer l'évolution des poids des concepts pondérés au cours des sessions de recherche. Cette étape consiste à accumuler les scores de similarité des pages Web dans les premiers 5 concepts les plus similaires. Par conséquent, le poids d'un concept de l'ontologie représente les scores de similarité agrégés issus de la classification de plusieurs pages Web sous ce concept.

La personnalisation du processus de recherche consiste ensuite à réordonner les résultats de recherche restitués par le système pour une requête q en utilisant le profil de l'utilisateur. Pour chaque document résultat r restitué par le système, le score final du document est calculé en fonction de son score original, et un score personnalisé. Le score personnalisé est calculé en fonction de la similarité du document avec les premiers concepts du profil de l'utilisateur ainsi que leur poids. Le score final new_{wt_r} d'un document r est calculé selon la formule suivante :

$$new_{wt_r} = wt_r * (0.5 + \frac{1}{4} \sum u_{c_{rl}}) \quad (7.4)$$

où

wt_r est le score original du document r restitué par le système,

$u_{c_{rl}}$ est le poids du concept c_{rl} dans le profil de l'utilisateur,

c_{rl} est un concept parmi les l concepts les plus similaires au document r .

7.5.2 Évaluation comparative et résultats

Dans le but de mettre en place une évaluation comparative fiable, nous adoptons un scénario d'évaluation qui consiste à utiliser les mêmes sources d'information pour la construction du profil de l'utilisateur dans les deux approches sur l'ensemble des 15 requêtes principales de la collection TREC HARD 2003. Nous avons utilisé la source *pertinence-qrels* pour chaque requête du fait qu'il correspond à un ensemble de requêtes principales suffisant pour comparer la performance de deux systèmes. Le principe de l'évaluation est le suivant :

1. Dans notre approche, le profil de l'utilisateur est construit par combinaison des profils de ses sous-requêtes générées à partir de la source *pertinence-qrels*.
2. Dans l'approche de *Gauch03*, le même ensemble de documents pertinents dans la source *pertinence-qrels* est exploité dans la construction du profil de l'utilisateur comme une instance de l'ontologie.
3. Le profil est utilisé dans le réordonnement des résultats de recherche de la requête principale en appliquant la formule 5.19 (cf. chap.5 sect. 5.8

p.172) dans notre méthode et la formule 7.4 dans l'approche de *Gauch03* (p.242).

Nous présentons dans la figure 7.22 la moyenne des précisions à n documents restitués selon la recherche classique, la recherche personnalisée selon notre approche et selon l'approche de *Gauch03*. Nous calculons les valeurs de précision et de rappel à n documents restitués par le système pour $n \in \{10, 20, \dots, 100\}$. Les résultats montrent que les deux approches de personnalisation dépassent la performance de la recherche classique. Le pourcentage d'amélioration de la performance de recherche en termes de précision à n documents et rappel à n documents selon les deux approches est représenté dans la figure 7.23.

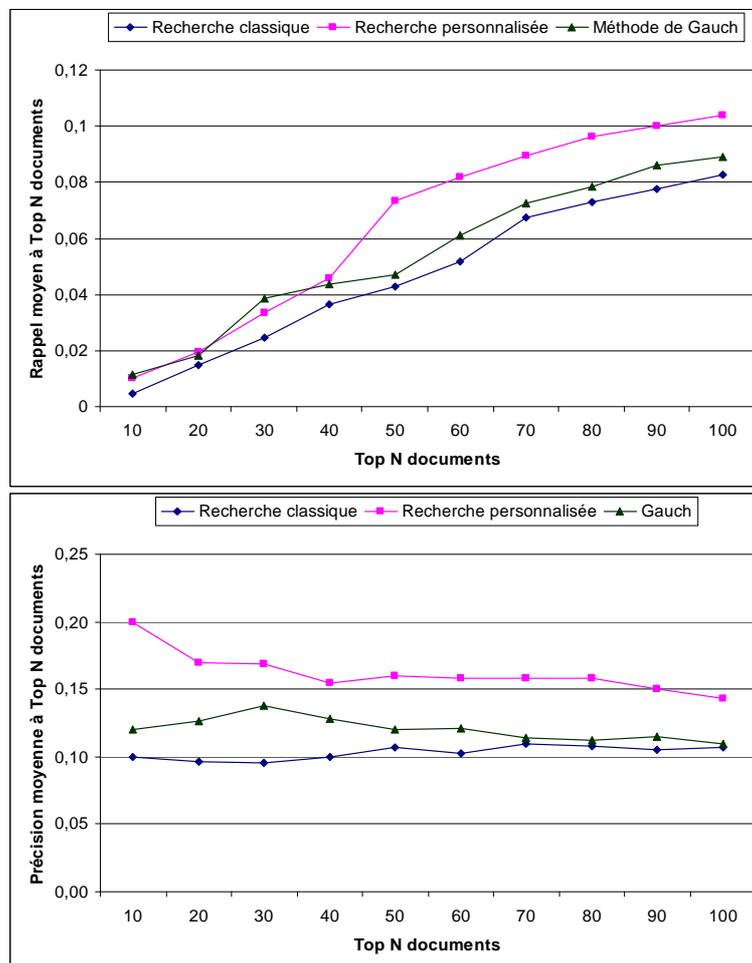


FIG. 7.22 – Comparaison de la performance de notre approche de personnalisation à l'approche de (gauch et al. 03) en termes de la précision moyenne et rappel moyen sur la séquence des requêtes principales de test

Nous constatons que notre approche permet d'obtenir des performances plus élevées que l'approche de *Gauch03*. Le meilleur gain de performance obtenu selon notre approche dépasse 100% à P_{10} et à R_{10} . L'approche de *Gauch03* donne un gain de performance plus petit à P_{10} estimée à 20% mais un meilleur gain de

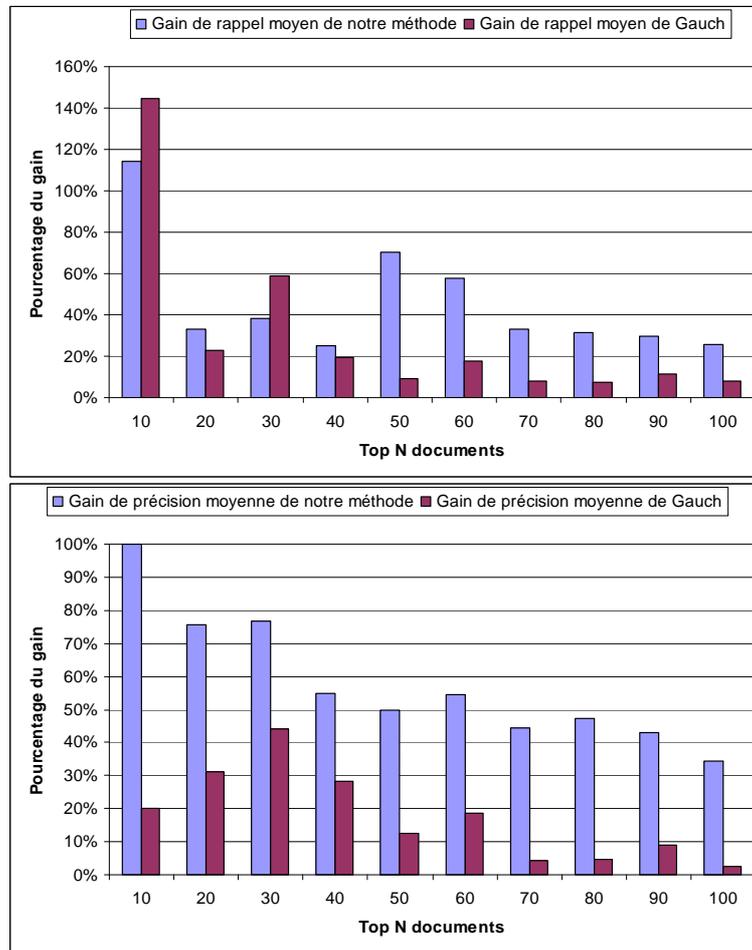


FIG. 7.23 – Comparaison du pourcentage de gain de performance obtenu par notre approche de personnalisation et l’approche de (Gauch et al. 03) sur l’ensemble des requêtes principales de TREC HARD 2003

performance à $R10$ dépassant 100%. La différence de performance entre les deux approches est due à plusieurs facteurs liés principalement à la représentation et la dérivation du profil de l’utilisateur d’une part et au calcul du score personnalisé du document d’autre part.

1. notre modèle de représentation enrichit la représentation du profil de l’utilisateur par des concepts activés qui sont inscrits dans une variété des thématiques proches selon une méthode de propagation des scores via les liens de référence de l’ontologie. Ceci permet à la fois, selon notre point de vue, de diversifier les aspects thématiques d’un même centre d’intérêt d’une part et de représenter les relations entre les les différentes concepts thématiques du profil d’autre part.
2. Le réordonnement des résultats de recherche consiste dans notre approche à utiliser la similarité du document avec les concepts du profil ainsi que les poids des concepts du profil. L’approche de *Gauch03* ignore le score de similarité du document avec les concepts du profil en exploitant

seulement leur poids dans le calcul du score final du document.

7.6 Conclusion

Nous avons présenté au cours de ce chapitre l'évaluation de l'efficacité de notre approche d'accès personnalisé à l'information selon le cadre d'évaluation issu de TREC HARD 2003. Nous avons également comparé notre approche à une autre approche de personnalisation proposée dans le domaine intégrant également un profil utilisateur sémantique et un processus de réordonnement des résultats de recherche.

Plus précisément, nous avons évalué l'efficacité de recherche personnalisée à travers une séquence de sessions de recherche simulées traitant différents besoins en information. Les résultats expérimentaux montrent bien l'efficacité des différentes composantes de notre approche, notamment :

1. une précision significative de la qualité des profils construits sur la base des sujets de recherche portés par les requêtes principale de la collection TREC HARD 2003.
2. une précision du mécanisme de délimitation des sessions de recherche.
3. l'impact positif de la méthode d'expansion des concepts et d'extraction du graphe maximal lors de la dérivation du profil de l'utilisateur sur la qualité des profils ainsi que sur l'efficacité de la recherche personnalisée.
4. un gain de performance significatif de la recherche personnalisée comparativement à la recherche classique sur l'ensemble des requêtes principales de la collection ainsi que sur la séquence des sous-requêtes. Ce gain est d'autant plus important que les documents servant à la construction du profil figurent parmi les premiers documents retrouvés par le système en réponse à une requête donnée. Ceci montre l'efficacité de notre approche dans un cadre de recherche réaliste et où le profil est construit à partir de quelques documents *cliqués* ou *visualisés*.
5. la comparaison de notre approche à une approche proposée dans le domaine révèle également une performance de recherche plus élevée validant ainsi l'impact positif de différents paramètres de notre modèle. Nous citons l'impact de la représentation du profil selon un graphe de concepts, la méthode d'expansion et de désambiguïsation des concepts lors de la construction du profil de l'utilisateur, l'évolution du poids des concepts dans une même session ainsi que la fonction du calcul du score personnalisé du document.

Conclusion générale

Synthèse

Les travaux présentés dans ce mémoire se situent dans le contexte général de l'accès personnalisé à l'information et plus particulièrement dans le cadre de l'utilisation des profils utilisateurs sémantiques à court et à long terme pour une recherche d'information personnalisée. Nos contributions présentées dans ce mémoire ont porté sur trois volets : l'identification du type de besoin derrière la requête, la modélisation du profil utilisateur à court et à long terme et l'évaluation du processus d'accès personnalisé à l'information.

1. Le type de besoin derrière la requête peut être informationnel lié à la recherche documentaire classique, navigationnel lié à la recherche des sites d'accueil ou transactionnel lié à la recherche des services en ligne. Dans notre approche, la prédiction du type de besoin derrière la requête consiste essentiellement à exploiter non seulement les caractéristiques linguistiques et morphologiques de la requête mais aussi le contexte de la session défini par le type de besoin des requêtes précédentes issues de la même session. Les caractéristiques morphologiques de la requête sont définies généralement par des critères dépendant de la longueur de la requête, l'étiquetage grammatical (*POS tagging : part-of-speech tagging en Anglais*), l'existence des termes interrogatifs informationnels ou des termes transactionnels, etc. L'estimation du type de besoin selon les caractéristiques morphologiques de la requête consiste ainsi à évaluer le nombre de critères morphologiques satisfaits par la requête parmi les critères définis. Selon le contexte de la session, l'estimation du type de besoin est à la base du calcul d'une probabilité de basculement d'un type de besoin à un autre type en fonction d'une longueur moyenne des sessions du type des requêtes précédentes. L'estimation finale du type de besoin de la requête est basée sur la combinaison selon la règle de Bayes de son estimation selon la morphologie et selon le contexte de la session.
2. La modélisation du profil utilisateur est basée sur l'exploitation de son historique de recherche et d'une ontologie de domaines Web permettant de dégager une représentation sémantique de ses centres d'intérêts à court et à long terme, selon une liste de concepts ou un graphe de concepts.

La construction du profil à court terme consiste à collecter les documents jugés pertinents implicitement pour un ensemble de requêtes liées à un même besoin en information définissant ainsi une session de recherche, et ensuite les exploiter pour extraire les concepts les plus pertinents issus de l'ontologie. L'évolution du profil à court terme est basée sur l'augmentation de l'importance des concepts récurrents dans la session et l'extraction des concepts reliés par des relations sémantiques prédéfinies dans l'ontologie. Cette évolution est conjointement liée à un mécanisme de délimitation des sessions de recherche basée sur une mesure de corrélation des rangs des concepts qui permet d'évaluer le degré de changement des centres d'intérêts entre requête et profil. Les profils construits à court terme et qui sont récurrents au cours des sessions de recherche de l'utilisateur permettent ainsi de définir le profil à long terme. La personnalisation du processus de recherche consiste finalement à exploiter le profil utilisateur dans le réordonnement des résultats de recherche. Il s'agit d'identifier un profil à long terme adéquat à une requête récurrente initiant une session de recherche selon une mesure de distance sémantique entre graphes de profils issus de l'ontologie ou exploiter le profil construit à court terme pour une requête inscrite dans la même session.

3. Le dernier volet de notre contribution consiste en la définition d'un cadre d'évaluation approprié pour l'accès personnalisé basé-session. Le cadre proposé consiste à réutiliser les ressources de la campagne d'évaluation TREC, plus particulièrement la tâche HARD de TREC 2003. Le choix de cette collection a été motivé par le fait que les requêtes sont difficiles et que le but est d'exploiter des sources d'évidence additionnelles autre que la requête pour l'amélioration de la recherche. Toutefois, nous n'avons pas exploité les métadonnées prédéfinies pour chaque requête de la collection qui consistent en des préférences qualitatifs liées à l'information recherchée puisque le contexte défini dans notre approche décrit les centres d'intérêts évolutifs de l'utilisateur. Dans ce sens, nous avons associé à chaque requête de la collection, un profil utilisateur simulant ainsi un centre d'intérêt de l'utilisateur. L'évaluation de notre modèle d'accès personnalisé est effectuée selon un protocole simulant des sessions de recherche par alignement séquentiel des sous-requêtes générées à partir d'une même requête de la collection. A partir des jugements associés à la requête, on construit automatiquement un profil sur la base des sous-requêtes jugées inscrites dans une même session selon le mécanisme de délimitation des sessions de recherche tout en utilisant un seuil prédéfini. Nous avons alors appliqué ce cadre pour évaluer précisément notre modèle et avons présenté les résultats expérimentaux obtenus. Nous avons appliqué également un protocole d'évaluation précédemment proposé pour l'évaluation de notre modèle d'accès personnalisé à l'information basée sur la réutilisation de la collection TREC 1. Le choix de cette collection est basé sur le fait que les requêtes sont décrites par un champ particulier qui spécifie un domaine d'intérêt permettant ainsi de simuler un profil

utilisateur. Le protocole d'évaluation est basé sur le principe de la validation croisée indépendamment du mécanisme de délimitation des sessions de recherche. Les expérimentations menées portent essentiellement sur l'évaluation de la qualité des profils construits et l'évaluation des performances ainsi que l'impact des différents paramètres du modèle. Les résultats obtenus montrent que notre modèle est efficace et performant pour personnaliser l'accès à l'information. Il est d'autant plus efficace que les requêtes sont corrélées à un même centre d'intérêt. Nous avons pu également définir des jalons d'apprentissage des profils à court terme via le mécanisme de délimitation des sessions de recherche.

Limites et perspectives

L'évaluation expérimentale de notre contribution a montré son efficacité selon plusieurs aspects, et ouvrent des perspectives portant sur l'amélioration de la prédiction du type de besoin et la modélisation du profil de l'utilisateur d'une part et l'évaluation du modèle d'accès personnalisé à l'information dans un scénario de recherche réaliste d'autre part.

Plus particulièrement, nos perspectives portent essentiellement sur les volets suivants :

1. Prédiction du type de besoin :
 - (a) Au niveau de l'estimation du type de besoin selon la morphologie de la requête, le type de requête par défaut est informationnel. Ceci réduit la précision de la prédiction du type informationnel de besoin. Notre perspective à court terme consiste à définir des caractéristiques appropriées au type informationnel des requêtes.
 - (b) Concernant l'estimation du type de besoin selon le contexte de la session, le calcul de la probabilité de basculement est basé sur l'utilisation d'une longueur moyenne statique des sessions. Or la longueur des sessions diffère entre les utilisateurs et même pour un même utilisateur entre ses différentes sessions. Notre deuxième perspective à court terme consiste à dégager une longueur moyenne adaptative des sessions pour chaque type de besoin. Ceci permet de considérer une longueur moyenne propre à l'historique de recherche de chaque utilisateur.
 - (c) L'évaluation de la prédiction du type de besoin a été faite en utilisant des collections TREC et selon un scénario de simulation des sessions de recherche. Dans le but de valider l'efficacité de notre approche dans un cadre de recherche réel, notre perspective à long terme, consiste à évaluer cette étape en utilisant des données issus d'un fichier log d'un moteur de recherche.

2. Modélisation du profil utilisateur :
 - (a) Le mécanisme de délimitation des sessions de recherche est basé sur l'utilisation d'un seuil prédéterminé statique issu du calcul des corrélations requête-profil. La stabilité de ce seuil au cours des sessions de recherche de l'utilisateur peut réduire la précision de groupage des requêtes corrélées dans une même session. Notre perspective à court terme dans ce sens porte sur la définition d'un seuil adaptatif de délimitation des sessions de recherche qui soit spécifique à chaque utilisateur et qui s'adapte au cours de ses sessions de recherche selon la différence de généralité/spécificité des différents centres d'intérêts de l'utilisateur.
 - (b) quant à l'évolution du profil à long terme, notre deuxième perspective à court terme consiste à mettre en place un mécanisme qui permet de juger la persistance et la fraîcheur des centres d'intérêts de l'utilisateur au cours du temps.
 - (c) à long terme, nous envisageons à définir un profil comportant plusieurs dimensions du contexte dans un système unifié d'accès personnalisé à l'information. Nous citons le contexte spatio-temporel définissant une situation de recherche qu'on l'associe à un contexte cognitif défini par des centres d'intérêts de l'utilisateur et des préférences de recherche. L'union de ces dimensions définit ce qu'on appelle le profil situationnel. En outre, le contexte social défini par l'appartenance de l'utilisateur à un réseau d'utilisateurs permet d'inférer la pertinence des résultats de celle du groupe.
3. Evaluation du modèle d'accès personnalisé à l'information :
 - (a) Lors de l'évaluation de l'efficacité de personnalisation à long terme, nous avons adopté le principe de la validation croisée qui permet d'utiliser des requêtes de test appartenant à des domaines ayant servi à la création des profils à long terme et ignorons le seuil de distance sémantique qui permet de juger si un nouveau centre d'intérêt est à apprendre. Notre troisième perspective dans ce sens est à court terme et consiste à apprendre le seuil de la distance sémantique entre graphes de profils et évaluer son impact sur la précision de la mesure dans un contexte de basculement des centres d'intérêts vers un nouveau besoin en information non préalablement appris à long terme.
 - (b) à court terme, notre perspective porte sur l'évaluation de l'efficacité de recherche personnalisée dans un contexte de recherche réel en utilisant des données des vrais utilisateurs issus des fichiers *logs* d'un moteur de recherche. Ceci permet d'évaluer le mécanisme de délimitation des sessions combiné avec des approches basé-temps d'une part et de construire des profils utilisateurs à partir de leurs vrais interactions d'autre part, puis d'en évaluer l'efficacité de personnalisation par utilisation des profils des utilisateurs réels.

- (c) à long terme, nous envisageons à mettre en place un cadre d'évaluation standard pour la RI personnalisée où l'évaluation est vigoureusement contrôlée en utilisant une collection de test réelle dont les requêtes sont émises par des vrais utilisateurs et leurs interactions sont exploitables pour apprendre des profils réels.

Bibliographie personnelle

- [1] L. Tamine, M. Daoud, B.-D. Dinh, and M. Boughanem. Contextual query classification in web search. In Baumeister and Atzmüller, editors, *International Workshop on Information Retrieval Learning, Knowledge and Adaptability (LWA), Hildesheim (Germany), 09/10/2008-11/10/2008*, pages 65–68, <http://www.uni-wuerzburg.de>, octobre 2008. Université Würzburg.
- [2] M. Daoud, L. Tamine, B.-D. Dinh, and M. Boughanem. Contextual Query Classification For Personalizing Informational Search. In *Web Information Systems Engineering, kerkennah Island, Sfax, Tunisia, 12/06/09-14/06/09*, <http://www.editions-hermes.fr/>, June 2009. Hermès.
- [3] M. Daoud, L. Tamine, and M. Boughanem. Detecting session boundaries for personalizing search using a conceptual user profile. In Springer, editor, *Advances in Electrical Engineering and Computational Science*, volume 39, pages 471–482. Springer, <http://www.springerlink.com>, 2008.
- [4] M. Daoud, L. Tamine, and M. Boughanem. Towards a graph based user profile modeling for a session-based personalized search. *Knowledge and Information Systems*, 2009.
- [5] M. Daoud, L. Tamine, M. Boughanem, and B. Chebaro. Learning Implicit User Interests Using Ontology and Search History for Personalization. In *International Web Information Systems Engineering - International Workshop on Personalized Access to Web Information (WISE-PAWI), Nancy, 03/12/07-07/12/07*, pages 325–336, <http://www.springerlink.com>, décembre 2007. Springer.
- [6] M. Daoud, L. Tamine, and M. Boughanem. Learning user interests for session-based personalized search. In Borlund, Schneider, Lalmas, and Tombros, editors, *ACM Information Interaction in context (IIx), London, 14/10/08-17/10/08*, pages 57–64, <http://www.acm.org/>, octobre 2008. ACM.
- [7] M. Daoud, L. Tamine, and M. Boughanem. Using a concept-based user context for search personalization. In *International Conference of Data Mining and Knowledge Engineering (ICDMKE), London, 02/07/08-*

- 04/07/08, pages 293–298, <http://www.iaeng.org>, juillet 2008. International Association of Engineers (IAENG).
- [8] M. Daoud, L. Tamine, M. Boughanem, and B. Chebaro. Construction des profils utilisateurs à base d’une ontologie pour une recherche d’information personnalisée. In *Conférence francophone en Recherche d’Information et Applications (CORIA), Trégastel, 12/03/2008-14/03/2008*, pages 225–240, <http://www.univ-rennes1.fr>, mars 2008. Université de Rennes 1.
- [9] M. Daoud, L. Tamine, M. Boughanem, and B. Chebaro. A Session Based Personalized Search Using An Ontological User Profile. In *ACM Symposium on Applied Computing (SAC 2009), Hawaii (USA), 08/03/09-12/03/09*, pages 1031–1035, <http://www.acm.org/>, mars 2009. ACM.
- [10] M. Daoud, L. Tamine, and M. Boughanem. Using a graph-based ontological user profile for personalizing search (Poster). In *Conference on Information and Knowledge Management (CIKM), Napa Valley, 26/10/08-28/10/08*, pages 1495–1496, <http://portal.acm.org/dl.cfm>, octobre 2008. ACM DL.
- [11] M. Daoud and L. Tamine. Proposition des cadres d’évaluation d’un système de RI personnalisé. In *Conférence francophone en Recherche d’Information et Applications (CORIA), Belambra de la Presqu’île de Giens, France, 05/05/09-07/05/09*, may 2009.
- [12] L. Tamine, M. Boughanem, and M. Daoud. Evaluation of contextual information retrieval : overview of issues and research. *Knowl Inf Syst (Kais)*, in press, 2009.
- [13] M. Daoud. Recherche Contextuelle d’Information. In *Conférence francophone en Recherche d’Information et Applications (CORIA), Saint-Etienne, 28/03/07-30/03/07*, pages 467–472. Université de Saint-Etienne, mars 2007.

Bibliographie

- [14] E. Agichtein, E. Brill, and S. T. Dumais. Improving web search ranking by incorporating user behavior information. In *Proceedings of the 29th International SIGIR conference on research and development in information retrieval*, pages 19–26, 2006.
- [15] C. P. Alexandru, N. Wolfgang, P. Raluca, and K. Christian. Using odp metadata to personalize search. In *SIGIR '05 : Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 178–185, 2005.
- [16] J. Allan and al. Challenges in information retrieval and language modeling : report of a workshop held at the center for intelligent information retrieval, university of massachusetts amherst, september 2002. *SIGIR Forum*, 37(1) :31–47, 2003.
- [17] S. arvi, J. Puolam, and a Kaski. Relevance feedback from eye movements for proactive information retrieval, 2004.
- [18] R. Baeza-Yates and R. A. Ribeiro-Neto. *Modern Information Retrieval*. New York : ACM Press ; Harlow England : Addison-Wesley, cop., 1999.
- [19] J. Bai, J.-Y. Nie, G. Cao, and H. Bouchard. Using query contexts in information retrieval. In *SIGIR '07 : Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 15–22, New York, NY, USA, 2007. ACM.
- [20] M. Balabanovic and Y. Shoham. Fab : Content-based collaborative recommendation. *Communications of the ACM*, 3(40) :66–72, March 1997.
- [21] C. L. Barry. User-defined relevance criteria : an exploratory study. *Journal of the American Society for Information Science*, 45 :149–159, 1994.
- [22] I. M. Begg, J. Gnocato, and W. E. Moore. A prototype intelligent user interface for real-time supervisory control systems. In *IUI '93 : Proceedings of the 1st international conference on Intelligent user interfaces*, pages 211–214, New York, NY, USA, 1993. ACM.
- [23] N. J. Belkin and W. B. Croft. Information filtering and information retrieval : two sides of the same coin? *Communications of the ACM*, 35(12) :29–38, 1992.
- [24] H. Beorko and M. Bernick. Automatic document classification. *Journal of the Association for Computing Machinery*, pages 151–161, 1963.

-
- [25] K. Bharat, T. Kamba, and M. Albers. Personalized, interactive news on the web. *Multimedia Syst.*, 6(5) :349–358, 1998.
- [26] D. Bilal. Children’s use of the yahooligans! web search engine : cognitive, physical, and affective behaviors on fact-based search tasks. *Journal of American Society in Information Science*, 51(7) :646–665, 2000.
- [27] P. Borlund. *Evaluation of interactive information retrieval systems*. Ph.d. thesis, Abo Akademi University, Finland, 2000.
- [28] P. Borlund and P. Ingwersen. Measures of relative relevance and ranked half-life : performance indicators in interactive ir. In *Croft W.B et al Eds*, pages 324–331. Proceedings of the 21st ACM SIGIR International Conference on Research and Development, Août 1998.
- [29] M. Boughanem, K. Sauvagnat, and C. Laffaire. Mercure at trec 2003 web track - topic distillation task. In *TREC 2003 : The Twelfth Text Retrieval Conference*, pages 343–348, 2003.
- [30] M. Boughanem and J. Savoy, editors. *Recherche d’information états des lieux et perspectives*. Hermès Science Publications, <http://www.editions-hermes.fr/>, 2008.
- [31] S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, 30(1-7) :107–117, 1998.
- [32] A. Broder. A taxonomy of web search. *SIGIR Forum*, 36(2) :3–10, 2002.
- [33] B. Brookes. The developping cognitive viewpoint in information science. In *De Mey, M. and al Editors*, pages 195–203, septembre 1977.
- [34] C. Buckley and E. M. Voorhees. Evaluating evaluation measure stability. In *SIGIR ’00 : Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval*, pages 33–40, New York, NY, USA, 2000. ACM.
- [35] C. Buckley and E. M. Voorhees. Retrieval evaluation with incomplete information. In *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 25–32, New York, 2004. ACM Press.
- [36] J. Budzik and K. Hammond. User interactions with every applications as context for just-in-time information access. In *Proceedings of the 5th international conference on intelligent user interfaces*, pages 44–51, Mars 2000.
- [37] H. Bunke. Recent developments in graph matching. *Pattern Recognition, International Conference on*, 2 :2117, 2000.
- [38] H. Bunke, X. Jiang, and A. Kandel. On the minimum common supergraph of two graphs. *Computing*, 65(1) :13–25, 2000.
- [39] K. Bystrom and K. Jarvelin. Task complexity affects information seeking and use. *Information Processing and Management*, 31(2) :191–213, 1995.

- [40] R. V. K. carlos A. Cuadra. Opening the black box of relevance. *Journal of Documentation*, 23(4) :291–303, 1967.
- [41] L. Chen and K. Sycara. Webmate : A personal agent for browsing and searching. In *Proceedings of the 2nd international conference on autonomous agents and multi agent systems, Minneapolis*, pages 10–13, 1998.
- [42] M.-S. Chen, J. S. Park, and P. S. Yu. Efficient data mining for path traversal patterns. *IEEE Trans. on Knowl. and Data Eng.*, 10(2) :209–221, 1998.
- [43] D. N. Chin. Empirical evaluation of user models and user-adapted systems. *User Model User-Adapted interact*, 11(1-2) :181–194, 2001.
- [44] P. A. Chirita, D. Olmedilla, and W. Nejdl. Pros : a personalized ranking platform for web search. In *Proceedings of the 3rd International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems*, pages 34–43, Eindhoven, Netherlands, August 2004.
- [45] C. Cleverdon. The cranfield test on index language devices. In *Aslib*, pages 173–194, 1967.
- [46] C. W. Cleverdon. The significance of the cranfield tests on index languages. In *SIGIR '91 : Proceedings of the 14th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 3–12, New York, NY, USA, 1991. ACM.
- [47] C. Cool. The concept of situation in information science. *Annual review of information science and technology*, 35 :5–42, 2001.
- [48] C. Cool and A. Spink. Issues of context in information retrieval : an introduction to the special issue. *Journal of Information Processing and Management*, 38(55) :605–611, 2002.
- [49] R. Cooley, B. Mobasher, and J. Srivastava. Data preparation for mining world wide web browsing patterns. *Knowledge and Information Systems*, 1 :5–32, 1999.
- [50] F. Crestani and C. J. Van Rijsbergen. A model for adaptive information retrieval. *Journal of Intelligent Information Systems*, 8(1) :29–56, 1997.
- [51] H. Cui, J.-R. Wen, J.-Y. Nie, and W.-Y. Ma. Query expansion by mining user logs. *IEEE Trans. Knowl. Data Eng.*, 15(4) :829–839, 2003.
- [52] W. Daelemans, J. Zavrel, K. Slot, and A. Bosch. Timbl : Tilburg memory based learner, version 3.0, reference guide. In *ILK Technical report*, 2000.
- [53] J. Dean and M. R. Henzinger. Finding related pages in the world wide web. *Computer Networks (Amsterdam, Netherlands : 1999)*, 31(11–16) :1467–1479, 1999.
- [54] F. Diaz and R. Jones. Using temporal profiles of queries for precision prediction. In *SIGIR '04 : Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 18–24, 2004.

- [55] C. Ding and J. Patra. User modeling for personalized web search with self-organizing map. *Journal of American Society in Information Science and Technology*, 58(4) :494–507, 2007.
- [56] C. Ding, J. C. Patra, and F. C. Peng. Personalized web search with self-organizing map. In *EEE '05 : Proceedings of the 2005 IEEE International Conference on e-Technology, e-Commerce and e-Service (EEE'05) on e-Technology, e-Commerce and e-Service*, pages 144–147, Washington, DC, USA, 2005. IEEE Computer Society.
- [57] S. Dumais, E. C. J. J. Cadiz, G. Jancke, R. Sarin, and C. R. Daniel. Stuff i've seen : a system for personal information retrieval and re-use. In *SIGIR'03 : Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval (SIGIR '03)*, pages 72–79. ACM Press, 2003.
- [58] E. N. Efthimiadis. Query expansion. *Annual Review of Information Systems and Technology*, 31 :121–187, 1996.
- [59] M. Eirinaki and M. Vazirgiannis. Web mining for web personalization. *ACM Trans. Interet Technol.*, 3(1) :1–27, 2003.
- [60] Y. El-Sonbaty and M. A. Ismail. A new error-correcting distance for attributed relational graph problems. In *Proceedings of the Joint IAPR International Workshops on Advances in Pattern Recognition*, pages 266–276, London, UK, 2000. Springer-Verlag.
- [61] Q. Feng and C. Junghoo. Automatic identification of user interest for personalized search. In *WWW '06 : Proceedings of the 15th international conference on World Wide Web*, pages 727–736, 2006.
- [62] M.-L. Fernández and G. Valiente. A graph distance metric combining maximum common subgraph and minimum common supergraph. *Pattern Recogn. Lett.*, 22(6-7) :753–758, 2001.
- [63] A. Foss, W. Wang, and O. R. Zaane. A non-parametric approach to web log analysis. In *Workshop on WebMining in First International SIAM Conference on Data Mining*, pages 41–50, April 2001.
- [64] S. Fox, K. Karnawat, M. Mydland, S. Dumais, and T. White. Evaluating implicit measures to improve web search. *ACM Trans. Inf. Syst.*, 23(2) :147–168, 2005.
- [65] E. Frias-Martinez, S. Y. Chen, R. D. Macredie, and X. Liu. The role of human factors in stereotyping behavior and perception of digital library users : a robust clustering approach. *User Modeling and User-Adapted Interaction*, 17(3) :1573–1391, 2007.
- [66] N. Fuhr. Information retrieval : introduction and survey. post-graduate course on information retrieval, university of Duisburg-Essen, Germany, 2000.
- [67] S. Gauch, J. Chaffee, and A. Pretschner. Ontology-based personalized search and browsing. *Web Intelli. and Agent Sys.*, 1(3-4) :219–234, 2003.

- [68] G. Gentili, A. Micarelli, and F. Sciarrone. Infoweb : An adaptive information filtering system for the cultural heritage domain. *Applied Artificial Intelligence*, 17(8-9) :715–744, 2003.
- [69] J. Gowan. *A multiple model approach to personalised information access*. Master thesis in computer science, Faculty of science, Université de College Dublin, February 2003.
- [70] L. Gravano, V. Hatzivassiloglou, and R. Lichtenstein. Categorizing web queries according to geographical locality. In *CIKM '03 : Proceedings of the twelfth international conference on Information and knowledge management*, pages 325–333, 2003.
- [71] A. Gulli and A. Signorini. The indexable web is more than 11.5 billion pages. In *WWW '05 : Special interest tracks and posters of the 14th international conference on World Wide Web*, pages 902–903, New York, NY, USA, 2005. ACM.
- [72] A. Gupta and N. Nishimura. Finding largest subtrees and smallest supertrees. *Algorithmica*, 21 :183–210, 1998.
- [73] D. Gusfield. *Algorithms on strings, trees, and sequences : computer science and computational biology*. Cambridge University Press, New York, NY, USA, 1997.
- [74] R. Harrathi and S. Calabretto. Un modèle de qualité de l'information. In G. Ritschard and C. Djeraba, editors, *EGC*, volume RNTI-E-6 of *Revue des Nouvelles Technologies de l'Information*, pages 299–304. Cépaduès-Éditions, 2006.
- [75] T. H. Haveliwala. Topic-sensitive pagerank. In *Proceedings of the Eleventh International World Wide Web Conference (WWW 2002)*, Honolulu, Hawaii, May 2002.
- [76] T. H. Haveliwala, A. Gionis, D. Klein, and P. Indyk. Evaluating strategies for similarity search on the web. In *WWW'02 : Proc. of the Eleventh International World Wide Web Conference*, pages 432–442, 2002.
- [77] D. He. Detecting session boundaries from web user logs. In *In Proceedings of of the BCS-IRSG 22nd Annual Colloquium on Information Retrieval Research*, pages 57–66, 2000.
- [78] T. Hofmann. Probabilistic latent semantic analysis. In *Proceedings of Uncertainty in Artificial Intelligence, UAI'99*, Stockholm, 1999.
- [79] I. Hsieh-Yee. Effects of the searcher and experience and subject knowledge on the search tactics of novice and experienced searchers. *Journal of the American Society for Information Science*, 44(3) :161–174, 1993.
- [80] X. Huang, F. Peng, A. An, and D. Schuurmans. Dynamic web log session identification with statistical language models. *J. Am. Soc. Inf. Sci. Technol.*, 55(14) :1290–1303, 2004.
- [81] M. E. Hupfer and B. Detlor. Gender and web information seeking : A self-concept orientation model. *Journal of the American Society for Information Science and Technology*, 57(8) :1105 – 1115, 2006.

- [82] P. Ingwersen. Polyrepresentation of information needs and semantic entities : elements of a cognitive theory for information retrieval interaction. In *Proceedings of the 17th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 101–110, New York, NY, USA, 1994. Springer-Verlag New York, Inc.
- [83] P. Ingwersen. Cognitive perspectives of information interactions : Elements of a cognitive ir theory. *Annual review of information science and technology*, 52(1) :3–50, 1996.
- [84] P. Ingwersen. Cognitive perspectives of information interactions : Elements of a cognitive ir theory. *Annual review of information science and technology*, 52(1) :3–50, 1996.
- [85] P. Ingwersen and K. Järvelin. *The TURN : Integration of Information Seeking and Retrieval in Context*. SPRINGER, August 2005.
- [86] Y. R. Jacques Savoy. Hyperliens et recherche d'information sur le web. In *JADT'04 : proceedings of JADT 2004*, pages 325–333, 2003.
- [87] B. J. Jansen, D. L. Booth, and A. Spink. Determining the user intent of web search engine queries. In *WWW '07 : Proceedings of the 16th international conference on World Wide Web*, pages 1149–1150, 2007.
- [88] B. J. Jansen, A. Spink, J. Bateman, and T. Saracevic. Real life information retrieval : a study of user queries on the web. *SIGIR Forum*, 32(1) :5–17, 1998.
- [89] B. J. Jansen, A. Spink, and V. Kathuria. How to define searching sessions on web search engines. In *Advances in Web Mining and Web Usage Analysis, 8th International Workshop on Knowledge Discovery on the Web, WebKDD 2006, Philadelphia, PA, USA*, pages 92–109, 2006.
- [90] K. Jarvelin and J. Kekalainen. Ir evaluation methods for highly relevant documents. In *Proceedings of the 23rd annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 41–48. Belkin and al, 2000.
- [91] K. Jarvelin and J. Kekalainen. Cumulative gain-based evaluation of ir techniques. *ACM Transactions on Information Systems (ACM TOIS)*, 20(4) :422–446, 2002.
- [92] T. Joachims. Optimizing search engines using clickthrough data. In *KDD '02 : Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 133–142, New York, NY, USA, 2002. ACM.
- [93] K. S. Jones. *Information Retrieval Experiment*. Butterworth-Heinemann, Newton, MA, USA, 1981.
- [94] K. Järvelin. On information, information technology and the development of society : An information science perspective. In *Ingwersen, P., Kajberg, L. and Pejtersen, A.M Editors*, pages 35–55, 1986.

- [95] I.-H. Kang. Transactional query identification in web search. In *AIRS'05 : Proceedings Information Retrieval Technology, Second Asia Information Retrieval Symposium, Jeju Island, Korea*, pages 221–232, 2005.
- [96] I.-H. Kang and G. Kim. Query type classification for web document retrieval. In *SIGIR '03 : Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*, pages 64–71, 2003.
- [97] G. Kazai and N. Milic-Frayling. Trust, authority and popularity in social information retrieval. In *CIKM '08 : Proceeding of the 17th ACM conference on Information and knowledge management*, pages 1503–1504, New York, NY, USA, 2008. ACM.
- [98] G. Kazai and N. Milic-Frayling. Effects of social approval votes on search performance. *Information Technology : New Generations, Third International Conference on*, 0 :1554–1559, 2009.
- [99] J. Kekalainen and K. Jarvelin. Evaluating information retrieval systems under the challenges of interaction and multidimensional dynamic relevance. In *Proceedings of the 4th CoLIS conference*, pages 253–270. P. Ingwersen and P. Vakkari, 2004.
- [100] D. Kelly and N. J. Belkin. Reading time, scrolling and interaction : Exploring implicit sources of user preferences for relevance feedback during interactive information retrieval. In *Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 408–409, New Orleans, Louisiana, United States, 2001.
- [101] D. Kelly and J. Teevan. Implicit feedback for inferring user preference : A bibliography. *SIGIR Forum*, 32(2), 2003.
- [102] H. R. Kim and P. K. Chan. Learning implicit user interest hierarchy for context in personalization. In *IUI '03 : Proceedings of the 8th international conference on Intelligent user interfaces*, pages 101–108, New York, NY, USA, 2003. ACM.
- [103] K. Kim. Effects of emotion control and task on web searching behavior. *Information Processing and Management*, 44(1) :373–385, 2008.
- [104] J. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 5(46) :604–632, 1999.
- [105] A. Kobsa. Generic user modeling systems. *User Modeling and User-Adapted Interaction*, 11(1-2) :49–63, 2001.
- [106] A. Kobsa and W. Wahlster. *User Models in Dialog Systems*. Springer-Verlag, 1989.
- [107] J. Konstan, M. Miller, B. ans Maltz, J. Herlocker, L. Gordon, and J. Riedl. GroupLens : Applying collaborative filtering to usenet news. In *Communications of the ACM*, 40(3), pages 77–87, 1997.

- [108] D. Kostadinov. La personnalisation de l'information, définition de modèle de profil utilisateur. rapport de dea. Master's thesis, Université de Versailles, France, 2003.
- [109] G. Koutrika and Y. Ioannidis. A unified user profile framework for query disambiguation and personalization. In *Proceedings of Workshop on New Technologies for Personalized Information Access*, July 2005.
- [110] R. Kraft, F. Maghoul, and C. Chang. Y!q : contextual search at the point of inspiration. In *CIKM '05 : Proceedings of the 14th ACM international conference on Information and knowledge management*, pages 816–823, New York, NY, USA, October 2005. ACM Press.
- [111] M. Lalmas, A. MacFarlane, S. M. Rüger, A. Tombros, T. Tsirikika, and A. Yavlinsky, editors. *Advances in Information Retrieval, 28th European Conference on IR Research, ECIR 2006, London, UK, April 10-12, 2006, Proceedings*, volume 3936 of *Lecture Notes in Computer Science*. Springer, 2006.
- [112] K. Lang. NewsWeeder : learning to filter netnews. In *Proceedings of the 12th International Conference on Machine Learning*, pages 331–339. Morgan Kaufmann publishers Inc. : San Mateo, CA, USA, 1995.
- [113] E. Law, T. Klobucar, and M. Pipan. User effect in evaluating personalized information retrieval systems. In *EC-TEL*, pages 257–271. Springer-Verlag, 2006.
- [114] U. Lee, Z. Liu, and J. Cho. Automatic identification of user goals in web search. In *WWW '05 : Proceedings of the 14th international conference on World Wide Web*, pages 391–400, 2005.
- [115] H. J.-M. Leplat J. Tâche et activité dans l'analyse psychologique des situations. *Cahiers de Psychologie Cognitive*, 3(1) :49–63, 1983.
- [116] C. W. Leung, S. C. Chan, and F. Chung. A collaborative filtering framework based on fuzzy association rules and multiple-level similarity. *Knowl. Inf. Syst.*, 10(3) :357–381, 2006.
- [117] G. Levi. A note on the derivation of maximal common subgraphs of two. directed or undirected graphs. *Calcolo*, 9 :341–354, 1972.
- [118] X. Li and W. B. Croft. Time-based language models. In *CIKM '03 : Proceedings of the twelfth international conference on Information and knowledge management*, pages 469–475, 2003.
- [119] Y. Li and N. J. Belkin. A faceted approach to conceptualizing tasks in information seeking. *Inf. Process. Manage.*, 44(6) :1822–1837, 2008.
- [120] Y. Li, R. Krishnamurthy, S. Vaithyanathan, and H. V. Jagadish. Getting work done on the web : supporting transactional queries. In *SIGIR '06 : Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 557–564, 2006.

- [121] H. Lieberman. Letizia : an agent that assists web browsing. In *IJCAI 95 : Proceedings of International Joint Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence.*, pages 924–929, 1995.
- [122] H. Lieberman. Autonomous interface agents. In *CHI*, pages 67–74, 1997.
- [123] F. Liu, C. Yu, and W. Meng. Personalized web search for improving retrieval effectiveness. *IEEE Transactions on Knowledge and Data Engineering*, 16(1) :28–40, 2004.
- [124] Z. Ma, G. Pant, and O. R. L. Sheng. Interest-based personalized search. *ACM Trans. Inf. Syst.*, 25(1) :5, 2007.
- [125] A. G. Maguitman, F. Menczer, H. Roinestad, and A. Vespignani. Algorithmic detection of semantic similarity. In *WWW '05 : Proceedings of the 14th international conference on World Wide Web*, pages 107–116, New York, NY, USA, 2005. ACM.
- [126] C. D. Manning, P. Raghavan, and H. Schütze. *Introduction to Information Retrieval*. Cambridge University Press, July 2008.
- [127] M. E. Maron and J. L. Kuhns. On relevance, probabilistic indexing and information retrieval. *J. ACM*, 7(3) :216–244, 1960.
- [128] J. J. Mcgregor. Backtrack Search Algorithms and the Maximal Common Subgraph Problem. *SOFTWARE-PRACT. AND EXPER.*, 12(1) :23–34, 1982.
- [129] A. Micarelli and F. Sciarrone. Anatomy and empirical evaluation of an adaptive web-based information filtering system. *User Modeling and User-Adapted Interaction*, 14(2-3) :159–200, 2004.
- [130] T. Mitchell. *Machine Learning*. McGraw-Hill Higher Education, 1997.
- [131] S. Mizzaro. How many relevances in information retrieval? *Interacting with Computers*, 10(3) :303–320, 1998.
- [132] D. Mladenic. Personal webwatcher : design and implementation. In *Technical Report IJS-DP-7472*. J. Stefan Institute, Department for Intelligent Systems, 1998.
- [133] J. Mostafa, S. Mukhopadhyay, and M. Palakal. Simulation studies of different dimensions of users' interests and their impact on user modeling and information filtering. *Information Retrieval*, 6(2) :199–223, 2003.
- [134] G. Muresan, C. L. Smith, M. Cole, L. Liu, and N. J. Belkin. Detecting document genre for personalization of information retrieval. In *HICSS '06 : Proceedings of the 39th Annual Hawaii International Conference on System Sciences*, page 50.3, Washington, DC, USA, 2006. IEEE Computer Society.
- [135] D. F. Nettleton, L. Calderón-Benavides, and R. A. Baeza-Yates. Analysis of web search engine query session and clicked documents. In *WEBKDD*, pages 207–226, 2006.

- [136] D. Nichols. Implicit rating and filtering. In *In Proc. 5th DELOS Workshop on Filtering and Collaborative Filtering*, pages 31–36, Budapest, Hungary, November 1997.
- [137] D. W. Oard and J. Kim. Modeling information content using observable behavior. In *In Proceedings of the 64th Annual Meeting of the American Society for Information Science and Technology*, pages 38–45, USA, 2001.
- [138] T. K. Park. Toward a theory of user-based relevance : a call for a new paradigm of inquiry. *J. Am. Soc. Inf. Sci.*, 45(3) :135–141, 1994.
- [139] M. J. Pazzani, J. Muramatsu, and D. Billsus. Syskill & webert : Identifying interesting web sites. In *Proceedings of the 30th National Conference on Artificial Intelligence*, pages 54–61, Portland, 1996.
- [140] B. Pia. The concept of relevance in ir. *J. Am. Soc. Inf. Sci. Technol.*, 54(10) :913–925, 2003.
- [141] V. V. Raghavan and H. Sever. On the reuse of past optimal queries. In *SIGIR '95 : Proceedings of the 18th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 344–350, New York, NY, USA, 1995. ACM.
- [142] I. A. Research, J. Zien, J. Meyer, and J. Tomlin. Web query characteristics and their implications on search engines jason zien, j org meyer, john tomlin. In *In Proceedings of the 10th International WWW Conference. Hong Kong*, 2001.
- [143] B. J. Rhodes and P. Maes. Just-in-time information retrieval agents. *IBM Syst. J.*, 39(3-4) :685–704, 2000.
- [144] E. Rich. User modeling via stereotypes. pages 329–342, 1998.
- [145] S. Robertson. The probability ranking principle in ir. In *Journal of documentation*, 33(4) :294–304, 1977.
- [146] S. Robertson and K. Sparck Jones. Relevance weighting for search terms. *Journal of The American Society for Information Science*, 27(3) :129–146, 1976.
- [147] J. J. Rocchio. Relevance feedback in information retrieval. In Salton [155].
- [148] H. Rode and D. Hiemstra. Using query profiles for clarification. In Lalmas et al. [111], pages 205–216.
- [149] J. rong Wen, N. Lao, and W.-Y. Ma. Probabilistic model for contextual retrieval. In *In : Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2004*, pages 57–63. ACM Press, 2004.
- [150] D. E. Rose and D. Levinson. Understanding user goals in web search. In *WWW '04 : Proceedings of the 13th international conference on World Wide Web*, pages 13–19, New York, NY, USA, 2004. ACM.

- [151] I. Ruthven and M. Lalmas. A survey on the use of relevance feedback for information access systems. *Knowledge Engineering review*, 18(2) :95–145, 2003.
- [152] W. W. Ryen, I. Ruthven, J. M. Jose, and C. J. Van Rijsbergen. Evaluating implicit feedback models using searcher simulations. *ACM Transactions on Information Systems*, 23(3) :325–361, 2005.
- [153] H. Sakagami and T. Kamba. Learning personal preferences on online newspaper articles from user behaviors. In *Proceedings of the 6th International WWW Conference*, Santa Clara, California, April, 7-11 1997.
- [154] G. Salton. A comparison between manual and automatic indexing methods. *Journal of American Documentation*, 20(1) :61–71, 1971.
- [155] G. Salton, editor. *The SMART Retrieval System - Experiments in Automatic Document Processing*. Prentice Hall, Englewood, Cliffs, New Jersey, 1971.
- [156] G. Salton and M. McGill. *The concept of "relevance" in information science : A historical review*. R.R. Bowker, New York, 1970.
- [157] G. Salton and M. McGill. *Introduction to Modern Information Retrieval*. McGraw-Hill, New York, 1983.
- [158] G. Salton and C. Yang. On the specification of term values in automatic indexing. *Journal of documentation*, pages 351–372, 1973.
- [159] T. Saracevic. The stratified 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.
- [160] Y. Shang and L. Li. Precision evaluation of search engines. *World Wide Web*, 5(2) :159–173, 2002.
- [161] D. Shen, Z. Chen, Q. Yang, H.-J. Zeng, B. Zhang, Y. Lu, and W.-Y. Ma. Web-page classification through summarization. In *SIGIR '04 : Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 242–249, New York, NY, USA, 2004. ACM.
- [162] X. Shen, B. Tan, and C. Zhai. Context-sensitive information retrieval using implicit feedback. In *SIGIR '05 : Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 43–50, New York, NY, USA, 2005. ACM.
- [163] X. Shen, B. Tan, and C. Zhai. Implicit user modeling for personalized search. In *CIKM '05 : Proceedings of the 14th ACM international conference on Information and knowledge management*, pages 824–831, New York, NY, USA, 2005. ACM.
- [164] A. Sieg, B. Mobasher, and R. Burke. Web search personalization with ontological user profiles. In *CIKM'07 : Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*, pages 525–534, New York, NY, USA, 2007. ACM.

- [165] A. Sieg, B. Mobasher, R. Burke, G. Prabu, and S. Lytinen. Using concept hierarchies to enhance user queries in web-based information retrieval. In *The IASTED International Conference on Artificial Intelligence and Applications. Innsbruck, Austria, 2004*.
- [166] M. Smith, V. Barash, L. Getoor, and H. W. Lauw. Leveraging social context for searching social media. In *SSM '08 : Proceeding of the 2008 ACM workshop on Search in social media*, pages 91–94, New York, NY, USA, 2008. ACM.
- [167] B. Smyth and E. Balfe. Anonymous personalization in collaborative web search. *Information retrieval*, 9(2) :165–190, 2006.
- [168] P. Solomon. Children’s information retrieval behavior : a case analysis of opac. *Journal of the American Society for Information Science*, 44(5) :245, 1988.
- [169] D. Sonnenwald and A. Pejtersen. Towards a framework to support information needs in design : A concurrent engineering example. *Knowledge organization and mangement*, pages 161–172, 1994.
- [170] E. Sormunen. Liberal relevance criteria of trec - : counting on negligible documents? In *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 324–330, New York, NY, USA, 2002. ACM.
- [171] K. Sparck-Jones. *The Cranfield tests*. Butterworth-Heinemann, 1981.
- [172] K. Sparck Jones and P. Willett, editors. *Readings in information retrieval*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1997.
- [173] M. Speretta and S. Gauch. Personalized search based on user search histories. In *WI '05 : Proceedings of the 2005 IEEE/WIC/ACM International Conference on Web Intelligence*, pages 622–628, Washington, DC, USA, 2005. IEEE Computer Society.
- [174] S. Sriram, X. Shen, and C. Zhai. A session-based search engine. In *SIGIR'04 : Proceedings of the International ACM SIGIR Conference*, 2004.
- [175] L. Tamine. *Optimisation de requêtes dans un système de recherche d'information : approche basée sur l'exploitation de techniques avancées de l'algorithmique génétique*. Thèse de doctorat en informatique, Université Paul Sabatier, decembre 2000.
- [176] L. Tamine. *De la recherche d'information orientée système à la recherche d'information orientée contexte : Verrous, contributions et perspectives*. Habilitation à diriger des recherches, Université Paul Sabatier, Toulouse, France, novembre 2008.
- [177] L. Tamine, M. Boughanem, and M. Daoud. Evaluation of contextual information retrieval : overview of issues and research. *Knowl Inf Syst (Kais)*, in press, 2009.

- [178] L. Tamine and S. Calabretto. Recherche d'information contextuelle et web. In M. Boughanem and J. Savoy, editors, *Recherche d'information : état des lieux et perspectives*, volume 1, chapter 7, pages 201–224. Hermès, <http://www.editions-hermes.fr/>, avril 2008.
- [179] L. Tamine, W. N. Zemirli, and W. Bahsoun. Approche statistique pour la définition du profil d'un utilisateur de système de recherche d'information. *Information - Interaction - Intelligence*, 7(1) : (en ligne), 2007.
- [180] L. Tamine-Lechani, M. Boughanem, and N. Zemirli. Personalized document ranking : exploiting evidence from multiple user interests for profiling and retrieval. In *Journal of Digital Information Management.*, 2008.
- [181] B. Tan, X. Shen, and C. Zhai. Mining long-term search history to improve search accuracy. In *KDD '06 : Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 718–723, New York, NY, USA, 2006. ACM.
- [182] F. Tanudjaja and L. Mui. Persona : A contextualized and personalized web search. In *Proc 35th Hawaii International Conference on System Sciences*, page 53, Big Island, Hawaii, January 2002.
- [183] J. tao Sun, H.-J. Zeng, H. Liu, and Y. Lu. Cubesvd : A novel approach to personalized web search. In *In Proc. of the 14 th International World Wide Web Conference (WWW)*, pages 382–390. Press, 2005.
- [184] J. Teevan, S. T. Dumais, and E. Horvitz. Personalizing search via automated analysis of interests and activities. In *SIGIR '05 : Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 449–456, New York, NY, USA, 2005. ACM.
- [185] C. G. Thomas and G. Fischer. Using agents to personalize the web. In *IUI '97 : Proceedings of the 2nd international conference on Intelligent user interfaces*, pages 53–60, New York, NY, USA, 1997. ACM.
- [186] A. Tombros, I. Ruthven, and J. M. Jose. How users assess web pages for information seeking. *Journal of American Society of Information Science and Technology (JASIST)*, 56(4) :327–344, 2005.
- [187] J. Trajkova and S. Gauch. Improving ontology-based user profiles. In *Proceedings of RIAO 2004*, Vaucluse, France, 2004.
- [188] J. R. Ullmann. An algorithm for subgraph isomorphism. *J. ACM*, 23(1) :31–42, 1976.
- [189] P. Vakkari. A theory of the task-based information retrieval process : a summary and generalisation of a longitudinal study. *Journal of documentation*, 57(1) :44–60, 2001.
- [190] D. Vallet, M. Fernández, P. Castells, P. Mylonas, and Y. Avrithis. Personalized information retrieval in context. 21st National Conference on Artificial Intelligence - 3rd International Workshop on Modeling and Retrieval of Context, Boston, USA, 16-17 July 2006, 2006.

- [191] E. M. Voorhees. The philosophy of information retrieval evaluation. In *CLEF '01 : Revised Papers from the Second Workshop of the Cross-Language Evaluation Forum on Evaluation of Cross-Language Information Retrieval Systems*, pages 355–370, London, UK, 2002. Springer-Verlag.
- [192] A. Waern. User involvement in automatic filtering : An experimental study. *User Modeling and User-Adapted Interaction*, 14(2-3) :201–237, 2004.
- [193] W. Wang and O. R. Zaïane. Clustering web sessions by sequence alignment. In *In Proceedings of the 13th international workshop on database and expert systems applications (DEXA 2002). Aix-en-Provence*, pages 394–398. Springer-Verlag, 2002.
- [194] C. Watters and G. Amoudi. Geosearcher : location-based ranking of search engine results. *J. Am. Soc. Inf. Sci. Technol.*, 54(2) :140–151, 2003.
- [195] G. I. Webb, M. J. Pazzani, and D. Billsus. Machine learning for user modeling. *User Modeling and User-Adapted Interaction*, 11(1-2) :19–29, 2001.
- [196] J.-R. Wen, J.-Y. Nie, and H.-J. Zhang. Clustering user queries of a search engine. In *WWW '01 : Proceedings of the 10th international conference on World Wide Web*, pages 162–168, New York, NY, USA, 2001. ACM.
- [197] R. W. White and D. Kelly. A study on the effects of personalization and task information on implicit feedback performance. In *CIKM '06 : Proceedings of the 15th ACM international conference on Information and knowledge management*, pages 297–306, New York, NY, USA, 2006. ACM.
- [198] R. W. White, I. Ruthven, and J. M. Jose. A study of factors affecting the utility of implicit relevance feedback. In *Proceedings of the 28th annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 15–19. Marchionini, G. Moffat, A Tait, J Baeza-Yates, R Ziviani, N Eds, August 2003.
- [199] D. Widyanoro, J. Yin, M. El Nasr, L. Yang, A. Zacchi, and J. Yen. Alipes : A swift messenger in cyberspace. In *Proceedings of Spring Symposium Workshop on Intelligent Agents in Cyberspace*, pages 62–67, Stanford, March 22-24 1999.
- [200] H. I. Xie. Users' evaluation of digital libraries (dls) : Their uses, their criteria, and their assessment. *Information Processing and Management*, 44(3) :1346–1373, 2008.
- [201] S. A. Yahia, M. Benedikt, L. V. S. Lakshmanan, and J. Stoyanovich. Efficient network aware search in collaborative tagging sites. *Proc. VLDB Endow.*, 1(1) :710–721, 2008.

-
- [202] O. Zamir and O. Etzioni. Grouper : a dynamic clustering interface to Web search results. *Computer Networks : The International Journal of Computer and Telecommunications Networking*, 31(11) :1361–1374, 1999.
- [203] W. N. Zemirli. *Modèle d'accès personnalisé à l'information basé sur les diagrammes d'influence intégrant un profil multidimensionnel*. Thèse de doctorat, Université Paul Sabatier, Toulouse, France, juin 2008.