



THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par *l'Université Toulouse III - Paul Sabatier*
Discipline ou spécialité : *Informatique*

Présentée et soutenue par *Mme RESSAD-BOUIDGHAGHEN Ourdia*
Le *12 Decembre 2011*

Titre : *Accès contextuel à l'information dans un environnement mobile :
approche basée sur l'utilisation d'un profil situationnel de l'utilisateur et
d'un profil de localisation des requêtes.*

JURY

*Mme. M. FRANCINE-MOENS (Professeur), Mme. S. CALABRETTO (Professeur), M. W. Kraaij
(Professeur), Claude CHRISMENT (Professeur), M. Mohand BOUGHANEM (Professeur),
et Mme. Lynda TAMINE-LECHANI (HDR).*

Ecole doctorale : *Mathématiques Informatique Télécommunications (MITT)*

Unité de recherche : *IRIT-UMR 5505*

Directeur(s) de Thèse : *M. Mohand BOUGHANEM et Mme. Lynda TAMINE-LECHANI*

Rapporteurs : *Mme. M. FRANCINE-MOENS et Mme. S. CALABRETTO*

Résumé

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

Notre contribution porte sur la conception d'un système de RI contextuel dans un cadre mobile. Plus spécifiquement, notre contribution se décline en trois principaux points : la modélisation et construction de profil situationnel de l'utilisateur, la caractérisation de la sensibilité de la requête à la localisation de l'utilisateur, ainsi que la définition d'un cadre de combinaison de ces éléments contextuels pour calculer un score de pertinence multidimensionnelle des documents.

Nous nous sommes intéressés en premier lieu à exploiter le profil situationnel de l'utilisateur dans un processus d'accès personnalisé à l'information. Le profil situationnel est composé de centres d'intérêts de l'utilisateur appris pour chaque situation de recherche. Une situation de recherche est caractérisée par une représentation sémantique de la localisation et de temps de l'utilisateur lors de sa recherche. Les centres d'intérêts sont construits en exploitant les documents jugés pertinents par l'utilisateur et une ontologie générale. Nous avons proposé d'utiliser l'approche par raisonnement à partir de cas pour sélectionner le centre d'intérêt à exploiter pour la personnalisation sur la base de la comparaison de la similarité des situations de recherche. Le centre d'intérêt sélectionné est utilisé dans le ré-ordonnement des résultats de recherche des requêtes appartenant à une situation de recherche similaire.

Nous exploitons ensuite le contexte de la requête dans un mécanisme de prédiction de la sensibilité de la requête à la localisation de l'utilisateur. Notre approche de prédiction de la sensibilité de la requête à la localisation se base sur la construction d'un modèle de langue de localisation de la requête. Ce modèle nous a servi comme source d'évidence pour calculer des caractéristiques pour la classification des requêtes selon leur sensibilité à la localisation. Nous avons également intégré notre approche de détection de la sensibilité de la requête à la localisation dans un processus d'adaptation des résultats de recherche selon le type de la requête.

En vue d'intégrer ces deux types d'adaptation dans un SRI contextuel, nous nous sommes proposés d'appliquer un modèle d'agrégation prioritaire pour la combinaison de pertinence multidimensionnelle pour la RI mobile. Ce modèle de pertinence multidimensionnelle présente la particularité d'exploiter deux opérateurs d'agrégation prioritaire permettant d'adapter les résultats de recherche selon les préférences de l'utilisateur exprimées sur les critères de pertinence.

Vu qu'il n'existe pas de cadre d'évaluation standard d'accès contextuel à l'information, plus particulièrement adapté au contexte mobile, nous avons proposé des cadres d'évaluation orientés-contexte basés sur des approches par "simulation de contexte" et "par étude journalière". Nous avons exploité ces cadres d'évaluation pour valider notre contribution dans le domaine. En particulier, nous avons évalué expérimentalement notre approche de personnalisation en utilisant notre profil situationnel en comparaison à un SRI standard, et avons montré que notre approche est à l'origine d'un gain de performance significatif. Nous avons validé notre approche de détection de la sensibilité de la requête à la localisation de l'utilisateur sur une collection de requêtes annotées manuellement issue du *log* de recherche d'AOL, en testant plusieurs classificateurs du domaine et par comparaison à une approche de l'état de l'art, et nous avons montré son efficacité à améliorer la performance de la recherche par comparaison à un SRI standard. Nous avons également comparé notre cadre de combinaison de pertinence à une approche de combinaison linéaire standard et montré son efficacité.

Abstract

Contextual information retrieval aims at combining knowledge about the query context and the user context in the same framework in order to better meet the user information needs. We propose a contextual search approach integrating a query location intent prediction method and a situational user profile modelling approach in order to improve the retrieval effectiveness for mobile search.

We propose an approach to personalize search results for mobile users by exploiting both cognitive and spatio-temporal context of the user. We propose to model the user on three semantic dimensions : time, location and interests. A case based reasoning approach is adopted to select the appropriate user profile for re-ranking the search results.

In order to identify the user intent global, local explicit and local implicit, we exploit the top N search results returned by a general Web search engine to build a location query profile using language models. Two measures namely location Kullback-Leibler Divergence and Kurtosis defined on this profile, allow us to effectively classify the three types of queries.

We also propose a multidimensional ranking model based on the standard relevance dimension of topic and the contextual dimensions of interests and location to personalise search results for a mobile user. The peculiarity of our multidimensional ranking lies in a “prioritized combination” of the considered criteria, using the “prioritized scoring” and “prioritized and” operators, which allow flexible personalization of search results according to users’ preferences.

As there is no standard evaluation protocol for evaluating contextual access retrieval, we have proposed context-oriented evaluation protocols ranging from simulation frameworks to user studies. We have exploited these protocols to evaluate our contributions in the domain and have shown the effectiveness of our approaches.

Dédicaces

à mes parents
à mon Nour
et au futur arrivant ...

Remerciements

Je tiens à remercier très sincèrement monsieur le Professeur Claude Chrisment, responsable de l'équipe SIG, pour m'avoir accueillie au sein de son 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 la confiance qu'il m'a accordé en me proposant le sujet de ma thèse, et pour tous les moyens qu'il a mis à ma disposition pour mener à bien mes travaux. Ses précieuses remarques, son exigence et ses commentaires ont permis d'améliorer la qualité de mes travaux.

Je tiens à remercier vivement, peut être jamais assez, 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. Son expérience et ses compétences ont permis l'accomplissement de ce travail. Qu'elle trouve ici les marques de mes sincères reconnaissances et de mon grand respect.

Je remercie vivement madame Marie-Francine Moens, Professeur à l'Université catholique, Belgique et madame Sylvie Calabretto, Professeur à LIRIS-INSA, Lyon pour avoir accepté d'être rapporteurs de ce mémoire, et pour l'honneur qu'ils me font en participant au jury. Merci également à monsieur Wessel Kraaij, Professeur à l'Université Radboud Nijmegen, Pays Bas, et à monsieur Claude Chrisment, Professeur à l'Université de Toulouse III, d'avoir accepté de juger ce travail et de faire partie du jury. Je les remercie tous pour leurs évaluations scientifiques, leurs avis éclairés sur le travail accompli et sur les directions de recherche futures.

Cette thèse a été financée est réalisée dans le cadre du projet QUAERO financée par OSEO, l'association française pour l'innovation, je tiens à leurs exprimer ma profonde reconnaissance.

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 exprimer ma reconnaissance à Cécile Laffaire et Guillaume Cabanac. Je les remercie pour leurs aides, leurs disponibilités et leurs générosités pour faire avancer mes

expérimentations. Je remercie mes amis de l'équipe qui ont contribué à la finalisation de quelques tâches d'évaluation dans ce mémoire. Je remercie Ihab, Lamjed, Duy, Saad, Mariam, Mouna et Dana pour leur collaboration et leur disponibilité. Je remercie également toutes les personnes qui ont participé de façon volontaire aux expérimentations menées dans cette thèse.

Je remercie du fond du cœur mes parents, mes sœurs et frères qui n'ont jamais cessé de croire en moi pendant toutes mes années d'études et qui m'ont toujours encouragé à aller de l'avant. Merci à mon Père, qui a semé en moi l'amour du savoir, qui a financé mes études et qui m'a encouragé à affronter les difficultés de l'étranger pour réussir mon avenir.

Enfin et surtout, je remercie avec grand amour mon mari Nour qui m'a soutenu dans les moments difficiles et incertains. Il était la lumière étincelle qui se fortifiait jour après jour et qui animait mon chemin avec plein d'émotions, d'enthousiasme et de bonheur. Grâce à toi je n'ai jamais senti la solitude à l'étranger. C'est la première fois que je sens le goût du succès accompagné par un bonheur complet.

Table des matières

Introduction générale	1
I Accès contextuel à l'information : aperçu sur les modèles et les cadres d'évaluation en RI mobile contextuelle	10
1 Concepts de base de la RI classique et émergence de la RI contextuelle	13
1 Introduction	13
2 Les fondements de la recherche d'information	14
2.1 Concepts de base de la recherche d'information	15
2.2 Processus général de la RI	17
2.2.1 La phase d'indexation	18
2.2.1.1 L'extraction des mots simples	19
2.2.1.2 Élimination des mots vides	19
2.2.1.3 La normalisation (lemmatisation ou radicalisation)	19
2.2.1.4 La pondération des mots	20
2.2.2 La phase de recherche ou d'appariement document-requête	21
2.3 Aperçu des principaux modèles de RI	21
2.3.1 Modèle booléen	23
2.3.2 Modèle vectoriel	23
2.3.3 Modèle probabiliste	24
2.4 Évaluation des performances des systèmes de RI	25
2.4.1 Collection de test	25
2.4.2 Protocole d'évaluation	27
2.4.3 Mesures d'évaluation	27
3 De la RI classique à la RI adaptative	31
3.1 Re-formulation des requêtes	32
3.2 Désambiguïsation du sens des mots de la requête	32
3.3 Regroupement thématique des résultats de recherche	33

3.4	Limites de la RI adaptative	33
4	Vers la RI orientée utilisateur	33
5	Émergence de la RI contextuelle	35
5.1	Facteurs d'émergence de la RI contextuelle	36
5.2	Verrous scientifiques et technologiques	37
5.2.1	Conception	38
5.2.2	Évaluation	38
5.2.3	Mise en œuvre	38
5.3	Notion de contexte en RI : définition et taxonomie	39
5.3.1	Définition du contexte en RI	39
5.3.2	Taxonomie du contexte	40
5.4	Architecture d'un système de RI contextuel	43
5.4.1	La modélisation du contexte	44
5.4.2	L'accès contextuel à l'information	46
5.4.2.1	Reformulation de la requête	47
5.4.2.2	Fonction d'appariement	48
5.4.2.3	L'ordonnancement des résultats	49
6	Conclusion	50
2	RI contextuelle dans un environnement mobile	52
1	Introduction	52
2	Émergence de la RI mobile	54
2.1	Spécificités de la RI mobile	55
2.1.1	Du point de vue du dispositif mobile	55
2.1.2	Du point de vue de l'usage et de la pratique de la recherche d'information	56
2.2	Notion de contexte dans la RI mobile	58
2.3	Problématique de la RI contextuelle et mobile	60
3	Modélisation du contexte en RI mobile	61
3.1	Construction du contexte mobile	61
3.1.1	Les sources d'information et leurs stratégies de collecte	62
3.1.2	Techniques d'acquisition du modèle contextuel	65
3.2	Représentation du contexte	69
3.2.1	Ressources de modélisation	70
3.2.2	Modèles de représentation du contexte	72
3.2.2.1	Modèles de représentation basés sur l'his- torique de recherche	73
3.2.2.2	Modèles attribut-valeur	73
3.2.2.3	Modèles orientés objet	74
3.2.2.4	Modèles hiérarchiques et de balisages	75

	3.2.2.5	Modèles sémantiques et conceptuels . . .	77
	3.2.2.6	D'autres modèles de structures spécifiques	79
	3.3	Synthèse des approches de modélisation du contexte mobile	81
4		Approches et modèles d'accès contextuel dans un environnement mobile	83
	4.1	Tâches d'accès à l'information en RI mobile et contextuelle	83
	4.2	Les modèles d'accès par type de contexte	84
	4.2.1	Techniques d'adaptation au dispositif mobile . .	84
	4.2.1.1	Optimiser l'accès à l'information sur les réseaux mobiles	85
	4.2.1.2	Adapter la visualisation de la liste des résultats de recherche	85
	4.2.1.3	Faciliter la saisie des requêtes	86
	4.2.2	Techniques d'adaptation au contexte utilisateur	87
	4.2.2.1	Modèles de reformulation de la requête .	88
	4.2.2.2	Modèles d'appariement documents-profil	89
	4.2.2.3	Modèles de ré-ordonnement des résultats de la recherche	90
	4.2.3	Techniques d'adaptation au contexte spatial . .	91
	4.2.3.1	Modèles de reformulation de la requête .	92
	4.2.3.2	Modèles d'appariement documents-localisation	92
	4.2.4	Techniques d'adaptation au contexte social . . .	94
	4.2.5	Techniques d'adaptation multidimensionnelle .	95
	4.2.5.1	Combinaison visuelle	96
	4.2.5.2	Combinaison des scores	97
	4.3	Synthèse des modèles d'accès contextuel à l'information en RI mobile	98
5		Évaluation de l'efficacité de la RI contextuelle dans un environnement mobile	99
	5.1	Limites de l'évaluation orientée-laboratoire	99
	5.2	Approches d'évaluation orientées-contexte	100
	5.2.1	Approches basées simulation du contexte	101
	5.2.2	Approches basées sur les fichiers <i>logs</i> de recherche	101
	5.2.3	Approches basées sur des études utilisateurs . .	102
	5.3	Discussion	104
6		Conclusion	105

II Contribution à la définition d'un modèle d'accès contextuel à l'information dans un environnement mo-

3	Modélisation d'un profil flexible pour un utilisateur mobile : proposition et évaluation	109
1	Introduction	109
2	Problématiques et motivations	110
3	Architecture générale de notre approche d'accès personnalisé à l'information	113
3.1	La modélisation du contexte	113
3.2	L'accès contextuel à l'information	115
4	Terminologie, définitions et notations	115
5	Modélisation et apprentissage du profil flexible d'un utilisateur mobile	116
5.1	Modélisation de la situation de l'utilisateur	117
5.1.1	Modélisation de la localisation	117
5.1.2	Modélisation du temps	119
5.1.3	Agrégation dans un modèle de situation	121
5.2	Modélisation des centres d'intérêt	121
5.2.1	Initialisation du contexte de la requête	122
5.2.2	Désambiguïsation du profil de la requête	124
5.2.3	Évolution du profil utilisateur au sein d'une même situation de recherche	126
6	Approche de raisonnement à partir de cas pour l'accès personnalisé à l'information	127
6.1	Représentation des cas	128
6.2	Processus de raisonnement	129
6.2.1	Phase de recherche d'un cas similaire	129
6.2.1.1	Similarité pour la dimension de localisation	130
6.2.1.2	Similarité pour les dimensions de temps	130
6.2.2	Réutilisation : Ré-ordonner les résultats de recherche	131
6.2.3	Révision	132
6.2.4	Apprentissage et mise à jour de la base des cas	132
6.2.4.1	Ajout d'un nouveau cas	133
6.2.4.2	Mise à jour d'un cas existant	133
7	Évaluation expérimentale	133
7.1	Évaluation expérimentale selon un cadre d'évaluation par simulation contextuelle	134
7.1.1	Principe	134
7.1.2	Ressources expérimentales	135
7.1.3	Stratégie d'évaluation	137

7.1.4	Métriques d'évaluation	139
7.1.5	Résultats et discussion	140
7.2	Évaluation expérimentale basée étude journalière	142
7.2.1	Méthodologie de construction du cadre d'évaluation	143
7.2.2	Collection de test	145
7.2.2.1	Requêtes contextuelles	145
7.2.2.2	Documents	145
7.2.2.3	Jugements de pertinence en contexte	146
7.2.3	Stratégie de validation	146
7.2.3.1	Apprentissage/test dans un ordre chronologique	146
7.2.3.2	Validation croisée <i>k-fold cross validation</i>	147
7.2.4	Métriques d'évaluation	147
7.2.5	Application du cadre d'évaluation	148
7.2.6	Résultats et discussions	149
7.2.6.1	Estimation expérimentale du paramètre de la mise en correspondance des situations	149
7.2.6.2	Évaluation de l'efficacité de la recherche personnalisée	150
8	Conclusion	153
4	Identification de la sensibilité de la requête à la localisation	156
1	Introduction	156
2	Problématique et motivations	157
3	Architecture de notre approche d'accès contextuel guidé par le type de la requête	160
4	Prédiction du type du besoin derrière la requête	162
4.1	Modélisation du profil local de la requête	162
4.2	Les attributs de classification basés sur le profil local de la requête	164
4.2.1	La mesure Kullback-Leibler Divergence	165
4.2.2	La mesure de Kurtosis	166
5	Accès contextuel à l'information guidé par la sensibilité de la requête à la localisation	167
5.1	Adaptation par re-formulation de la requête	168
5.2	Adaptation par ré-ordonnancement des résultats de la recherche	168
6	Évaluation expérimentale	169
6.1	Description des ressources expérimentales	169
6.1.1	Collection de requêtes	170

6.1.2	Outils et ressources d'extraction des caractéristiques de la requête	170
6.1.3	Outil de classification des requêtes par apprentissage supervisé basé sur le logiciel Weka	172
6.2	Protocole d'évaluation	172
6.2.1	Protocole d'évaluation de l'efficacité de la technique de classification	173
6.2.2	Protocole d'évaluation de l'efficacité de l'accès contextuel guidé par le type de la requête	173
6.3	Résultats et discussion	174
6.3.1	Évaluation de l'efficacité de nos attributs de classification	175
6.3.2	Une évaluation comparative avec l'approche de Gravano et al. [89]	178
6.3.2.1	Aperçu de l'approche de Gravano et al. [89]	178
6.3.2.2	Résultats de l'évaluation comparative	179
6.3.3	Évaluation de l'efficacité de l'accès contextuel guidé par le type de la requête	179
6.3.3.1	Estimation expérimentale du paramètre de ré-ordonnement	180
6.3.3.2	Comparaison de l'efficacité de l'accès contextuel sur toutes les requêtes	181
6.3.3.3	Comparaison de l'efficacité de l'accès contextuel sur les requêtes mal classées	182
7	Conclusion	183
5	Agrégation prioritaire de plusieurs éléments contextuels : proposition et évaluation	185
1	Introduction	185
2	Problématique et motivations	186
3	Présentation de l'approche d'agrégation prioritaire multi-critères	187
3.1	La modélisation du problème : agrégation prioritaire multi-critères	188
3.2	Les opérateurs d'agrégation prioritaire	189
3.2.1	L'opérateur prioritaire " <i>scoring</i> "	189
3.2.2	L'opérateur prioritaire " <i>and</i> "	190
4	Combinaison multi-dimensionnelle pour la RI mobile	191
4.1	La dimension thème	192
4.2	La dimension centres d'intérêts	192
4.3	La dimension localisation	193

5	Évaluation expérimentale	194
5.1	Cadre expérimental	194
5.1.1	Collection de documents	195
5.1.2	Contextes utilisateurs	195
5.1.3	Collection de requêtes	197
5.1.4	Scénario d'évaluation et jugements de pertinence	198
5.1.5	Métriques d'évaluation	199
5.2	Résultats et discussion	199
5.2.1	Évaluation de l'efficacité de la combinaison de la pertinence multi-dimensionnelle	200
5.2.2	Évaluation de l'efficacité des opérateurs d'agrégation prioritaire	201
6	Conclusion	203
	Conclusion générale	205
	A	211
	Annexe A	211
	B	217
	Annexe B	217
	Bibliographie	224

Liste des tableaux

2.1	Synthèse des différents modèles de représentation du contexte en RI mobile	82
3.1	La liste des “10” premiers termes et leurs poids représentant le concept “ <i>Business and Economy/Restaurants and Bars</i> ” de l’ontologie ODP.	123
3.2	Les contextes géographiques et leurs requêtes associées	135
3.3	Enrichissement du scénario d’utilisation “ <i>musée</i> ” par une partie narrative décrivant la pertinence d’un document dans le contexte.	137
3.4	Comparaison des moyennes des précisions@X entre notre approche de personnalisation et la <i>baseline</i> sur toute les requêtes de test.	141
3.5	Un exemple des requêtes récupérées de l’étude journalière.	148
3.6	Comparaison des moyennes des Précision@X entre notre approche personnalisée et la <i>baseline</i> en utilisant les deux stratégies d’évaluation.	152
3.7	Comparaison des moyennes des nDCG@X entre notre approche personnalisée et la <i>baseline</i> en utilisant les deux stratégies d’évaluation.	152
4.1	Exemple de valeurs de <i>locationKL</i> pour différents types de requêtes	166
4.2	Exemple de valeurs de <i>kurtosis</i> pour différents types de requêtes	167
4.3	Les 5 plus faibles et/ou plus fortes valeurs de LocationKL et Kurtosis	176
4.4	Résultats de classification des requêtes de test.	177
4.5	Performance de la classification sur les classes locale et globale : comparaison entre notre approche et l’approche de [89] sur l’ensemble de nos requêtes de test.	179
5.1	La liste des 30 concepts testés représentant les centres d’intérêts des utilisateurs.	196
5.2	Un exemple de requêtes formulées selon chaque stratégie.	197
5.3	Comparaison de la MAP calculée sur les résultats de recherche des opérateurs “ <i>scoring</i> ” et “ <i>weighted average</i> ” pour chaque scénario d’évaluation.	202

5.4	Comparaison de la MAP calculée sur les résultats de recherche des opérateurs “ <i>scoring</i> ” et “ <i>weighted average</i> ” pour chaque stratégie de formulation des requêtes.	203
A.1	Collection de requêtes extraites du <i>log</i> d’AOL.	211
B.1	Collection de requêtes simulées.	217

Table des figures

1.1	Approche générale de la RI : Processus en U	17
1.2	Classification des modèles de RI selon Kuroopka [138]	22
1.3	Rappel et précision.	28
1.4	Forme générale de la courbe de précision-rappel d'un SRI	29
1.5	Le concept du contexte multidimensionnel en RI	40
1.6	Architecture de base d'un SRI contextuel	43
2.1	Domaines liés à notre problématique de recherche.	53
2.2	Pourcentage des requêtes par catégorie selon l'analyse de [122].	57
2.3	Pourcentage des requêtes selon le type du besoin informationnel de l'utilisateur mobile.	57
2.4	Pourcentage des différents facteurs contextuels derrière le besoin informationnel de l'utilisateur mobile.	58
2.5	Réseau bayésien pour inférer le type de place visitée à partir de la trace temporelle (les nœuds blancs sont fournis en entrée du réseau).	67
2.6	Processus de prédiction du type d'activité de l'utilisateur mobile.	69
2.7	L'interface de l'ontologie générale ODP.	70
2.8	Exemple d'ontologie extraite pour la requête "apple".	71
2.9	Représentation partielle d'un concept "Option Contract" et ses propriétés.	72
2.10	Exemple d'ontologie propriétaire développée dans [107].	73
2.11	Un schéma simplifié du modèle objet du profil situationnel.	75
2.12	Exemple du modèle de contexte d'objet.	76
2.13	Exemple de modèle utilisateur avec des balises de préférences de contenu et des balises de zones temporelles.	77
2.14	Descripteur de contexte.	79
2.15	Modèle d'arbre de contexte dimensionnel.	80
2.16	Schéma de causalité modélisant les facteurs de temps, de centres d'intérêts et de localisation.	81
2.17	Exemple de trajectoires sémantiques.	81
2.18	Exemple de profils construits dans [266] et leurs exploitation pour la reformulation des requêtes.	88

2.19	Un aperçu de l’interface d’affichage par canaux contextuels selon [52].	96
2.20	Un snapshot de l’interface de recherche intégrant les dimensions du contexte selon [52].	96
3.1	Architecture générale de notre approche de RI personnalisé . . .	113
3.2	Modèle simplifié de l’ontologie spatiale.	118
3.3	Modèle simplifié de l’ontologie temporelle.	120
3.4	Dimensions agrégées dans le modèle de situation.	121
3.5	Le cycle de raisonnement à partir de cas (RàpC).	128
3.6	Stratégie de validation croisée appliqué sur chaque cas d’utilisation.	139
3.7	Cadre d’évaluation basé étude journalière.	141
3.8	Comparaison de l’efficacité entre notre modèle personnalisé et Yahoo BOSS sur chaque requête en termes. (a) de la DCG@10, (b) de la DCG@20.	142
3.9	Cadre d’évaluation basé étude journalière.	144
3.10	Effet du paramètre Beta sur la pertinence de la similarité entre situation.	150
3.11	Effet du paramètre gamma dans la combinaison des scores sur la précision en utilisant.	151
3.12	Effet du paramètre gamma dans la combinaison des scores sur la nDCG en utilisant.	151
4.1	Un aperçu des résultats de Google retournés pour la requête locale implicite “ <i>looking to hire a dj</i> ”.	158
4.2	Un aperçu des résultats de Google Local retournés pour la requête locale implicite “ <i>looking to hire a dj</i> ”.	159
4.3	Architecture générale du système s’accès contextuel intégrant notre approche de classification.	161
4.4	Schéma général de notre modèle de classification des requêtes. .	161
4.5	Comparaison entre le profil de la collection- $P(l C)$ et celui de la requête “ <i>airport shuttle</i> ”- $P(l Q)$	164
4.6	Comparaison entre le profil local d’une requête globale “ <i>zig-zone</i> ”, une requête locale implicite “ <i>jobs</i> ” et une requête locale explicite “ <i>nj lottery</i> ”.	165
4.7	Un snapshot des résultats de la recherche retournés par l’API Google pour la requête “ <i>airport shuttle</i> ” : il illustre la présence et la diversité des localisations (surlignées en rouge).	171
4.8	Variation de l’efficacité de la recherche en termes de précision moyenne en fonction du paramètre de ré-ordonnancement α . . .	180

4.9	Variation de l'efficacité de la recherche en termes de nDCG moyenne en fonction du paramètre de ré-ordonnement α	180
4.10	Comparaison des moyennes de Précision@X entre nos approches contextuelles et la <i>baseline</i> sur toutes les requêtes de test.	181
4.11	Comparaison des moyennes de nDCG@X entre nos approches contextuelles et la <i>baseline</i> sur toutes les requêtes de test.	182
4.12	Comparaison des moyennes de Précision@X entre nos approches contextuelles et la <i>baseline</i> sur les requêtes mal classées.	183
4.13	Comparaison des moyennes de nDCG@X entre nos approches contextuelles et la <i>baseline</i> sur les requêtes mal classées.	183
5.1	Un aperçu de la structure hiérarchique de l'ODP.	195
5.2	Performance en termes de (a) précision, (b) nDCG à n moyennée sur l'ensemble des scénarios d'évaluation.	200
5.3	Performance en termes de précision à n moyennée sur l'ensemble des scénarios d'évaluation.	201
5.4	Performance en termes de nDCG à n moyennée sur l'ensemble des scénarios d'évaluation.	202

Introduction générale

Contexte et problématique

Les systèmes de recherche d'information (SRI) permettent de sélectionner à partir d'une collection de documents, des informations pertinentes répondant à des besoins utilisateurs, exprimés sous forme de requêtes.

La recherche d'information (RI), domaine déjà ancien, n'a cessé d'évoluer dans le but de rationaliser le processus complexe permettant l'identification, au sein de volume de plus en plus importants d'information, celles qui sont potentiellement intéressantes pour l'utilisateur. En effet, la popularisation d'Internet ainsi que l'explosion des services de l'information de nos jours ont propulsé la RI au premier plan. Cependant, la surabondance de l'information et sa large accessibilité à travers le web, ont engendré la dégradation de la qualité des résultats retournés par les SRI [215]. En effet, les modèles classiques de RI se caractérisent 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 des documents à l'utilisateur qui se trouve ainsi face à une surcharge informationnelle qui le désoriente où il doit prendre la charge de sélectionner ce qui est pertinent de ce qui ne l'est pas. En clair, le problème n'est pas tant la disponibilité de l'information, mais sa pertinence relativement à un contexte d'utilisation spécifique. Dans [42] les auteurs montrent que l'approche généraliste des outils disponibles de RI sur le web 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, est à l'origine des problèmes évoqués.

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 au besoin précis de l'utilisateur en termes des documents pertinents associés. Les techniques dérivées de ce cadre concernent la reformulation de requête par ré-injection de pertinence [97], l'expansion des requêtes par désambiguïsation [198] ou le regroupement thématique des

résultats [272]. 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.

Suite aux limitations des techniques de la RI adaptative et dans le but d'améliorer la performance des SRI, plusieurs études [38, 120, 102] 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. 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 [104]. 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*”.

L'objectif principal des travaux en recherche contextuelle d'information (RCI) est d'optimiser la pertinence des résultats de recherche, en impliquant deux étapes complémentaires : définition du contexte du besoin en information de l'utilisateur, puis adaptation de la recherche en le prenant en considération dans le processus de sélection de l'information. Les recherches actuelles en RCI se sont focalisées sur la modélisation et l'exploitation du contexte cognitif de l'utilisateur [225, 70, 152, 220, 162] défini par les centres d'intérêts de l'utilisateur et ses préférences de recherche sur la notion de pertinence.

Cependant, l'essor des technologies mobiles (en termes de capacité des dispositifs et de réseaux) de nos jours, permet à un utilisateur nomade d'accéder à des informations depuis pratiquement toute place et à tout moment. Ce nouveau cadre d'utilisation accentue le besoin et la nécessité de prendre en considération d'autres facteurs contextuels afin d'offrir des systèmes de recherche mobiles capables d'adapter les informations délivrées à l'utilisateur et à son contexte. Des travaux récents [86, 118] tentent de répondre aux problématiques inhérentes à ce nouveau cadre d'utilisation et d'améliorer ainsi les performances de recherche dans cet environnement en modélisant et intégrant le contexte dans le processus de recherche mobile. 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 dans ce domaine concernent principalement la modélisation du contexte, l'évolution du contexte

au cours des activités 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. Parmi les approches de RI contextuelle dans ce domaine, nous citons particulièrement les modèles qui répondent aux contraintes liées au dispositif mobile, les modèles qui répondent au contexte physique défini par la localisation géographique de l'utilisateur, 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 d'adaptation multidimensionnelle qui modélisent et combinent plusieurs éléments contextuels.

Les approches basées sur le dispositif mobile ont abordé les questions liées à l'adaptation de la recherche aux contraintes imposées par les fonctionnalités limitées des appareils mobiles. Des approches sont proposées pour adapter la visualisation de la liste des résultats de recherche [21, 228] et pour faciliter la saisie des requêtes [124, 242]. Les approches basées sur la localisation exploitent la capacité des appareils mobiles à fournir la localisation physique de l'utilisateur pour adapter les résultats de la recherche. Alors que certains travaux proposent de filtrer l'information basée sur la proximité géographique [165] d'autres reformulent la requête en incluant le contexte géographique de l'utilisateur mobile [94, 35]. Les approches basées sur les centres d'intérêt ont porté sur la modélisation des centres d'intérêt de l'utilisateur mobile pour améliorer la précision des résultats de recherche [266, 172, 29, 27]. Le défi soulevé par cette catégorie de travaux est de tenir compte de la dynamique des centres d'intérêt de l'utilisateur liée au changement de son environnement lors du processus de personnalisation. Les approches d'adaptation multidimensionnelle [139, 244, 28, 44] proposent des techniques de combinaison de la pertinence issue de l'exploitation de plusieurs dimensions contextuelles. Cela nécessite de réfléchir le modèle de combinaison adéquat pour prendre en compte différentes fonctions de *ranking*.

L'émergence de la RI contextuelle 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 compte des interactions de l'utilisateur avec le système et son environnement en considérant que son besoin en information et sa perception de pertinence comme des notions statiques. On distingue trois principaux types d'évaluations adaptés à la RI contextuelle : l'évaluation par simulation de contextes, l'évaluation basée sur les *logs* de recherche et l'évaluation par des études utilisateurs. Dans ces évaluations, les cadres proposés sont fortement liés au SRI contextuel à évaluer quant à la spécification des données exploitées pour la construction du modèle de contexte et les mesures d'évaluation adoptées.

Contribution

Les travaux présentés dans ce mémoire se situent dans le contexte précis de l'accès contextuel à l'information dans un environnement mobile. Plus précisément, nos contributions dans le domaine porte sur cinq volets : (1) la définition du contexte selon deux dimensions ; le profil utilisateur et le type de besoin (global, local explicite et local implicite), (2) la modélisation du contexte de recherche, plus précisément, la construction et l'évolution du profil situationnel de l'utilisateur et la modélisation de la sensibilité de la requête à la localisation, (3) l'exploitation du profil situationnel de l'utilisateur et du modèle de la requête pour l'adaptation du processus de RI, (4) la proposition d'un modèle d'agrégation prioritaire de pertinence multidimensionnelle pour personnaliser la RI mobile et (5) la proposition de cadres d'évaluation adaptés qui permettent d'évaluer l'efficacité de nos approches de recherche contextuelles.

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 profil situationnel de l'utilisateur et la sensibilité de la requête à la localisation. 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) 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 désambiguïser la recherche, de répondre à des besoins en information dans des situations de recherche spécifiques. Les défis majeurs à ce niveau consistent à représenter les situations de recherche de l'utilisateur et ses centres d'intérêts attachés, puis à les exploiter dans un modèle d'accès personnalisé à l'information.
 - (b) La sensibilité de la requête à la localisation est importante pour améliorer la performance de recherche en tenant compte de la localisation de l'utilisateur mobile. En effet, ce facteur contextuel est d'une grande influence sur les besoins en information des utilisateurs en mobilité. Les défis majeurs à ce niveau consistent à définir un modèle de localisation des requêtes et des caractéristiques spécifiques à chaque type de besoin.
2. **Modélisation du contexte de recherche** :
 - (a) Modélisation du profil de l'utilisateur :

Nous définissons une représentation sémantique du profil de l'utilisateur basée sur l'exploitation de la situation spatio-temporelle de l'utilisateur ainsi que le sous-ensemble de documents jugés pertinents

implicitement par l'utilisateur et des ontologies appropriées. Notre objectif est de dégager à partir des interactions de l'utilisateur avec le système et de son évolution dans son contexte situationnel, une représentation conceptuelle de ses centres d'intérêts liés à chaque situation de recherche. A cet effet, la modélisation du profil utilisateur dans notre approche est caractérisée par :

- la représentation sémantique du contexte temporel et spatial de recherche via des l'ontologies spatiale et temporelle et leur intégration dans un modèle unique de situation.
- la représentation sémantique du profil de l'utilisateur selon un ensemble de concepts sémantiquement liés via l'ontologie de référence utilisée.
- un procédé d'évolution du profil situationnel au sein d'une même situation de recherche qui permet de scruter le changement des centres d'intérêts au cours du temps. Ce procédé est conjointement lié à un mécanisme de comparaison des situations de recherche basé sur une mesure de similarité de situations.

- (b) Identification de la sensibilité de la requête à la localisation : nous proposons de classer les requêtes selon trois classes : globale, locale explicite et locale implicite. Nous définissons pour cela un modèle de langue de localisation de la requête à partir des “ n ” premiers résultats de recherche retournés par un moteur web général. La prédiction du type de la requête se base sur le calcul de deux caractéristiques, définies sur le modèle de langue de localisation de la requête, issues de la théorie de l'information et de la théorie des probabilités, notamment les mesures de “*Kullback-Leibler Divergence*” et de “*Kurtosis*”.

3. Exploitation du modèle du contexte dans le processus de RI :

- (a) le profil situationnel de l'utilisateur est exploité dans le ré-ordonnement des résultats de recherche des requêtes traitant du même besoin en information. Le profil situationnel est identifié selon une technique de RàPC et exploité dans le processus de ré-ordonnement des résultats des requêtes émises dans la même situation de recherche.
- (b) Le modèle de classification des requêtes selon leur sensibilité à la localisation de l'utilisateur est intégré dans un modèle d'accès à l'information guidé par le type de la requête. Il est exploité selon deux techniques par reformulation de requête et par ré-ordonnement des résultats de recherche.

4. Agrégation prioritaire et personnalisée de pertinence multidimensionnelle :

Nous avons proposé d'appliquer un modèle d'agrégation prioritaire pour la combinaison de pertinence multidimensionnelle pour personnaliser la RI mobile. La particularité de ce modèle d'agrégation réside dans l'exploitation de deux opérateurs d'agrégation prioritaire permettant d'adapter les résultats de recherche selon les préférences de l'utilisateur exprimées sur les critères de pertinence. Nous avons appliqué ce modèle pour la combinaison de trois pertinences selon : la thématique de la requête, les centres d'intérêts de l'utilisateur et sa localisation.

5. Évaluation des performances d'un processus d'accès contextuel à l'information :

En l'absence de cadre d'évaluation standard pour la RI contextuelle, nous avons proposé des cadres d'évaluation adaptés pour l'évaluation de chacune de nos contributions. Un aperçu de ces cadres d'évaluation est le suivant :

- (a) Deux cadres d'évaluation sont proposés pour évaluer l'efficacité de notre approche de personnalisation intégrant le profil situationnel de l'utilisateur. Le premier cadre est élaboré selon une approche d'évaluation par simulation des contextes. Il est basé sur la simulation d'un ensemble de scénarios de recherche adaptés au cas d'un utilisateur mobile. Chaque scénario de recherche est défini par un contexte géographique, et un ensemble de requêtes associées. La construction des profils utilisateurs sur les scénarios de recherche simulées est basée sur des jugements de pertinence donnés par des utilisateurs en tenant compte du contexte de la recherche. L'évaluation de la recherche personnalisée est basée sur un protocole de cross-validation au sein de chaque scénario de recherche simulée. Le second cadre d'évaluation proposé est élaboré selon une approche d'évaluation par étude journalière. Il est basé sur l'exploitation de requêtes et de contextes d'utilisateurs réels en situations de recherche mobile. Les jugements de pertinence sont donnés par les participants à l'étude journalière en tenant compte de leurs contextes réels de recherche.
- (b) Le cadre d'évaluation proposé pour valider notre approche de détection de la sensibilité de la requête à la localisation est basé sur une collection de requêtes annotées manuellement issue du *log* de recherche d'AOL. L'évaluation de l'efficacité de la classification des requêtes est basée sur le protocole de la cross-validation et une comparaison à une approche de l'état de l'art. L'évaluation de l'efficacité de la recherche selon un scénario de recherche intégrant le type de la requête est basée sur des jugements de pertinence des documents effectués en tenant compte du type de la requête.
- (c) Le cadre d'évaluation proposé pour évaluer notre approche pour la combinaison de pertinence multidimensionnelle est également élaboré

selon une approche d'évaluation par simulation des contextes. La collection de test est construite en se basant sur une partie de la branche de l'ontologie générale ODP¹. Les profils utilisateurs et leurs localisations sont simulés par des concepts de l'ontologie. Les jugements de pertinence sont simulés selon un algorithme automatique qui se base sur des évidences issues de la structure de l'ontologie. Dans ce cadre, l'évaluation de l'efficacité de la recherche est basée sur des scénarios de recherche simulant différents ordres de priorité sur les pertinences contextuelles considérées.

Plan de la thèse

Ce mémoire est constitué de deux parties principales : la première présente le contexte général dans lequel se situe notre travail, à savoir la recherche d'information et plus précisément la recherche d'information contextuelle dans un environnement mobile ; la seconde partie comprend le détail de notre contribution dans le domaine.

L'objectif de la première partie, "***Accès contextuel à l'information : aperçu sur les modèles et les cadres d'évaluation en RI mobile contextuelle***", est de porter la lumière sur le domaine de la recherche d'information contextuelle, puis son application pour embrasser la recherche dans l'environnement mobile. La première partie regroupe deux chapitres.

Le chapitre 1 présente les notions et concepts de base de la RI ainsi que les facteurs d'émergence de la RI contextuelle. 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 contextuelle. Nous présentons ensuite la notion de contexte en RI ainsi que les modèles de sa présentation et les approches de son exploitation dans le cadre de l'adaptation du processus de RI à la notion du contexte multidimensionnel.

Le chapitre 2 traite de l'émergence de la RI mobile et contextuelle selon deux principales pistes de recherche ; les approches et modèles de la RI contextuelle mobile ainsi que les approches d'évaluation des techniques contextuelles. Selon les approches de la RI contextuelle mobile, les différentes techniques d'acquisition du contexte, les différents modèles de représentation et de procédés de construction et d'évolution du modèle du contexte, ainsi que les différents modèles d'accès contextuel à l'information intégrant le modèle du contexte dans

¹<http://www.dmoz.org/>

l'une des phases du processus de recherche sont abordés. Selon la démarche d'évaluation contextuelle, nous présentons d'abord les limites du paradigme d'évaluation classique de *Cranfield* dans l'évaluation des SRI contextuel. Ensuite, nous traitons de l'émergence des trois grandes tendances des approches d'évaluation orientées-contexte ainsi que des mesures d'évaluation orientées-contexte. Plus particulièrement, il s'agit des approches d'évaluation par simulation de contexte, des approches d'évaluation basées sur les fichiers *logs* de recherche et des approches d'évaluation basées sur des études utilisateurs.

La seconde partie du mémoire intitulée : “***Contribution à la définition d'un modèle d'accès contextuel à l'information dans un environnement mobile***”, détaille notre contribution dans le domaine de la RI contextuelle et mobile. Elle comprend trois chapitres.

Le chapitre 3 présente notre contribution à la définition d'un profil situationnel de l'utilisateur pour une recherche d'information personnalisée. Deux modèles de représentation sémantique de la situation de recherche et des centres d'intérêts de l'utilisateur sont présentés. Ensuite, l'application de la technique de raisonnement à partir de cas (RàPC) pour l'exploitation du profil situationnel dans le processus de recherche y est décrite. Concernant, l'évaluation expérimentale, deux nouveaux cadres d'évaluation sont présentés. Le premier cadre suit la démarche d'évaluation par simulation contextuelle et le second cadre est basé sur une étude journalière. Ce chapitre présente également les résultats expérimentaux obtenus pour l'évaluation de notre approche de personnalisation selon ces deux cadres d'évaluation.

Le chapitre 4 présente notre approche de prédiction de la sensibilité de la requête à la localisation de l'utilisateur ainsi qu'une évaluation expérimentale de cette approche. Le principe de l'identification de la sensibilité de la requête à la localisation basé sur un modèle de langue de localisation des requêtes et les attributs de classification définis y sont décrits. Deux techniques d'adaptation des résultats de recherche selon le type de la requête y sont présentées. Concernant l'évaluation expérimentale, un protocole d'évaluation intégrant un échantillon de requêtes extraites d'un *log* de recherche web ainsi que les résultats de l'évaluation y sont présentés.

Le chapitre 5 présente notre approche pour une adaptation contextuelle multidimensionnelle pour la RI mobile ainsi qu'une évaluation expérimentale de cette approche. Le principe de cette approche d'adaptation multidimensionnelle reposant sur l'application d'un modèle d'agrégation prioritaire pour la combinaison de pertinence multidimensionnelle y est présenté. En particulier nous décrivons deux opérateurs d'agrégation prioritaire permettant d'adapter les résultats de recherche selon les préférences de l'utilisateur exprimées sur les critères de pertinence. Nous présentons ensuite l'application de ce modèle pour la combinaison de trois dimensions de pertinence notamment : thématique,

centres d'intérêts et localisation. Concernant l'évaluation expérimentale, un protocole d'évaluation basé simulation contextuelle ainsi que les résultats de l'évaluation y sont présentés.

En conclusion, nous dressons le bilan de nos travaux réalisés dans le cadre de l'accès contextuel à l'information dans un environnement mobile. Nous introduisons ensuite les perspectives de ces travaux selon la contextualisation de la RI mobile et les cadres d'évaluation des SRI contextuels.

Première partie

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

Introduction

La prolifération de la technologie mobile (PDAs, téléphones mobiles, ...) a rendu l'accès à une grande masse de documents sur le web, possible à toute place et à tout moment. Ce nouveau cadre d'utilisation, caractérisé par : 1) une difficulté de saisie des requêtes, 2) une limitation en temps pour consulter les résultats de recherche et 3) par la mobilité de l'utilisateur, accentue le besoin et la nécessité de prendre en considération des informations du contexte pour améliorer la précision de recherche.

Les moteurs de recherche traditionnels considèrent peu le contexte de la recherche et ne sont pas adaptés à l'environnement mobile. Des travaux récents en RI contextuelle tentent d'améliorer les performances de recherche dans cet environnement. Le principe de base de la RI contextuelle consiste à modéliser puis intégrer le contexte de recherche comme une structure informationnelle additionnelle dans la chaîne d'accès à l'information. La mise en œuvre des approches de RI contextuelle est basée sur deux principales composantes : la modélisation du contexte et l'accès contextuel à l'information. Nous résumons les aspects critiques de ces deux composantes comme suit :

- La modélisation du contexte repose sur des techniques et des outils permettant non seulement de capturer et de modéliser le contexte de recherche de l'utilisateur mais aussi de gérer son évolution de manière dynamique au cours du temps.
- L'accès contextuel à l'information consiste à intégrer le contexte de recherche dans le but d'améliorer la précision de recherche. L'efficacité de la recherche contextuelle est potentiellement dépendante de la représentation du contexte, sa précision et son efficacité à l'égard des besoins en information spécifiques de l'utilisateur.

L'émergence de la RI contextuelle a remis en cause le paradigme d'évaluation classique des SRI. Les limites majeures de ce paradigme sont principalement liées 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 en réponse à une requête donnée, et au protocole d'évaluation qui ne tient pas compte des interactions de l'utilisateur avec le système et avec son environnement en considérant que son besoin en information et sa perception de pertinence sont des notions statiques.

A cet effet, le défi majeur dans l'évaluation de l'efficacité de la recherche contextuelle consiste à évaluer la performance de recherche dans des situations de recherche réelles en tenant compte de la perception de l'utilisateur de la pertinence de l'information vis-à-vis de ses préférences et son contexte de recherche.

Nous nous intéressons dans cette partie à introduire l'émergence de la RI contextuelle et mobile selon les principales pistes de recherche suivantes :

- Nous présentons un aperçu des principaux modèles et techniques proposées pour la modélisation du contexte dans cet environnement,
- Nous nous intéressons ensuite aux approches et modèles proposés pour l'exploitation du contexte dans le processus d'accès à l'information,
- Nous abordons également un aperçu des différentes approches d'évaluation orientées-contexte adaptées à l'évaluation de l'efficacité de la recherche contextuelle.

Cette partie est organisée comme suit. Le chapitre 1 présente les concepts de base de la RI classique ainsi que l'émergence de la RI contextuelle. Le chapitre 2 traite de l'émergence de la RI mobile et de l'application de la RI contextuelle dans cet environnement en présentant les différents modèles et techniques de représentation et d'exploitation du contexte ainsi que les démarches d'évaluation contextuelles proposées dans ce domaine.

Chapitre 1

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

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 modèles 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 [42] montrent que l'origine de ces limites réside en partie dans le fait que ces modèles 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'information.

Les premières techniques développées en RI dans le but de remédier à cette problématique s'apparentent à la RI adaptative. Il s'agit du développement

de techniques de re-formulation de requêtes [198], de désambiguïsation du besoin derrière les requêtes [220] ou du regroupement thématique des résultats de la recherche (Vivisimo¹, Exalead²). Plus récemment, les travaux en RI se sont orientés vers une nouvelle génération d’approches de recherche basées sur l’accès contextuel à l’information. Il s’agit du développement de techniques de modélisation et d’exploitation du contexte de recherche, notamment les approches de personnalisation des résultats de la recherche basées sur la modélisation des centres d’intérêts de l’utilisateur (Persona [237], Yahoo Y!Q³ [136], *Google personalized search*⁴). L’objectif de la RI contextuelle est de mieux répondre aux besoins en information de l’utilisateur en intégrant le contexte de recherche dans le processus d’accès à l’information.

Ce chapitre traite des concepts de base de la RI classique ainsi que de l’émergence de la RI contextuelle. La section 2 présente tout d’abord les fondements de la RI classique. Nous abordons les notions et les modèles de base de la RI classique, puis nous présentons la démarche d’évaluation classique des systèmes de recherche d’information. La section 3 présente l’orientation des travaux vers la RI adaptative pour laquelle nous donnons un aperçu des techniques développées. La section 4 décrit l’évolution de la RI vers la RI utilisateur en mettant l’accent sur les trois révolutions importantes qui ont marqué cette évolution. La section 5 traite de l’émergence de la RI contextuelle. Plus particulièrement, nous abordons les problématiques majeures de la RI classique en présence du contexte, nous essayons ensuite de définir la notion de contexte et l’architecture générale d’un système d’accès contextuel à l’information, puis nous donnons un aperçu des principaux modèles et techniques de modélisation et d’exploitation du contexte en RI, et nous terminons par un aperçu des verrous scientifiques et technologiques qui ont marqué les travaux de recherche en RI contextuelle. La dernière section conclut le chapitre.

2 Les fondements de la recherche d’information

La Recherche d’Information (RI) ([250, 205, 22]) est la branche de l’informatique qui s’intéresse à l’acquisition, l’organisation, le stockage, la recherche et la distribution de l’information. En bref, un système de recherche d’information (SRI) permet de sélectionner à partir d’une collection de documents, des informations pertinentes répondant à des besoins utilisateurs, exprimés sous forme de requêtes. Nous abordons dans la suite de cette section les concepts

¹<http://clusty.com/>

²<http://www.exalead.com/search/>

³<http://yq.search.yahoo.com/>

⁴<https://www.google.com/history/>

de base de la recherche d'information, puis nous décrivons le processus général d'un SRI, ensuite nous passons en revue les principaux modèles de RI et nous présentons la démarche classique d'évaluation des SRI.

2.1 Concepts de base de la recherche d'information

Plusieurs concepts clés s'articulent autour de la définition d'un SRI :

1. **Collection de documents** : la collection de documents (ou fond documentaire, corpus) constitue l'ensemble des informations (des documents) exploitables et accessibles. Nous utiliserons dans la suite les termes : corpus, collection ou fonds documentaires de manière indifférente.
2. **Document** : le document constitue l'information élémentaire d'une collection de documents. L'information élémentaire, appelée aussi granule de document, peut représenter tout ou une partie d'un document. Nous utiliserons dans la suite les termes information ou document pour désigner un granule documentaire.
3. **Besoin en information** : cette notion est souvent assimilée au besoin de l'utilisateur. Ingwersen [100] définit trois types de besoin utilisateur :
 - *Besoin vérificatif* : l'utilisateur cherche à vérifier le texte avec les données connues qu'il possède déjà. Il recherche donc une donnée particulière, et sait même souvent comment y accéder. La recherche d'un article sur Internet à partir d'une adresse connue serait un exemple d'un tel besoin. Un besoin de type vérificatif est dit stable, c'est-à-dire qu'il ne change pas au cours de la recherche.
 - *Besoin thématique connu* : l'utilisateur cherche à clarifier, à revoir ou à trouver de nouvelles informations dans un sujet et domaine connus. Un besoin de ce type peut être stable ou variable; il est très possible en effet que le besoin de l'utilisateur s'affine au cours de la recherche.
 - *Besoin thématique inconnu* : pour ce type de besoins, l'utilisateur cherche de nouveaux concepts ou de nouvelles relations hors des sujets ou domaines qui lui sont familiers. Le besoin est intrinsèquement variable et est toujours exprimé de façon incomplète.
4. **Requête** : la requête est l'expression du besoin en information de l'utilisateur. Elle représente l'interface entre le SRI et l'utilisateur. Divers types de langages d'interrogation sont proposés dans la littérature. Une requête

est un ensemble de mots clés, mais elle peut être exprimée en langage naturel, booléen ou graphique.

5. **Pertinence** : la pertinence est une notion fondamentale dans le domaine de la RI. Une définition simple de cette notion complexe est donnée dans [34] : “*La pertinence est la correspondance entre un document et une requête, ou encore une mesure d’informativité du document à la requête*”. On retrouve dans [179] une définition plus large de la notion de pertinence, dépendant de nombreux critères liés au contexte de la recherche, tels que : le degré de correspondance (*aboutness*), l’utilité (*usefulness/utility*), rentabilité (*usability*), l’importance, ... des résultats retournés par rapport aux objectifs, aux intérêts, à la situation intrinsèque du moment. Ces différents critères ont amené à la catégorisation de la pertinence utilisateur principalement en 5 classes de pertinence : la pertinence algorithmique, la pertinence thématique, la pertinence cognitive, la pertinence situationnelle et la pertinence motivationnelle (ou affective) [208]. Elles peuvent être définies comme suit :

- *la pertinence algorithmique (ou système)* : souvent présentée par un score de l’adéquation du contenu des documents vis-à-vis de celui de la requête. Pour mesurer cette adéquation, le SRI procède à la comparaison de la représentation interne de la requête et de la représentation interne des documents. Le degré de similitude du document et de la requête mesure la pertinence du document pour cette requête. Le but de tout SRI est de rapprocher la pertinence algorithmique calculée par le système aux jugements de pertinence donnés par des vrais utilisateurs.
- *la pertinence thématique* : traduit le degré d’adéquation de l’information retrouvée au thème évoqué par le sujet de la requête ; c’est la mesure de pertinence utilisée par les assesseurs de la campagne d’évaluation TREC.
- *la pertinence cognitive* : représente la relation entre l’état de la connaissance intrinsèque de l’utilisateur et l’information portée par les documents telle qu’interprétée par l’utilisateur ; cette pertinence se caractérise par une dynamique qui permet d’améliorer la connaissance de l’utilisateur via l’information renvoyée le long de sa recherche.
- *la pertinence situationnelle* : est vue comme l’utilité de l’information retrouvée par rapport à la tâche ou le problème posé par l’utilisateur.
- *la pertinence motivationnelle (ou affective)* : décrit la relation entre les intentions, les buts et les motivations de la recherche tels que fixés par l’utilisateur d’une part et les informations retrouvées d’autre part.

2.2 Processus général de la RI

Pour établir la mise en correspondance des besoins en information des utilisateurs d'une part et des informations contenues dans les fonds documentaires d'autre part, les systèmes de RI adoptent une démarche simple, couramment appelée : Processus en U de recherche d'information. Ce processus consiste en deux phases principales (comme l'illustre la figure 1.1 qui sont : l'indexation et l'appariement.

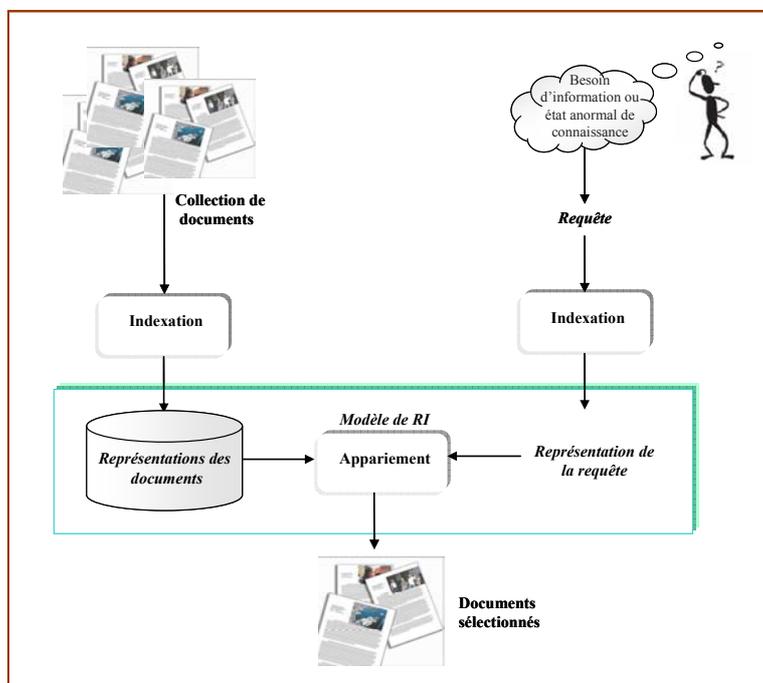


FIG. 1.1 – Approche générale de la RI : Processus en U

1. **Processus de représentation ou d'indexation** : Comme il n'est pas pratique de chercher l'information désirée par un parcours complet de tous les textes des collections, les documents et les requêtes sont représentés par un ensemble de termes caractérisant leurs contenus. C'est une opération fondamentale en RI, à l'issue de laquelle un descripteur du document ou de la requête, est construit. Un descripteur est une liste de termes significatifs pour l'unité textuelle correspondante, auxquels sont associés généralement des poids pour différencier leur degré de représentativité.
2. **Processus de recherche ou d'appariement requête-document** : Il représente le processus noyau d'un SRI. Il mesure un degré de correspondance (pertinence) d'un document vis-à-vis d'une requête. Il est

étroitement lié au modèle de représentation des documents et des requêtes. La problématique majeure des SRI est de retrouver les quelques dizaines ou milliers de documents pertinents parmi des millions de documents. Cet écart de cardinalité rend cette tâche encore plus difficile.

2.2.1 La phase d'indexation

L'objectif principal de cette étape est de fournir des représentations des documents et des requêtes qui soient facilement exploitable par la machine dans la phase de recherche. En effet, les documents (sous forme de texte libre) sont difficiles à exploiter tels quels lors de la phase de recherche. Cette représentation est souvent une liste de mots clés significatifs, que l'on nomme : descripteur du document (ou de la requête). L'indexation peut être : manuelle, semi-automatique ou automatique.

1. **Indexation manuelle** : chaque document est analysé par un spécialiste du domaine ou un documentaliste, qui effectue le choix des mots en utilisant un vocabulaire contrôlé (liste hiérarchique, thésaurus, lexique, ...). L'indexation manuelle fournit une terminologie pour indexer et rechercher des documents, assurant ainsi de meilleure pertinence dans les réponses apportées par le SRI [191]. Cependant, elle présente plusieurs inconvénients liés notamment à l'effort et le prix qu'elle exige (en temps et en nombres de personnes). De plus, un degré de subjectivité lié au facteur humain fait que pour un même document, des termes différents peuvent être sélectionnés par des indexeurs différents. Il peut même arriver qu'une personne, à des moments différents, indexe différemment le même document.
2. **Indexation semi-automatique** : les termes du document sont extraits en un premier temps par un processus automatique. Puis l'indexeur- un spécialiste du domaine, intervient pour effectuer le choix final des termes significatifs et établir les relations entre les mots clés, généralement en utilisant un vocabulaire contrôlé sous forme de thésaurus ou de base terminologique.
3. **Indexation automatique** : ici le processus d'indexation est complètement automatisé. L'indexation automatique peut se faire selon une méthode linguistique ou statistique. L'approche courante est plutôt statistique.

D'une façon générale, un processus d'indexation automatique comprend un ensemble de traitements automatisés sur les documents, qui sont les suivants : extraction des mots simples, élimination des mots vides, normalisation et pondération des mots.

2.2.1.1 L'extraction des mots simples

Cette opération consiste à extraire du document un ensemble de termes ou de mots simples par une analyse lexicale permettant d'identifier les termes en reconnaissant les espaces de séparation des mots, des caractères spéciaux, des chiffres, les ponctuations, etc.

2.2.1.2 Élimination des mots vides

La liste des mots simples extraite précédemment peut contenir des mots non significatifs, appelés "mots vides", tels que : les pronoms personnels, les prépositions, ..., ou même des mots athématiques qui peuvent se retrouver dans n'importe quel document (par exemple des mots comme contenir, appartenir, etc). L'élimination de ces mots peut se faire en utilisant une liste dressée de mots vides (également appelée anti-dictionnaire), ou en écartant les mots dépassant un certain nombre d'occurrences dans la collection. Bien que ce traitement présente l'avantage de diminuer le nombre de termes d'indexation, il peut cependant induire des effets de silence (par exemple, en éliminant le mot a de vitamine a).

2.2.1.3 La normalisation (lemmatisation ou radicalisation)

C'est un processus morphologique permettant de regrouper les variantes d'un mot. En effet, on peut trouver dans un texte différentes formes d'un mot désignant le même sens. Ils seront représentés par un seul mot désignant le concept véhiculé (ex : écologie, écologiste, écologique ? écologie). On distingue quatre principaux types de lemmatisation : par analyse grammaticale en utilisant un dictionnaire (ex : Tree-tagger ⁵), par utilisation de règles de transformation de type condition action surtout pour l'anglais (ex : l'algorithme de Porter [182]), par troncature des suffixes à X caractères (ex : la troncature à 7 caractères), ou encore par la méthode des n-grammes utilisée pour le chinois et très intéressante pour la radicalisation. Il reste cependant à mentionner que ces traitements peuvent induire certains inconvénients tels que la production de normalisation agressive, par exemple les mots university/universe, organization/organ, policy/police sont normalisés par l'algorithme de Porter. Ou l'oubli de quelques normalisations intéressantes, par exemple : matrices/matrix, Europe/European, machine/machinery ne sont pas normalisés. Il existe des techniques (analyse de corpus) pour réduire ces effets négatifs.

⁵<http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/>

2.2.1.4 La pondération des mots

La pondération est une fonction fondamentale puisqu'elle traduit le degré d'importance des termes dans les documents. Cette importance est souvent calculée sur la base d'aspects statistiques, si l'on ne veut pas citer certaines méthodes qui proposent d'introduire des éléments linguistiques dans l'indexation des documents. Parmi les nombreuses formules de pondération définies dans le domaine, la mesure *TF-IDF* est de loin la plus connue et utilisée. Elle est basée sur la combinaison des deux facteurs : fréquence du terme (*TF*) quantifiant la représentativité locale d'un terme dans le document, et fréquence inverse de document (*IDF*), mesurant la représentativité globale du terme vis-à-vis de la collection des documents. Elle est donnée par la multiplication des deux mesures TF et IDF comme suit :

$$TF * IDF = \log(1 + TF) * IDF \quad (1.1)$$

Les mesures *TF* et *IDF* sont définies comme suit :

1. **TF (Term Frequency)** : cette mesure a été introduite pour tenir compte de la fréquence d'un terme dans un document. L'idée sous-jacente est que plus un terme est fréquent dans un document plus il est important dans sa description. Elle représente une "pondération locale" d'un terme dans un document. On trouve plusieurs variantes de cette mesure. Soit le document d_j et le terme t_i , alors la fréquence TF_{ij} du terme dans le document est donnée selon l'une des formulations suivantes :

$$TF_{ij} = 1 + \log(td_{ij}), TF_{ij} = \frac{td_{ij}}{\sum_k td_{kj}} \quad (1.2)$$

où td_{ij} est le nombre d'occurrences du terme t_i dans le document 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. **IDF (Inverse Document Frequency)** : ce facteur mesure la fréquence d'un terme dans toute la collection, c'est la "pondération globale". En effet, un terme fréquent dans la collection, a moins d'importance qu'un terme moins fréquent. Cette mesure est exprimée selon l'une des déclinaisons suivantes :

$$IDF_i = \log(N/n_i), IDF_i = \log(N - n_i/n_i) \quad (1.3)$$

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

2.2.2 La phase de recherche ou d'appariement document-requête

La phase de recherche implique un processus d'interaction de l'utilisateur avec le SRI. 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 mécanisme d'indexation adopté, (3) la correspondance entre la requête et les documents par exploitation de l'index et l'affichage des résultats de recherche. En particulier, le processus de mise en correspondance ou d'appariement document-requête permet de mesurer la pertinence d'un document vis-à-vis d'une requête, en calculant un score de correspondance entre la représentation de chaque document et celle de la requête. Ce score traduit un degré de pertinence système, que l'on essaye de rapprocher le plus possible du jugement de pertinence de l'utilisateur vis-à-vis du document. Cette valeur est calculée à partir d'une probabilité ou une similarité appelée : $RSV(q, d)$ (Retrieval Status Value), où q est une requête et d un document. Cette mesure tient compte des poids des termes.

Le processus d'appariement est étroitement lié au processus d'indexation et de pondération des termes des requêtes et des documents du corpus. Notons également que d'une façon générale, le modèle de représentation des documents et des requêtes ainsi que l'appariement document-requête, permettent de caractériser et d'identifier un modèle de recherche d'information.

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 flou. Dans la suite, nous présentons les principaux modèles développés dans les travaux dans ce domaine.

2.3 Aperçu des principaux modèles de RI

Un modèle de RI se définit principalement, par sa modélisation de la mesure de la pertinence document-requête, mais aussi par sa représentation des documents et sa représentation des requêtes. Plus formellement, un modèle de RI est défini par le quadruple $[D, Q, F, R(q_i, d_j)]$ [22], où :

- D est un ensemble des représentations des documents (vue logique),
- Q est un ensemble de représentations des besoins informationnels de l'utilisateur ou ses requêtes,
- F représente le cadre de modélisation de la représentation des documents, des requêtes et de leurs relations,
- $R(q_i, d_j)$ est une fonction d'ordonnement qui associe un score de pertinence (un nombre $\in R$, généralement $R : Q \times D \rightarrow [0, 1]$), où : q_i est une représentation de la requête $\in Q$ et d_j est une représentation d'un document $\in D$.

Il existe une panoplie de modèles de RI, la figure 1.2, présente la classification des modèles de RI selon Kuropka [138]. Comme illustré dans cette figure, les modèles de RI peuvent être regroupés selon le type du modèle mathématique utilisé, en trois grandes classes, à savoir :

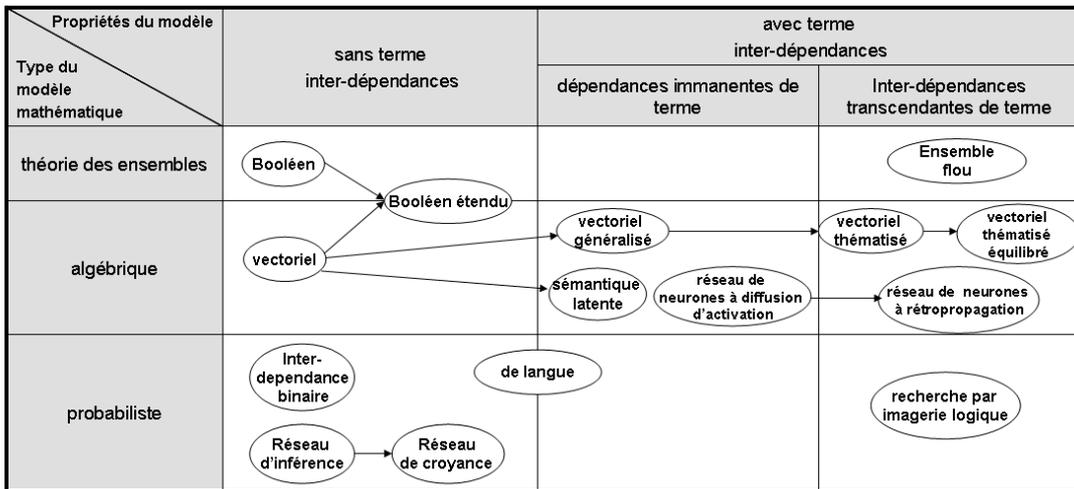


FIG. 1.2 – Classification des modèles de RI selon Kuropka [138]

- **les modèles ensemblistes** : ces modèles trouvent leurs fondements théoriques dans la théorie des ensembles. On distingue le modèle booléen pur (*boolean model*), le modèle booléen étendu (*extended boolean model*) et le modèle basé sur les ensembles flous (*fuzzy set model*).
- **les modèles vectoriels** : basés sur l’algébrique, plus précisément le calcul vectoriel. Ils englobent le modèle vectoriel (*vector model*), le modèle vectoriel généralisé (*generalized vector model*), *Latent Semantic Indexing* (LSI) et le modèle connexionniste.
- **les modèles probabilistes** : se basent sur les probabilités. Ils comprennent le modèle probabiliste général, le modèle de réseau de document ou d’inférence (*Document Network*) et le modèle de langue.

Nous présentons dans la suite les principaux modèles issus de chacune de ces trois classes. Nous référons aux nombreux ouvrages sur la RI [22, 205, 34] pour des présentations exhaustives des modèles de RI.

2.3.1 Modèle booléen

Le modèle booléen [203] est le premier qui s'est imposé dans le monde de la recherche d'information. Il est basé sur la théorie des ensembles et l'algèbre de Bool. Dans ce modèle, chaque document est représenté par une conjonction logique des termes non pondérés qui constitue l'index du document. En effet, le modèle booléen considère que les termes de l'index sont présents ou absents d'un document. En conséquence, les poids des termes dans l'index sont binaires, c'est à dire soit 0 soit 1. Une requête q_i est elle aussi composée de termes liés par les trois connecteurs logiques ET, OU, NON. La fonction d'ordonnement vérifie alors si l'index de chaque document d_j implique l'expression logique de la requête q_i . Le résultat de cette fonction est un score binaire décrit comme suit : $RSV(q_i, d_j) = \{1, 0\}$.

2.3.2 Modèle vectoriel

Initialement proposé par Salton dans le système SMART [204], ces modèles représentent les documents et les requêtes sous forme de vecteurs dans l'espace vectoriel à M dimensions, ces dimensions étant les termes du vocabulaire d'indexation. Chaque document est représenté par un vecteur : $d_j = (w_{1j}, w_{2j}, \dots, w_{Mj})$. De même chaque requête est représentée par un vecteur : $q_i = (w_{1i}, w_{2i}, \dots, w_{Mi})$. Avec : w correspond au poids d'un terme dans le document d_j ou dans la requête q_i . La pondération des composantes de la requête est soit la même que celle utilisée pour les documents, soit donnée par l'utilisateur lors de sa formulation. La pertinence requête-document est traduite par la similarité de leurs vecteurs associés. Le mécanisme de recherche consiste donc à retrouver les vecteurs documents qui s'approchent le plus du vecteur requête. Les principales mesures de similarité utilisées sont :

Le produit scalaire :

$$RSV(q_i, d_j) = \sum_{k=1}^M w_{ki} \cdot w_{kj} \quad (1.4)$$

La mesure de Jaccard :

$$RSV(q_i, d_j) = \frac{\sum_{k=1}^M w_{ki} \cdot w_{kj}}{\sum_{k=1}^M w_{ki}^2 + \sum_{k=1}^M w_{kj}^2 - \sum_{k=1}^M w_{ki} \cdot w_{kj}} \quad (1.5)$$

La mesure cosinus :

$$RSV(q_i, d_j) = \frac{\sum_{k=1}^M w_{ki} \cdot w_{kj}}{\left(\sum_{k=1}^M w_{ki}^2\right)^{1/2} \cdot \left(\sum_{k=1}^M w_{kj}^2\right)^{1/2}} \quad (1.6)$$

2.3.3 Modèle probabiliste

Le modèle probabiliste a été développé dans les années 70, et a connu des développements récents, car les approches basées sur ce modèle ont obtenu de très bons résultats dans TREC, par exemple le modèle BM25 [197].

La fonction de pertinence se base sur le calcul de probabilités de pertinence des documents pour les requêtes données. On distingue deux classes de documents pour une requête q_i : les pertinents (R) et les non pertinents (\bar{R}), ainsi deux mesures de probabilité seront calculées : $P(R|d_j)$ la probabilité que le document d_j soit dans R et $P(\bar{R}|d_j)$ la probabilité que ce document soit dans \bar{R} . 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. Alors la pertinence entre le document d_j et la requête q_i est donnée par :

$$RSV(q_i, d_j) = \frac{P(R|d_j)}{P(\bar{R}|d_j)}$$

En appliquant la règle de Bayes sur la formule précédente cela donnera :

$$RSV(q_i, d_j) = \frac{P(d_j|R)}{P(d_j|\bar{R})}$$

Pour estimer les probabilités $P(d_j|R)$ et $P(d_j|\bar{R})$ un document sera décomposé en un ensemble d'événements $d_j(t_1, t_2, \dots, t_N)$. Chaque événement dénotera la présence ou l'absence d'un terme t_i dans un document d_j . En supposant l'indépendance des termes des documents, la formule précédente devient :

$$RSV(q_i, d_j) = \sum_{i=1}^N \log \frac{P(w_{ij}/R)}{P(w_{ij}/\bar{R})} \quad (1.7)$$

où w_{ij} est le poids du terme t_i dans le document d_j . Le score de pertinence d'un document vis-à-vis d'une requête peut se calculer par différentes formules [194]. Par exemple, dans le modèle BM25 le calcul du poids d'un terme dans un document intègre différents aspects relatifs à la fréquence locale des termes, leur rareté et la longueur des documents :

$$w_{ij} = \log\left(\frac{N - df + 0.5}{df + 0.5}\right) \times \frac{(K_1 + 1) * TF}{K_1 \times ((1 - b) + b \times \frac{dl}{avgdl}) + tf} \quad (1.8)$$

où dl est la taille du document d_j , $avgdl$ est la moyenne des tailles des documents dans la collection et k_1 , b sont des paramètres qui dépendent de la collection ainsi que du type des requêtes.

2.4 Évaluation des performances des systèmes de RI

La validation empirique d'un modèle, d'une technique ou d'un algorithme en RI est une fonction transversale, fondée sur le paradigme de l'évaluation comparative des résultats qu'ils produisent dans des situations de recherche simulées à l'aide de collections de test.

Les premiers protocoles d'évaluation adoptés en RI se basent sur une approche de type laboratoire où (*laboratory-based model*) initié par Cleverdon [55] dans le cadre du projet Cranfield Project II, et qui constitue le cadre de référence dans lequel s'inscrivent les expérimentations et la validation des systèmes. Cette approche fournit des ressources de base pour l'évaluation des SRI : une collection de requêtes, une collection de documents, des jugements de pertinence associés à chaque requête et des mesures d'évaluation basées essentiellement sur le rappel et la précision. Elle est souvent adoptée dans les campagnes d'évaluation des SRI tels que TREC, CLEF, INEX, etc.

2.4.1 Collection de test

La collection ou corpus de test constitue le contexte d'évaluation, c'est-à-dire les éléments qui vont servir à tester le processus de sélection des documents par le SRI. Les corpus diffèrent par le nombre de documents et le nombre de requêtes, mais aussi sur leur domaine de spécialité, la façon de juger la pertinence, etc.

De nombreux projets basés sur les corpus de tests se multiplient depuis les années 1970, on peut citer la collection CACM, la collection CISI, la campagne CLEF ou encore la campagne INEX. La campagne la plus connue est sans conteste TREC (Text REtrieval Conference) organisée annuellement depuis 1992 par la NISI et la DARPA.

Dans TREC, les recherches étaient centrées au départ (de TREC1 à TREC6) sur deux tâches principales : la tâche de routage et la tâche *ad hoc*. La tâche *ad hoc* est constituée d'un ensemble de nouvelles requêtes qui sont lancées sur une collection de documents fixés, et la tâche de routage est composée d'un ensemble de requêtes fixes lancées sur une collection de documents en évolution perpétuelle. Autour de ces tâches, bon nombre de pistes ont été explorées, et de nouvelles tâches sont apparues. On peut par exemple citer des évaluations de recherche de documents écrits dans une autre langue que l'anglais (exemple : espagnol, français, ou encore chinois, arabe) (à partir de 1994), des évaluations de recherche à travers des langages multiples, des évaluations sur de très grands corpus (tâche Terabyte en 2004), ou des évaluations portant sur des aspects plus diversifiés (la tâche interactive depuis 1994, la tâche QA (question-réponse) en

1999 etc), ou encore des évaluations de recherche sur des vidéos (depuis 2001) ou des documents Web (depuis 1999).

Généralement, chaque collection de test est composée : d'une collection de documents, aussi appelée corpus de documents, d'une collection de requêtes ou corpus de requêtes, et des jugements de pertinence des documents par rapport à ces requêtes.

- **Collection de requêtes** : le corpus de requêtes, également appelées "*topics*", simule l'activité de recherche de l'utilisateur. Pour exploiter au mieux les caractéristiques de la collection de documents et avoir une évaluation assez objective, il est important de créer un ensemble de quelques dizaines de requêtes (TREC utilise de 25 à 50 *topics* comme une norme du nombre de requêtes de test), et qui soient adéquates par leur longueur, les thèmes abordés, leur forme, etc. Les requêtes sont généralement artificielles formulées par des assesseurs qui participent à la campagne d'évaluation, mais elle peuvent aussi être de vraies requêtes extraites à partir de log de recherche Web comme c'est le cas pour la tâche Web de Trec.
- **Corpus de documents** : c'est un ensemble de documents sur lesquels les SRI posent des requêtes et récupèrent les documents pertinents. Selon le TLFI⁶ (Trésor de la Langue Française Informatisé), un corpus de documents est défini comme : "le recueil réunissant ou se proposant de réunir, en vue de leur étude scientifique, la totalité des documents disponibles d'un genre donné, par exemple épigraphiques, littéraires, etc.". Il existe de très nombreux ensembles de documents en accès libre, notamment sur le Web : des documents plus ou moins vulgarisés, plus ou moins spécialisés dans un domaine, dans une langue ou une autre, etc. Le choix d'une collection ou autre dépend de la tâche de recherche que l'on veut évaluer, pour garantir une représentativité par rapport à la tâche. De même que la spécification du volume des collections de documents utilisées dans l'évaluation est relativement dépendante de la tâche de recherche impliquée dans le SRI à évaluer, pour garantir une diversité des sujets et du vocabulaire. Les premiers corpus de test développés dans les années 1970 renferment quelques milliers de documents. Les corpus de test plus récents (par exemple, ceux de TREC) contiennent en général plus de 100000 documents (considérés maintenant comme un corpus de taille moyenne), voir des millions de documents (corpus de grande taille). Dans tous les cas, il s'agit de choisir les contenus de manière adéquate. [91], avance "qu'il ne faut pas perdre de vue que c'est l'opération de choix raisonné parmi les composants disponibles qui crée un corpus", surtout pour la recherche d'information. Le travail concernant la sélection des

⁶<http://atilf.atilf.fr/>

documents des corpus est d'ailleurs très déterminant et fait l'objet de nombreuses recherches [88].

- **Jugements de pertinence** : pour la construction d'un corpus de test, les jugements de pertinence constituent la tâche la plus difficile. Les jugements de pertinence indiquent pour chaque document du corpus s'il est pertinent, et parfois même à quel degré il l'est, pour chaque requête. Pour établir ces listes de documents pour toutes les requêtes, les utilisateurs (ou des testeurs simulant des utilisateurs) doivent examiner chaque document de la base de document, et juger s'il est pertinent par rapport à une requête donnée. Dans les programmes d'évaluation tels que TREC, les collections de documents contiennent plus d'un million de documents, ce qui rend impossible le jugement exhaustif de pertinence. Ainsi, dans le cas de grandes collections, les jugements de pertinence sont construits selon la technique de *pooling*, effectuée à partir des 1000 premiers documents retrouvés par les systèmes participants.

2.4.2 Protocole d'évaluation

Le protocole d'évaluation dans le modèle d'évaluation orienté-laboratoire définit une méthodologie rigoureuse et efficace pour comparer plusieurs SRI, stratégies de recherche, ou algorithmes sur une même base, en spécifiant trois composants non indépendants qui sont : le nombre de *topics* utilisés, les mesures d'évaluation utilisées et la différence de performance requise pour considérer qu'une stratégie de recherche est meilleure qu'une autre [41].

L'évaluation de l'efficacité de chaque stratégie de recherche consiste à évaluer la liste des 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 assessseurs. 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 *topics* dans la collection de test.

Les protocoles d'évaluation utilisés se basent sur des métriques, nous présentons dans la suite les mesures les plus courantes dont les plus classiques "Rappel" et "Précision".

2.4.3 Mesures d'évaluation

En RI, la mise au point des modèles passe par une phase expérimentale qui suppose l'utilisation de métriques qui ont pour but de permettre la comparaison des modèles entre eux ou la mise au point de leurs paramètres. Cleverdon

[56] a défini six critères qui peuvent être utilisés pour l'évaluation de la performance d'un SRI : la couverture de l'univers du discours de la collection, le temps de réponse, la présentation des résultats, l'effort requis de l'utilisateur pour retrouver parmi les documents retournés ceux qui répondent à son besoin, le taux de rappel du système, et la précision du système.

Parmi ces critères, la précision et le rappel sont les plus populaires utilisés pour estimer notamment l'efficacité du SRI exprimée par sa capacité à sélectionner tous les documents pertinents et à rejeter tous les documents non pertinents, qui est considérée comme la plus importante du fait qu'elle traduit l'objectif premier des SRI. D'autres mesures d'évaluation d'un SRI existent, les travaux [160, 254, 195] présentent des études des différentes mesures d'évaluation utilisées en RI. Ceci a permis de dégager d'autres mesures de performance relativement importantes.

Nous présentons dans la suite les mesures les plus classiques de "Rappel" et de "Précision", ainsi qu'un ensemble de mesures les plus courantes.

1. **Rappel et précision** : étant donnée une requête q , les documents de la collection peuvent être classifiés en fonction de leur rapport à la requête (pertinents/non pertinents) comme le montre la figure 1.3. On considère $|S|$ le nombre de documents sélectionnés par un système de RI pour la requête q . On considère de plus, le nombre $|P|$ des documents pertinents dans la collection pour cette requête et $|PS|$ le nombre des documents pertinents sélectionnés par le système.

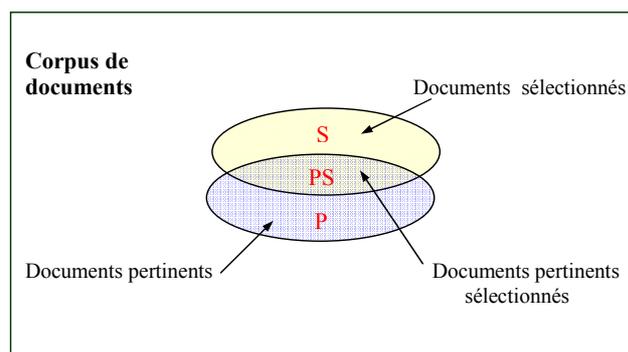


FIG. 1.3 – Rappel et précision.

La mesure de précision calcule alors la capacité du système à rejeter tous les documents non pertinents pour une requête. Elle est donnée par le rapport entre les documents sélectionnés pertinents et l'ensemble des documents sélectionnés :

$$Précision = \frac{|PS|}{|S|}$$

Le rappel mesure la capacité du système à renvoyer tous les documents

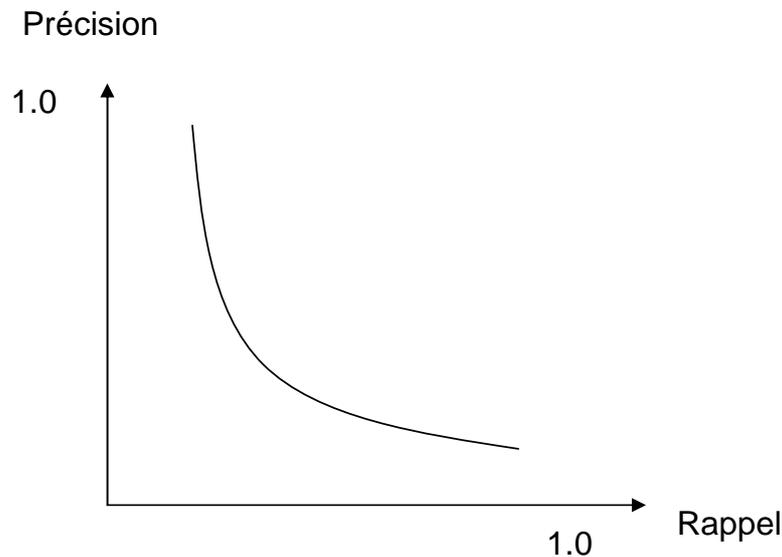


FIG. 1.4 – Forme générale de la courbe de précision-rappel d'un SRI

pertinents pour une requête. Il est donné par le rapport entre les documents pertinents sélectionnés et l'ensemble des documents pertinents pour la requête :

$$Rappel = \frac{|PS|}{|P|}$$

La précision mesurée indépendamment du rappel (et inversement) est peu significative. En pratique, les valeurs du rappel et de précision sont conjointement calculées à chaque document restitué pour les i premiers documents dans la liste des réponses du système. La figure 1.4 représente la forme générale de la courbe de rappel/précision qui en résulte. Ces deux mesures évoluent en sens inverse. Intuitivement, si on augmente le rappel en retrouvant plus de documents pertinents, on diminue la précision en retrouvant aussi plus de documents non pertinents. Inversement, une plus grande précision risque de rejeter des documents pertinents diminuant ainsi le rappel. Le comportement d'un système peut varier en faveur de la précision ou en faveur du rappel au détriment de l'autre métrique.

2. **F-mesure** : c'est une mesure qui combine la précision et le rappel, nommée F-mesure ou F-score introduite dans [250] et définie par :

$$F_{\beta} = \frac{(1 + \beta^2) \cdot (\text{précision} \cdot \text{rappel})}{(\beta^2 \cdot \text{précision} + \text{rappel})}$$

pour des valeurs réelles positives de β traduisant l'importance relative du rappel et de la précision. Un cas particulier de la mesure générale F_{β} est connu comme la mesure F1 ($\beta = 1$), dans ce cas particulier la F-mesure définit la moyenne harmonique du rappel et de la précision :

$$F = \frac{2 \cdot \textit{précision} \cdot \textit{rappel}}{\textit{précision} + \textit{rappel}}$$

3. **précision@X** : c'est la précision à différents niveaux de coupe. Cette précision mesure la proportion des documents pertinents retrouvés parmi les X premiers documents retournés par le système.
4. **R-précision** : 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.
5. **Précision moyenne (Mean Average Precision-MAP)** : c'est la moyenne des précisions moyennes (*Average precision-AP*) obtenues sur l'ensemble des requêtes à chaque fois qu'un document pertinent est retrouvé :

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

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 .

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

6. **La mesure BPREF** : dans le cas de collections volumineuses, la construction de jugements de pertinence complets est difficile voir impossible puisque elle est très coûteuse en terme de temps. Afin de pallier cet inconvénient, Buckley et Voorhees ont proposé la mesure *BPREF* (Binary PReference-based measure) [41]. Cette mesure ne considère que les documents jugés et elle prend en compte les documents pertinents et les documents non pertinents. Elle est donnée par la formule suivante :

$$bpref = \frac{1}{R} \sum_r 1 - \frac{|n \textit{ classés avant } r|}{R}$$

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

7. **Mean Reciprocal Rank (MRR)** : une autre mesure basée rang est la métrique *Mean Reciprocal Rank*. Elle 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 des 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 couramment utilisée dans les systèmes Questions-Réponses où l'utilisateur s'intéresse à recevoir la bonne réponse en premier rang.

3 De la RI classique à la RI adaptative

La recherche d'information classique se base principalement sur l'appariement des descripteurs des documents et des requêtes pour la sélection des informations pertinentes. Or, il est communément admis, dans la communauté en RI, qu'une problématique cruciale dans le domaine est la différence des univers de discours utilisés d'une part des utilisateurs lors de l'expression de leurs besoins en information, et d'autre part, des auteurs de documents (*term mismatch*). Il en résulte ainsi un défaut d'appariement qui engendre une dégradation des performances de recherche. D'autres facteurs viennent accentuer ce problème, notamment : des requêtes utilisateurs courtes, des volumes d'informations importants, des expressions plus ou moins vagues du besoin en information, etc [264].

La première direction des travaux en RI ayant apporté des solutions à ce problème s'inscrit dans le cadre de la RI adaptative. Celle-ci comprend l'ensemble des stratégies et techniques qui permettent de reformuler la requête dans le but de l'adapter au besoin précis de l'utilisateur en termes des documents pertinents associés. La RI adaptative tente alors d'exploiter des informations additionnelles, au-delà de la requête, généralement extraites des interactions de l'utilisateur avec le SRI, dans le but d'améliorer la recherche. On peut distinguer trois principales classes de techniques développées en RI adaptative : (1) les techniques de re-formulation de requête [198], les techniques de désambiguïsation du sens des mots de la requête [220] ou les techniques de regroupement thématique des résultats de recherche Grouper [272].

3.1 Re-formulation des requêtes

Afin de rapprocher au mieux la pertinence système de la pertinence utilisateur, quelques systèmes incorporent une étape supplémentaire de re-formulation de la requête. La re-formulation de la requête consiste à construire une nouvelle requête par ajout, suppression et/ou re-pondération des termes pour mieux représenter les besoins de l'utilisateur. On peut distinguer deux types de re-formulation :

1. **Re-formulation automatique** : également dite pseudo ré-injection de pertinence. Dans ce cas, l'utilisateur n'intervient pas, l'extension de la requête est effectuée en exploitant les premiers documents retournés par le SRI en réponse à la requête initiale (*blind feedback*) [198, 163] ou à partir d'une ressource externe qui peut être un thesaurus, une ontologie, etc [129, 19].
2. **Re-formulation manuelle** : il s'agit de la stratégie de re-formulation de la requête la plus populaire [65, 97]. On la nomme communément ré-injection de la pertinence ou *relevance feedback*. Dans un cycle de ré-injection de pertinence, on présente à l'utilisateur une liste de documents sélectionnés par le système comme réponse à la requête initiale. Après les avoir examinés, l'utilisateur indique ceux qu'il considère pertinents. L'idée principale de la ré-injection de pertinence est de sélectionner les termes importants appartenant aux documents jugés pertinents par l'utilisateur, et de renforcer l'importance de ces termes dans la nouvelle formulation de la requête.

3.2 Désambiguïsation du sens des mots de la requête

Cette technique aide l'utilisateur à mieux exprimer son besoin en information relatif à son intention de recherche. Elle lui permet de saisir le vrai sens évoqué par les termes de sa requête. L'adaptation du processus de recherche à des critères spécifiques de l'utilisateur par désambiguïsation du besoin en information derrière la requête est adoptée dans plusieurs approches dans le domaine de la RI adaptative. La plupart de ces techniques se basent sur l'exploitation des interfaces de clarification interactives à base d'ontologie [220]. D'autres approches se basent sur la définition des paramètres mesurables à partir de la requête ou à partir du profil des premiers documents retournés par la requête. Dans ce sens, un nouveau type de ré-injection 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 [199].

3.3 Regroupement thématique des résultats de recherche

Ces techniques visent à aider l'utilisateur à choisir l'information en faisant un regroupement thématique des résultats de recherche. Ces techniques permettent de grouper les documents similaires ensemble afin d'avoir une accessibilité et une navigation plus simple. Des techniques de regroupement des résultats de recherche dans des catégories ([272], Vivisimo, Kartoo⁷, Exalead, etc.) et des approches de classification du web dans des taxonomies de concepts ([214], ODP⁸, Google directory, etc.) sont développées pour faciliter la navigation.

3.4 Limites de la RI adaptative

Les travaux en RI adaptative ont certes apporté des gains de performance de recherche en apportant des solutions au problème du défaut d'appariement requête-document. Cependant une analyse fine des résultats d'expérimentations rapportées dans la littérature [232], montre que ces performances dépendent d'autres facteurs a priori non contrôlés de manière inhérente au processus de réécriture adaptative de la requête. Ces facteurs peuvent être liés : à l'utilisateur, à l'information portée par la requête et/ou le document, et à l'interaction entre l'utilisateur et le SRI. Il est clair que ces facteurs sont dépendants les uns des autres et qu'il est impératif de les intégrer dans le processus de RI dans le but d'améliorer les performances de la recherche. Ces points constituent précisément une partie des enjeux fixés dans le domaine de la RI orientée utilisateur, que nous présentons dans la section suivante.

4 Vers la RI orientée utilisateur

La fin des années 70 a connu la naissance d'un nouveau courant de recherche, matérialisé par "l'approche utilisateur" et "l'approche cognitive" de la RI, mettant l'utilisateur au centre de l'activité de recherche d'information. Les premiers travaux de ([38], [120], [102]) ont porté sur l'intégration de la perspective cognitive de l'utilisateur dans l'interprétation du concept information et dans l'interprétation du besoin en information dans le cadre d'une tâche ou d'une situation de recherche précise ; ces travaux s'intéressent également à l'appréciation de l'interdépendance des éléments de l'environnement de l'utilisateur et leur impact sur sa perception de la pertinence [25]. La RI orientée

⁷<http://www.kartoo.com/index.php3>

⁸<http://www.dmoz.org>

utilisateur marque le début de trois révolutions importantes dans le domaine de la RI, elles sont nommées explicitement dans [196] :

1. **la révolution cognitive** : cette révolution remet en cause l'isolation du besoin en information de l'acteur l'ayant exprimé. En ce sens que l'interprétation du besoin en information est subordonnée à la dimension cognitive de l'utilisateur qui l'a exprimé (centres d'intérêt, expertise, objectifs, tâche etc.) et de la situation de recherche correspondante,
2. **la révolution de la notion de pertinence** : cette révolution remet en cause l'aspect statique et dichotomique de la pertinence, qui est le concept clé en RI. La pertinence est, sous l'angle de la RI orientée utilisateur, un concept multidimensionnel couvrant divers niveaux (très pertinent, peu pertinent, marginalement pertinent et non pertinent) et types (pertinence situationnelle, pertinence affective, pertinence cognitive) [179], [180]
3. **la révolution de l'interaction** : cette révolution remet en cause la séparation entre l'utilisateur et le SRI en ce sens que l'interaction utilisateur-SRI est une dimension non négligeable, devant être clairement intégrée dans la conception globale du processus de recherche d'information.

Dans [230] un premier bilan des résultats fondamentaux issus de la RI cognitive/utilisateur est dressé, ils y sont notamment mentionnés les résultats suivants :

- le développement de modèles cognitifs de l'interaction en recherche d'information ([210], [248]),
- le développement de méthodes pour la multi-représentation de documents et requêtes représentant les différentes visions cognitives de leur contenu ([101], [141], [233]),
- le développement de modèles et stratégies de filtrage d'information ([26]) et de recommandation ([150], [178]),
- le développement de modèles et de stratégies d'accès contextuel/personnalisé à l'information ([156], [157]),
- le développement de modèles d'évaluation basé sur la pertinence situationnelle dans le contexte d'une recherche d'information interactive ([196]).

Un des résultats les plus importants issus de cette perspective d'adaptation du processus de RI à l'utilisateur et qui se trouve au cœur de nos travaux, est "la RI contextuelle". C'est une direction de recherches qui vise particulièrement la prise en compte du contexte de recherche, comme une nouvelle dimension, permettant à terme de déployer un appariement requête-contexte-document.

Dans la section suivante, nous nous focaliserons sur cette nouvelle branche de la RI qu'est la RI contextuelle. Nous essayerons de définir la notion de contexte dans le cadre particulier de la RI, et de donner un aperçu sur les problématiques et origines de la RI contextuelle ainsi que sur les spécificités d'un SRI contextuel.

5 Émergence de la RI contextuelle

L'objectif principal 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 à ses besoins en information. 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 need"*.

Le contexte est une notion complexe et difficilement formalisable, ayant de multiples définitions couvrant les différentes facettes qu'elle comprend, selon le cadre d'application associé [23]. On la trouve exploitée dans les domaines tels que : la modélisation utilisateur, l'intelligence artificielle, l'hypermédia adaptatif, l'informatique ubiquitaire, les bases de données ([211, 201, 71, 105, 134]).

Dans cette section, nous présentons en premier lieu l'essence des motivations des travaux vers la RI contextuelle. Dans la section 5.2, nous présentons un aperçu des verrous scientifiques et technologiques qui ont marqué les travaux de recherche en RI contextuelle. Nous essayons dans la section 5.3 de faire la lumière sur la notion de contexte dans le cadre de la RI. Nous décrivons dans la section 5.4 l'architecture générale d'un système d'accès contextuel à l'information, et nous terminons une revue des principaux modèles de contexte proposés dans la littérature.

5.1 Facteurs d'émergence de la RI contextuelle

On peut expliquer l'orientation des travaux de recherche vers la RI contextuelle par un ensemble de facteurs qui sont principalement liés : à la prolifération et l'hétérogénéité des ressources d'information, la diversité et l'ambiguïté des besoins en information des utilisateurs ainsi qu'à la diversité des environnements de recherche.

1. Accroissement du volume d'information

Dans le domaine général de la RI, la dimension du problème et le contexte d'utilisation ont changé dans des proportions considérables depuis quelques années. En effet, la production de données sous forme électronique ne cesse de croître, les documents électroniques contiennent de plus en plus souvent des informations hétérogènes (données structurées, documents textuels, composants logiciels, images, mais aussi audio et vidéo qui tendent à se généraliser), des méta-données sont générées et associées aux données de base afin d'en faciliter les accès ultérieurs, et les utilisateurs accèdent à des sources de plus en plus vastes et disséminées, le cas extrême étant le web (selon une étude statistique couvrant plusieurs moteurs de recherche, la taille du Web est estimée au minimum à 14,65 billion de pages web à la date du mois de Mai 2011⁹). L'accès à une information pertinente dans cet environnement qualifié par la prolifération des ressources constitue donc un vrai défi pour la RI classique. 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, la sélection de l'information consiste à considérer que tout document contenant les mots de la requête (dans n'importe quel ordre) est potentiellement pertinent. Cette liste des résultats retournés par le système est de plus en plus importante que le volume des ressources disponibles est important. Par conséquent, l'utilisateur se trouve face à une surcharge informationnelle, dans laquelle il lui incombe la charge de distinguer ce qui est pertinent de ce qui ne l'est pas.

2. Diversité et non expressivité des besoins en information

Les SRI classiques sont conçus pour servir le plus large public possible, tout on ne considérant que la requête dans la représentation du besoin en information de l'utilisateur. Dans ce cas, les SRI renvoient à une même requête soumise par des utilisateurs différents, la même liste de résultats qui correspond à divers centres d'intérêts ayant des estimations de pertinence comparables. De plus, les requêtes utilisateurs sont courtes et/ou

⁹<http://www.worldwidewebsize.com/>

ambigües, la longueur moyenne des requêtes des utilisateurs sur le *Web* est estimée à quelques mots (environ 3 mots d'après [274, 110]). Ce qui accentue la problématique de l'expressivité du besoin en information. Les études [42] ont montré que l'approche généraliste des outils disponibles en RI est à l'origine des problématiques évoquées des SRI classiques et leur incapacité à discriminer les utilisateurs en fonction de leurs centres d'intérêts ou de leurs préférences de recherche.

3. Diversité des environnements de recherche

L'émergence de nouveaux environnements de recherche tels que : la RI mobile (la téléphonie mobile, les PDA, etc), la RI social (réseaux sociaux, les agendas personnels, etc) et bien d'autres application (guides touristiques, la RI médicale, etc), font du contexte de l'utilisateur une source d'information importante permettant d'améliorer la précision de la recherche [256, 89]. En effet, ces nouveaux environnements de recherche imposent des contraintes spécifiques de recherche, qui peuvent être : géographiques généralement liées à la localisation de l'utilisateur, temporelles généralement liées à l'instant de soumission de la requête, matérielles généralement liées aux dispositifs d'accès à l'information dont dispose l'utilisateur, etc.

Pour surmonter ces problèmes et pouvoir adapter les résultats de la recherche aux besoins spécifiques d'un utilisateur donné dans un contexte d'utilisation particulier, les travaux en RI contextuelle ont vu le jour. Ils mettent respectivement en avant les aspects liés au contexte, aux échanges d'information entre l'utilisateur, son environnement et le système. Et proposent l'adaptation de la recherche à l'utilisateur principalement, mais aussi aux ressources, au système ou à l'environnement.

5.2 Verrous scientifiques et technologiques

En tant que nouveau domaine de recherche, l'accès contextuel à l'information pose des problématiques nouvelles allant de la modélisation du contexte jusqu'à la modélisation de la pertinence cognitive en passant par la modélisation de l'interaction entre un utilisateur et un système de recherche d'information. Nous synthétisons dans ce qui suit les principaux verrous tels qu'ils sont identifiés dans [230], ils peuvent être projetés sur trois niveaux : conception, évaluation et mise en œuvre.

5.2.1 Conception

Le contexte constitue une nouvelle dimension à modéliser puis décliner dans le modèle d'accès à l'information. Ceci pose alors des problèmes liés à la définition, la formalisation, la mise en relation et exploitation conjointe de différents éléments potentiels : centres d'intérêts et préférences de l'utilisateur, son but, sa tâche, l'expression de sa requête, sa perception de la pertinence, ses interactions, etc. Les recherches actuelles s'orientent vers la définition de modèles de contexte permettant à même de préciser les résultats d'une activité de recherche d'information. Les problèmes posés s'affinent précisément vers : (1) l'identification de sources d'informations appropriées traduisant la perception de l'utilisateur de l'information pertinente, (2) la représentation des informations jugées et manipulées par l'utilisateur, (3) la maintenance et l'évolution du contexte à travers les interactions de l'utilisateur avec le SRI. Ce verrou est loin d'être levé comme en témoigne les nombreuses manifestations scientifiques du domaine telles que les conférences *Information, Interaction in ConteXt* (IiX), *Information Retrieval In conteXt* (IRIX) et le workshop *Context based Information retrieval* annexé à la conférence *User modelling*.

5.2.2 Évaluation

L'avènement de la RI contextuelle a mis en exergue les limites des protocoles d'évaluation à la Cranfield et des mesures sous-jacentes. Les principales critiques qui lui sont reprochées [93], sont : 1) qu'elle suppose que l'utilisateur émet sa requête dans une session unique, le système est donc évalué en mode *batch*, 2) qu'elle ne tient pas compte du profil de l'utilisateur, et 3) que les jugements de pertinence sont binaires (pertinent/non pertinent) et portent uniquement sur la pertinence thématique. Les méthodologies d'évaluation classiques des systèmes d'accès à l'information devraient donc être révisées dans le sens de l'intégration du contexte. Cette révision portera essentiellement sur la définition de nouveaux protocoles d'évaluation basés sur des utilisateurs spécifiques, de nouvelles métriques d'évaluation tenant compte de différents niveaux de pertinence, de nouvelles collections de tests décrites par des méta-données descriptives des contextes de recherche. C'est particulièrement l'objet de *workshops* très récents tels que CIRSE (2009 et 2010), "*future of IR evaluation*" dans le cadre de SIGIR'2010, et d'autres *workshops* dans ECIR'2011.

5.2.3 Mise en œuvre

L'exploitation de systèmes d'accès personnalisés pose un problème fondamental, d'ordre technologique, qui porte sur la protection de la vie privée. En

effet, la définition, l'utilisation et la dissémination des profils constituent à la fois un atout et une contrainte pour assurer la portée des systèmes qui les supportent. C'est tout d'abord un atout dans le sens où les différents profils sont maintenus et sont accessibles, pouvant donc contribuer à mettre en œuvre une recherche d'information collective et dynamique par l'introduction de techniques d'apprentissage. C'est ensuite une contrainte dans le sens où elle doit être impérativement soutenue par une réflexion sur les droits des personnes pour assurer la sécurité globale des profils. L'objectif fondamental est donc de trouver le compromis entre la personnalisation en tant que service permettant de fidéliser les utilisateurs et la protection de leur vie privée, relativement à une diffusion élargie de leurs informations personnelles pour des fins non explicitement approuvées.

5.3 Notion de contexte en RI : définition et taxonomie

5.3.1 Définition du contexte en RI

Les premières définitions de la notion de contexte en RI remontent aux travaux de Ingerwersen [102] et de Saracevic [209] qui ont placé le contexte en amont de l'interaction utilisateur-SRI. Le contexte y est défini comme l'ensemble des facteurs cognitifs et sociaux ainsi que les buts et intentions de l'utilisateur au cours d'une session de recherche.

Par la suite, de nombreux travaux ont identifié des niveaux de description du contexte selon un modèle basé sur des infrastructures incluant de larges classes contextuelles. Plusieurs définitions du contexte sont alors proposées dans la littérature, elle diffèrent essentiellement par ses éléments constitutifs et sa portée à court ou à long terme.

Wilson [261] place l'utilisateur au centre du processus de recherche d'information et identifie son contexte de recherche en analysant les pratiques de recherche d'information (*information seeking behaviour*) et les besoins d'information de l'utilisateur en fonction des besoins personnels de celui-ci (physiologiques, affectifs, cognitifs), de son rôle social (travail, performance) et de son environnement (de travail, socio-culturel, politicoéconomique et physique).

D'autres travaux [17, 224] précisent que le contexte couvre des aspects larges tels que 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, ou alors selon les familles de facteurs caractéristiques tels que le niveau environnement, le niveau utilisateur et le niveau interaction [61].

C'est le sens générique du contexte qui a été largement exploré cette dernière décennie en RI contextuelle [103].

5.3.2 Taxonomie du contexte

La notion de contexte en RI peut être reliée à plusieurs entités intervenantes dans le processus de recherche d'information. On distingue principalement trois entités qui interviennent dans le processus de RI :

- *L'utilisateur* : il peut être caractérisé par ses connaissances, ses buts et ses intentions de recherche d'information.
- *Le système* : il se lie aux caractéristiques relatives à lui même tel que : le temps de réponse, le coût, etc.
- *L'environnement de recherche* : il présente des caractéristiques liées à des critères sociaux, organisationnels et situationnels.

Pour mieux comprendre les facteurs contextuels qui sont nécessaires en vue d'une meilleure application du contexte dans les systèmes de RI, diverses taxonomies du contexte ont été proposées dans la littérature [83, 60, 104, 231]. Ci-dessous, nous détaillons la taxonomie présentée dans [231], qui représente une synthèse des dimensions du contexte abordées ou explorées en RI contextuelle (illustrée dans la figure 1.5). Elle comprend cinq dimensions principales : dispositif d'accès à l'information, contexte spatio-temporel, contexte de l'utilisateur, contexte de la tâche et contexte du document.

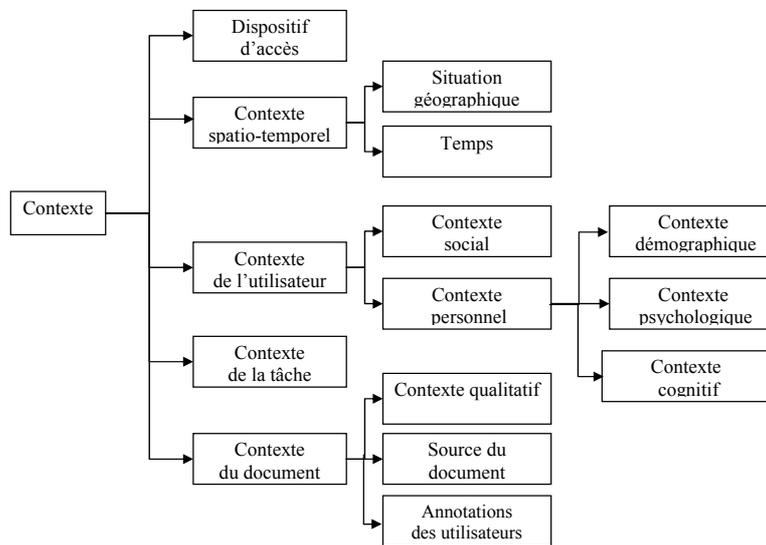


FIG. 1.5 – Le concept du contexte multidimensionnel en RI

1. **Dispositif d'accès à l'information** : cette dimension regroupe les aspects techniques des dispositifs d'accès à l'information. En effet, l'environnement matériel de recherche est devenu une composante importante du fait de l'apparition de nouveaux environnements de recherche, autres les PCs de bureau, tels que : les agendas personnels, les téléphones mobiles, les PDA, etc. Ce contexte définit alors les caractéristiques physiques du processus d'interaction liées aux capacités des moyens matériels dont

dispose l'utilisateur : le terminal disponible, l'état du réseau auquel l'utilisateur est connecté (bande passante, débit, etc.). L'adaptation du processus de RI aux caractéristiques de l'outil physique d'accès s'est imposée particulièrement pour des utilisateurs mobiles présentant des contraintes physiques (ressources mémoires limitées, zone d'affichage des résultats réduite) [85], ayant un impact potentiel sur la perception de pertinence de l'utilisateur vis-à-vis de l'information retournée.

2. **Contexte spatio-temporel** : cette dimension comprend l'ensemble des attributs caractéristiques des deux sous-dimensions de la localisation spatiale et du temps, pouvant caractériser la situation de l'utilisateur. L'adaptation selon cette dimension concerne particulièrement les applications où les informations ont une pertinence dépendante de la localisation et du temps instanciés lors de l'activité de recherche de l'information (par exemple : le routage, l'actualité et la météo, les guides touristiques, etc.) [238, 86].
3. **Contexte de l'utilisateur** : cette dimension permet de modéliser et de stocker les données caractérisant l'utilisateur. Elle est la dimension du contexte la plus explorée en RI contextuelle. Elle définit à elle seule un sous-type de la RI contextuelle, appelée : "RI personnalisée". Où l'accent est mis souvent sur l'utilisation d'un modèle de l'utilisateur (profil) préalablement construit [152, 106]. Le contexte de l'utilisateur couvre deux sous-dimensions : le contexte personnel de l'utilisateur et son environnement social.
 - **Contexte personnel** : comprend à son tour les sous-dimensions suivantes :
 - *Contexte démographique* : fait référence à des facteurs de préférences personnelles tels que la langue (Google personalized, Yahoo) et le genre [98, 81], exploitées de manière à personnaliser la recherche pour des besoins spécifiques.
 - *Contexte psychologique* : regroupe des facteurs psychologiques tels que l'anxiété et la frustration qui sont considérés comme pouvant avoir un impact sur le comportement de l'utilisateur notamment son jugement de pertinence [30, 131].
 - *Contexte cognitif* : cette sous-dimension porte sur l'expertise [243] et les centres d'intérêt de l'utilisateur à court terme reflétant un domaine d'intérêt spécifique relatif à une session de recherche [95, 216, 14, 115, 69] ou ses centres d'intérêt à long terme désignant les domaines d'intérêt généraux de l'utilisateur [152, 218, 236].
 - **Contexte social** : cette dimension concerne les informations sur la communauté à laquelle appartient l'utilisateur telle que les amis, les voisins et les collègues. L'adaptation du processus de recherche d'infor-

mation consiste essentiellement à considérer les préférences et les profils partagés par la communauté de l'utilisateur plutôt que ses préférences et profil personnels [140, 222].

4. **Contexte de la tâche** : cette dimension signifie l'ensemble des activités de recherche liées à un même besoin en information ou but de l'utilisateur. Le contexte de la tâche est introduit du fait que la nature de la tâche influence le comportement de recherche de l'utilisateur [149], traduit par sa perception de l'information pertinente [260, 36]. A ce jour il n'existe pas de nomenclature reconnue ou de définition générique applicable aux tâches. A la place, on trouvera une multitude de travaux approfondissant une ou plusieurs facettes des tâches de recherche d'information. Des travaux tentent de définir des classes d'intention telles que la recherche ciblée (*fact-finding*) vs. de la recherche exploratoire (*exploration task*) [168] ou alors la tâche transactionnelle, la tâche informationnelle ou la tâche navigationnelle, évoquant particulièrement une recherche d'information sur le web [109]. Dans [68] les auteurs se sont focalisés sur la capture de l'intention commerciale derrière la recherche sur le web, et dans [89, 258, 247] les auteurs se sont intéressés à la classification des requêtes selon des intentions de recherche locales ou globales.

D'autres travaux proposent plutôt une classification de la nature des tâches dans des domaines d'intérêts. On peut alors trouver des types de tâches de recherche tels que : (1) *Academic Research*, (2) *News and Weather*, (3) *Shopping and Selling*. L'approche proposée dans [149] 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 tâche et la perception de l'utilisateur dans la tâche.

5. **Contexte du document** : le contexte du document trouve son origine dans le principe de polyreprésentation développé dans les travaux de Ingwersen [101]. Le principe de polyreprésentation de l'information s'appuie sur l'hypothèse que les informations sont classifiables selon un ensemble de critères (considérés comme des dimensions de l'espace de représentation de l'information) tels que : le genre de l'information, son auteur, sa structure, son style etc. et que la considération de l'ensemble de ces critères permet de mieux subordonner la pertinence de l'information à des situations/contextes de recherche spécifiques. Trois sous-dimensions peuvent être identifiées sous cette dimension. La première porte sur le contexte direct de l'information (forme, couleur, métadonnées de structure etc.) [245]. La deuxième correspond à toutes les informations qui peuvent être rajoutées par les utilisateurs des documents (*user annotated data*) [151, 267]. La troisième concerne les caractéristiques de la source d'information et sa perception par les utilisateurs [263], telles que :

sa crédibilité, sa fiabilité, sa fraîcheur, sa cohérence, sa complétude, sa sécurité, 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 l'application de recherche (web, locale, médicale, mobile, entreprise, etc). Dans ce qui suit, nous présentons la plateforme de base d'un SRI supportant la notion de contexte puis nous mettons en évidence sa déclinaison dans le processus d'accès à l'information.

5.4 Architecture d'un système de RI contextuel

Un SRI est dit contextuel ou sensible au contexte (context-aware) s'il exploite les données du contexte de recherche pour sélectionner l'information pertinente en réponse à une requête utilisateur. La pertinence de l'information retournée à l'utilisateur est alors dépendante de son adéquation à la requête et de plus aux dimensions du contexte (définies ci-haut) qui sont perceptibles dans la situation de recherche en cours. La figure 1.6 présente l'architecture de base d'un SRI contextuel selon [230]. On peut y distinguer deux fonctionnalités fondamentales : la modélisation du contexte et l'accès contextuel à l'information.

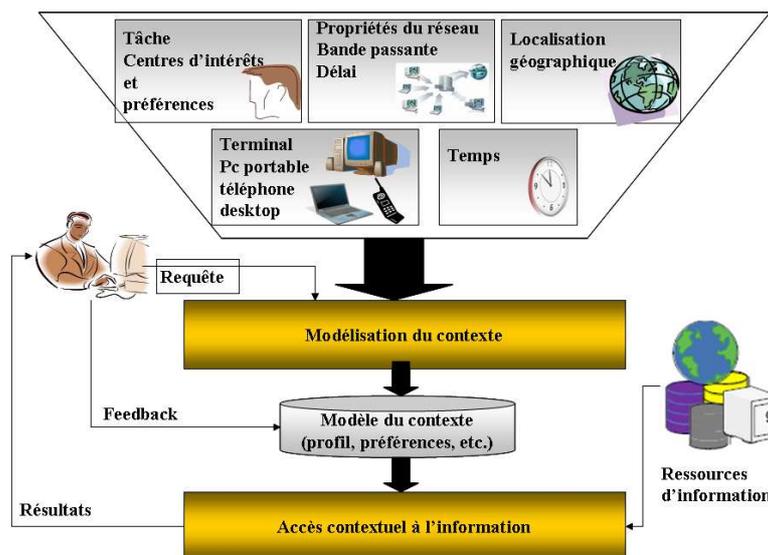


FIG. 1.6 – Architecture de base d'un SRI contextuel

Nous donnons dans la suite de cette section, un aperçu des principales techniques de modélisation du contexte utilisateur ainsi que les modèles d'accès

contextuel à l'information. Nous notons que la majorité des travaux dans le domaine [156, 152, 84, 219, 241] ont abordé la question critique portant sur la modélisation et l'exploitation du contexte cognitif de l'utilisateur. C'est la dimension du contexte qui est largement étudiées et considérées la plus importante dans l'amélioration de la précision de la recherche informationnelle.

5.4.1 La modélisation du contexte

Par opposition à la RI orientée-système qui s'appuie sur la requête comme unique source d'évidence à modéliser, permettant de spécifier le besoin en information, la RI contextuelle s'appuie sur une source d'évidence additionnelle exprimée à travers le contexte qu'il convient alors de modéliser. La nature et la portée du modèle dépendent des dimensions considérées du contexte. De manière générale, un modèle de contexte est défini par instanciation de chacun de ces éléments : les sources d'information exploitées, les stratégies de collecte de ces informations, les ressources de modélisation utilisées, et les modèles de représentation et d'évolution adoptés.

1. Les sources d'information

Peuvent être de différents types : documents visités, historique des interactions, environnement (temps, température etc.), etc. Une revue de la littérature montre que les sources les plus utilisées sont les suivantes :

- Le comportement de l'utilisateur perçu à l'aide d'indicateurs d'évaluation implicite [77] tels que l'historique des *clicks*, les données de navigation [241, 216] et le mouvement des yeux [115].
- Les pages et sites favoris [156].
- Des informations locales [74] et contextuelles telles que les sources accédées comme les journaux (*Reuters*, *New York Times*), les *Blog* sites et les sites de e-commerce [42].
- Les premières pages ou résumés de pages web retournés par un moteur de recherche [89, 68]

2. Les stratégies de collecte de ces informations

On peut distinguer principalement deux stratégies de collecte des données du contexte, les stratégies explicites et les stratégies implicites.

- *L'acquisition explicite* : repose principalement sur les techniques de *feedback* explicite largement utilisées dans de reformulation de requêtes par ré-injection de pertinence. L'approche dans [220] 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 re-formulation personnalisée de la requête. Dans [152] on 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 [105] 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.

Ces techniques d'acquisition explicite permettent une construction contrôlée du profil utilisateur. Cependant, elles présentent des limites à cause de l'effort supplémentaire imposé à l'utilisateur à spécifier explicitement ses besoins. Ces limitations ont orienté les travaux vers des techniques d'acquisition implicite des données du contexte utilisateur.

- *L'acquisition implicite* : consiste à collecter à l'aide d'algorithmes d'acquisition implicite les données de l'utilisateur en observant ses interactions avec le système durant les activités de recherche [237, 84, 74, 217]. L'avantage de cette approche est qu'elle ne nécessite aucune implication directe de l'utilisateur. Le répertoire d'information collectées constitue un riche répertoire de données qui peut éventuellement être exploité selon diverses techniques. 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 [76, 162]. D'autres études utilisent des machines d'apprentissage automatique [257] des données utilisateurs dans le but d'inférer son profil et améliorer ainsi l'efficacité de la recherche.

3. Les ressources de modélisation

Les ressources sémantiques (e.g. ontologies, dictionnaires, ...), sont parfois exploitées pour enrichir les données du modèle. Nous citons la hiérarchie de concepts de "Yahoo" [220], les portails en ligne tels que Magellan et Lycos [84] ou alors celle de l'ODP [84, 219, 69] comme les sources d'évidence le plus souvent utilisées. Certains travaux [249] exploitent des ontologies de domaine.

4. Les modèles de représentation

Permet de formaliser la représentation du contexte en qualité de structure unifiée (partie d'une ontologie, classe de vecteurs de termes, ensemble de concepts, ...) ou d'un ensemble d'informations avec des structures différentes et spécifiques, puis de les faire évoluer au cours du temps. Parmi les modèles de représentation proposés dans la littérature on peut distinguer :

- *Les représentations basées historique de recherche* : consistent en l'ensemble des requêtes et des pages web précédemment visitées ou cliquées de l'utilisateur (*Google Psearch*¹⁰, [189]) ou l'ensemble des requêtes et

¹⁰<http://www.google.com/psearch>

les résumés textuels de ses résultats associés [216] accumulés au cours des sessions de recherche de l'utilisateur.

- *Les représentations ensemblistes* : se basent sur un ensemble de mots clés (ou vecteurs de termes) pondérés représentés souvent selon le modèle vectoriel de Salton [206]. Les paquets de termes représentent généralement les centres d'intérêts de l'utilisateur. Nous pouvons distinguer entre les représentations ensemblistes qui utilisent un vecteur de termes pondérés représentant un centre d'intérêt [150, 234] et celles qui utilisent des classes de vecteurs de termes pondérés dont chacun représente un centre d'intérêt [156, 220].
- *Les représentations connexionnistes* : consistent non seulement à extraire des termes à partir des documents pertinents de l'utilisateur, mais à intégrer ces termes dans un réseau de nœuds pondérés. Cette représentation permet de résoudre les failles de la représentation ensembliste par la mise en place des relations de corrélation sémantiques entre les mots du vocabulaire utilisé. En effet, la richesse sémantique dans cette représentation permet de résoudre le problème de la polysémie des termes. Ce type de représentation est adopté dans [158, 24] par construction d'un réseau de termes reliés par des arcs, ou par des connecteurs logiques [105] selon des relations de co-occurrence.
- *Les représentations conceptuelles* : se basent sur l'exploitation des ontologies de domaines ou des hiérarchies de concepts préalablement définies. L'approche de représentation conceptuelle 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. A la fin le contexte utilisateur sera présenté par un réseau de nœuds conceptuels reliés entre eux en respectant la topologie des liens définis dans les hiérarchies ou les ontologies utilisées. Nous pouvons distinguer entre les représentations conceptuelles sous forme de hiérarchie de concepts [130], une partie d'une ontologie de référence [84, 219], des matrices de concepts [152] ou encore des graphes de concepts [70].

5.4.2 L'accès contextuel à l'information

Le processus classique de RI est projeté selon une dimension additionnelle liée au contexte de recherche. L'objectif est alors de sélectionner l'information pertinente à la requête adressée au SRI, en tenant compte de la requête de l'utilisateur d'une part et du contexte de recherche en cours d'autre part. Le

contexte peut être exploité à différentes phases du processus de RI : dans la formulation de la requête, dans la fonction de pertinence, dans l'ordonnement des résultats de recherche.

5.4.2.1 Reformulation de la requête

Les éléments du contexte peuvent être utilisés pour reformuler une requête. La reformulation de requête consiste à augmenter la requête avec des informations du contexte avant de lancer le processus d'appariement.

Dans [220] la reformulation de requêtes est faite en intégrant les termes représentant le contexte de l'utilisateur dans l'algorithme de Rocchio comme suit :

$$Q2 = \alpha.Q1 + \beta. \sum T_{sel} - \gamma T_{desel} \quad (1.9)$$

Dans cette formule, T_{rel} est un vecteur de termes représentant le concept pertinent à la requête et T_{Nrel} est un vecteur de termes représentant le concept non pertinent sélectionné 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.

Une autre approche de reformulation de la requête [135] exploite un profil connexionniste. Le processus de reformulation de la requête génère une nouvelle requête en appliquant toutes les écritures possibles définies par les arcs pondérés du profil et satisfaisants un critère du seuil de corrélation fixé.

Le processus de reformulation décrit dans [216] consiste à repondérer les termes de la requête selon le contexte représentant l'historique de recherche de l'utilisateur. Plus particulièrement, l'historique de recherche est composé de l'historique de requêtes défini selon un modèle de langage statistique $p(w/H_Q)$, l'historique de clics défini selon un autre modèle de langage $p(w/H_c)$, et le modèle de la requête défini par $p(w/Q_k)$. Ainsi, la reformulation de la requête est effectuée selon le modèle contextuel de la requête $p(w/\theta_k)$ défini par la combinaison interpolée de ces trois modèles de langages comme suit :

$$p(w/\theta_k) = \alpha * p(w/Q_k) + (1 - \alpha) * p(w/H_Q) + \beta * p(w/H_c) \quad (1.10)$$

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

Dans [259] l'approche proposée pour l'expansion des termes de la requête est basée sur des relations de corrélation entre les termes de la requête, les éléments contextuels et les termes des documents. Les relations de corrélation sont combinées selon quatre modèles probabilistes différents : un modèle orienté-contexte, un modèle indépendant requête-contexte, un modèle dépendant requête-

contexte et un modèle filtrant le 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 l'un des modèles probabilistes proposés. Enfin, les termes les mieux pondérés sont alors utilisés pour reformuler la requête de l'utilisateur.

5.4.2.2 Fonction d'appariement

Le contexte peut également intervenir dans la définition de la fonction de pertinence. Le calcul du score du document est alors une fonction qui assigne au document un score de pertinence en fonction non seulement de la requête mais aussi du contexte utilisateur.

Jeh et Widom [112] ont proposé une variante personnalisée de l'algorithme PageRank en l'occurrence PPV (Personalized PageRank Vector). Son principe fondamental consiste à privilégier les pages reliées aux pages préférées de l'utilisateur ou les pages citées par ces dernières au cours du processus de calcul des scores de sélection.

Dans le même sens, [78] ont proposé l'adaptation des paramètres de la fonction de pertinence au contexte de l'utilisateur. La méthode d'appariement document-requête proposée permet de générer automatiquement des stratégies d'ordonnancement pour des contextes différents en utilisant les techniques de programmation génétique.

Dans [273], le contexte de l'utilisateur est intégré dans la fonction d'appariement du modèle bayésien de RI. L'approche proposée 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 contexte de l'utilisateur. Plus précisément, 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 le contexte de 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. Le calcul du score de pertinence du document D vis-à-vis de la requête Q et l'utilisateur U est alors exprimé par :

$$RSV_U(Q, D) = \Psi_{k=1..n}(\mu_k(D, c_k) * p(Q/D, c_1, c_2, \dots, c_n)) \quad (1.11)$$

Dans [66, 67] les auteurs ont proposé de calculer une pertinence multidimensionnelle des documents intégrant des scores contextuels liés au profil utilisateur et à la source des documents. Le score contextuel relatif au profil de l'utilisateur est décliné selon deux fonctions “*coverage*” et “*appropriateness*” définies

comme suit :

$$Coverage(c, d) = \frac{\sum_{i=1}^{|T|} \min(w_{ic}, w_{id})}{\sum_{i=1}^{|T|} w_{ic}}, \quad (1.12)$$

$$Appropriateness(d, c) = 1 - \frac{\sum_{i=1}^{|T|} |w_{ic} - w_{id}|}{|T|}, \quad (1.13)$$

où w_{ic} et w_{id} représentent les poids des termes respectivement dans le document d et le profil de l'utilisateur c , $|T|$ est la taille du vocabulaire de la collection. La fonction de pertinence liée à la source des documents dénotée “*reliability*” est donnée par la formule suivante :

$$\Gamma_u(d) = T_u(s(d)), \quad (1.14)$$

où $s(d)$ représente la source du document d , et $T_u(s(d))$ représente le degré de confiance de l'utilisateur u en la source $s(d)$.

Le score traditionnel d'un document est calculé par une fonction notée “*aboutness*” donnée par la mesure standard de *cosinus* :

$$Aboutness(q, d) = \frac{\sum_{i=1}^{|T|} w_{iq} \cdot w_{id}}{\sqrt{\sum_{i=1}^{|T|} w_{iq}^2 \cdot \sum_{i=1}^{|T|} w_{id}^2}} \quad (1.15)$$

w_{iq} représente le poids des termes dans la requête q .

Pour la combinaison des scores de pertinences individuelles d'un document, les auteurs proposent un modèle d'agrégation prioritaire exploitant des opérateurs d'agrégation développés dans [265].

5.4.2.3 L'ordonnement des résultats

Cette phase peut également prendre en compte le contexte pour ré-ordonner les résultats fournis par le processus de sélection. De ce fait, l'ordre final des documents à présenter à l'utilisateur est une combinaison du score/rang produit par le processus de sélection classique et celui fourni par la similitude avec le contexte de l'utilisateur.

Dans les deux approches présentées dans [156, 249], le ré-ordonnement des résultats de recherche consiste à combiner le score d'appariement original du document avec le score de similarité entre le document et les centre d'intérêts de l'utilisateur représentant son contexte. Ce dernier est calculé en appliquant une mesure de similarité vectorielle basé sur le *cosinus* entre le document et le représentant du profil utilisateur.

Par opposition à ces approches, le ré-ordonnement des résultats de recherche dans [84, 219] est effectué par combinaison de rangs des documents.

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

$$FinalRank = \alpha * PersoRank + (1 - \alpha) \cdot KeywordRank \quad (1.16)$$

Le score personnalisé du document est calculé comme étant la somme de ses scores de similarité avec les N premiers concepts les plus similaires à ce document, et du poids des concepts dans le profil comme suit :

$$PersoScore = \sum_{i=1..N} score(c_i) \cdot \cos(\vec{d}, \vec{c}_i) \quad (1.17)$$

Une autre approche de ré-ordonnement est proposée dans [152]. Dans cette approche, N listes de résultats sont produites pour chaque concept du profil représentant l'utilisateur et pour le même ensemble des résultats initiaux restitués par le SRI classique. Le rang d'un document dans une liste dépend de sa similarité avec le concept représentant la liste. La fusion des listes des résultats est effectuée par 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 :

$$vote(d_i, L_j) = W_j \cdot (MM - i + 1) \quad (1.18)$$

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 :

$$W_j = rankC_j \cdot \sqrt{sim(q, C_j) \cdot numC_j} \quad (1.19)$$

et où $rankC_j$ est le rang de la catégorie associée à la liste L_j , $sim(q, C_j)$ est le score de similarité de la catégorie C_j avec la requête q et $numC$ 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.

6 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 vers la RI contextuelle en passons par la RI adaptative, notamment les techniques de la reformulation des requêtes, de la désambiguïsation du sens des mots des requêtes et du regroupement thématique des résultats de la recherche. Nous avons cité les problématiques de la RI classique en présence du contexte et les facteurs d'émergence de la RI contextuelle. Nous avons également donné un aperçu des principaux modèles d'accès contextuels à l'information.

Avec la prolifération des appareils mobiles, l'accès à l'information sur le Web est devenu possible à tout moment et à toute place. Ce nouveau cadre d'utilisation caractérisé par la variabilité et la dynamique du contexte d'accès à l'information présente un champs d'application par excellence de la RI contextuelle. C'est précisément à ce cadre applicatif que nous nous sommes intéressés dans les travaux de ma thèse. Dans le chapitre suivant, nous apportons un aperçu sur les enjeux de ce nouveau cadre d'utilisation pour la communauté RI et nous présentons les premiers travaux de RI contextuelles dans ce domaine.

Chapitre 2

RI contextuelle dans un environnement mobile

1 Introduction

L'évolution des supports mobiles d'accès à l'information (PDAs, téléphone mobile, etc) a transféré le domaine de la RI d'un environnement statique (PC-de bureau) vers un environnement de recherche en constante évolution du à la mobilité inhérente de l'utilisateur. L'accès au grand volume d'information qu'est la *web* est devenu possible depuis pratiquement toute place et à tout moment. Ce nouveau cadre d'utilisation caractérisé par une surabondance des informations, des requêtes plus courtes et ambiguës, des contraintes des appareils mobiles, et de spécificité des besoins informationnels typiquement conséquence de la situation dans laquelle est impliqué l'utilisateur, rend les SRI traditionnels incapables de satisfaire les requêtes des utilisateurs dans cet environnement. Il est alors nécessaire d'adapter les SRI traditionnels, de nouvelles technologies sont nécessaires pour la représentation, la modélisation, l'indexation et la recherche des données mobiles. L'adaptation des SRI à ce nouveau cadre d'utilisation passe nécessairement par la prise en compte du contexte mobile lors du processus de recherche d'information. La croyance est que l'exploitation d'informations contextuelles dans le processus de RI peut améliorer la précision et la présentation des résultats de recherche.

Les premiers travaux en recherche d'information contextuelle se sont principalement focalisés sur la modélisation et l'exploitation du contexte cognitif de l'utilisateur [156, 84, 152, 219, 241]. Les travaux de recherche se sont étendus ces dernières années pour inclure le contexte physique et environnementale de l'utilisateur mobile. Des travaux récents [21, 166, 118, 44] tentent de répondre aux problématiques inhérentes à ce nouveau cadre d'utilisation et d'améliorer ainsi

les performances de recherche dans cet environnement. Ces travaux peuvent être regroupés sous le domaine de recherche d'information mobile et contextuelle, comme illustré dans la figure 2.1. Ce dernier se trouve à la croisée de plusieurs domaines de recherche : la recherche d'information, l'informatique mobile et les systèmes sensibles au contexte.

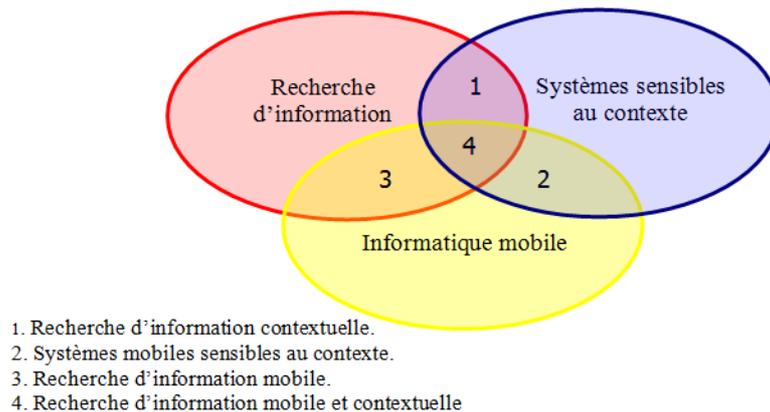


FIG. 2.1 – Domaines liés à notre problématique de recherche.

Comparativement à la RI contextuelle dans les contextes classiques de l'informatique de bureau, la RI contextuelle dans le cadre mobile vise à traiter le problème de la surcharge d'information en fournissant des résultats appropriés aux contraintes techniques des appareils mobiles d'une part et à des facteurs personnels et environnementaux de l'utilisateur d'autre part. Parmi les éléments contextuels les plus importants traités dans la littérature, nous citons le dispositif d'accès à l'information, la localisation géographique de l'utilisateur, le temps, les centres d'intérêt de l'utilisateur et le contexte social de l'utilisateur. Les approches dans le domaine se distinguent par différents aspects liés aux éléments contextuels effectivement pris en compte lors de la recherche. Nous distinguons la source et la nature des données exploitées dans la dérivation du contexte, le mode de construction et de représentation du contexte et le modèle d'accès contextuel à l'information qui intègre un ou plusieurs éléments contextuels dans une ou plusieurs phases du processus de RI.

La prise en compte du contexte dans la nouvelle génération des systèmes de recherche d'information mobile a remis en cause le modèle d'évaluation classique orienté-laboratoire. En effet, les limites de ce modèle d'évaluation sont principalement liées à la non implication du contexte de recherche dans le protocole d'évaluation des SRI. Ces limites ont donné naissance à une nouvelle génération de méthodologies d'évaluation intégrant le contexte dans le processus d'évaluation.

Ce chapitre traite de l'application de la RI contextuelle dans l'environnement mobile. Nous discutons en premier lieu de l'émergence de la RI mobile, pour aborder ensuite les différentes définitions et représentations du contexte

dans ce cadre, et terminer par une revue des cadres d'évaluation proposés dans la littérature. Nous nous sommes intéressés dans la section 2 à l'émergence de la RI mobile, où nous discutons des spécificités de ce nouveau cadre de recherche, de la notion du contexte telle qu'elle est appliquée dans la RI mobile et des diverses problématiques qu'engendre ce nouveau cadre pour la communauté RI. Dans la section 3 nous présentons les différents modèles de représentation du contexte adoptés dans les travaux du domaine. Dans la section 4, nous mettons l'accent sur l'accès contextuel à l'information dans le processus de RI où nous abordons les principaux modèles et techniques d'exploitation du contexte proposés dans ce domaine. La section 5, est consacrée aux cadres d'évaluation de l'efficacité de la RI contextuelle dans un environnement mobile où nous mettons en avant les limites des approches d'évaluation standard, pour ensuite présenter les approches d'évaluation proposées dans le domaine pour pallier à ces limites. La section 6 conclut le chapitre.

2 Émergence de la RI mobile

L'évolution de l'équipement et des usages ainsi que l'évolution technique des terminaux et des navigateurs et enfin l'évolution des forfaits proposés par les opérateurs, sont les trois facteurs majeurs qui ont permis le démarrage de l'Internet Mobile. Au-delà des sites web mobiles, simples versions adaptées de sites web existants, des services exclusivement dédiés à une utilisation spécifique sur mobile sont aussi apparus et leur nombre ne cessent d'augmenter. Selon le rapport [240], le nombre de sites web dédiés aux mobiles est passé de 326.600 à 440.100 de Décembre 2009 à Avril 2010. Cette figure représente une croissance de 35% et une croissance annuelle de 232% de 2009 à 2010. Selon le même rapport, le nombre de sites dédiés aux mobiles est estimé à plus 1,1 millions fin 2011.

Par conséquence, il est apparu une demande croissante pour des outils de recherche efficaces adaptés aux environnements mobiles. Pour répondre à ce besoin, les systèmes de recherche sur PC ont été transférés vers cette nouvelle plate forme de recherche avec ses propres caractéristiques et ses limites. Le concept de système ou de moteur de recherche mobile est alors apparu, il est défini dans [132] comme "un logiciel conçu pour un appareil mobile pour fournir un service ou un portail, par lequel l'utilisateur peut soumettre une requête (habituellement par l'entrée d'un ensemble de mots clés) et obtenir la liste des résultats correspondants aux critères de recherche". Selon [246], la RI mobile consiste en l'indexation et la recherche d'information textuelles et multimédia pour être retournées sur un dispositif mobile avec des connections sans fils. Cette définition est donnée comme suit : "*Mobile information retrieval is concerned with the indexing and retrieval of information such as text, graphics,*

animation, sound, speech, image, video, and their possible combinations for use in mobile devices with wireless network connectivity.”.

Les outils de recherche sur mobile sont apparus aux débuts des années 2000. Les grands moteurs du Web (eg. Google, Yahoo, AOL, Bing, etc) ont affiché dès cette date leur présence. Cependant, ils ne s'étaient contentés que d'une simple transposition de leurs versions de bureau vers la technologie mobile. Pour les systèmes de recherche, le domaine de la RI mobile reste encore un *challenge* à relever. En effet, les utilisateurs n'ont pas le même comportement ni les mêmes recherches que sur les PC de bureau. De plus, les appareils mobiles sont des dispositifs d'accès à l'information personnels qui permettent de stocker et de capturer régulièrement un grand nombre de données utilisateur, tels que sa localisation, la liste de ses contacts, ses préférences, etc.

Les verrous technologiques et scientifiques à lever alors sont de développer des systèmes de recherche qui sont adaptés d'un côté au dispositif mobile d'accès à l'information mais aussi de modéliser et d'exploiter les diverses sources d'information contextuelles qu'offre le dispositif mobile de l'utilisateur en vue d'améliorer la précision des résultats de la recherche.

Nous présentons dans cette section les spécificités de la RI mobile, la définition de la notion de contexte dans ce cadre d'application et les problématiques qu'il soulève pour la recherche d'information dans ce domaine.

2.1 Spécificités de la RI mobile

On peut distinguer deux types fondamentaux de différences entre la recherche mobile et la recherche de bureau : du point de vue du dispositif mobile et du point de vue de l'usage et de la pratique de la recherche d'information.

2.1.1 Du point de vue du dispositif mobile

Les appareils mobiles présentent des caractéristiques en termes de limites et d'atouts par rapport à leur homologues de bureau.

1. **Les limites** : malgré les avancées technologiques, les appareils mobiles présentent à ce jour quelques limites, parmi lesquelles on peut citer :
 - La taille des écrans des appareils mobiles est réduite et elle varie à travers les modèles.
 - La puissance de traitement des appareils mobile est faible et leur utilisation reste dépendante de la vie de la batterie.

- La difficulté de saisie des requêtes liée à la disposition des caractères du clavier.
2. **Les atouts** : parmi les atouts majeurs de ces appareils on peut citer :
- Le caractère personnel de l'appareil mobile qui permet de bien identifier le profil de l'utilisateur.
 - L'omniprésence de l'appareil mobile rend l'accès à l'information possible à tout moment et à toute place.
 - La possibilité de fournir des informations de localisation de l'utilisateur avec des précisions très importantes.
 - La possibilité de définir le réseau social de l'utilisateur à partir des listes de contacts, des personnes à côté, etc.

2.1.2 Du point de vue de l'usage et de la pratique de la recherche d'information

Diverses études [122, 123, 268, 223, 54, 53] ont été menées dans le but de comprendre la nature des besoins informationnels des utilisateurs mobiles et leur pratique de recherche. Divers enseignements peuvent en être tirés, on peut les grouper selon les points suivant :

1. **Usage de la recherche** : selon les analyses faites ([122, 268]) sur les *logs* de recherche, il a été montré que :
 - les requêtes des utilisateurs mobiles sont plus courtes en moyenne 2.3 termes (vis-à-vis 2.6 sur les PC de bureau).
 - il y a moins de requêtes par session¹ en moyenne 1.6 requêtes (vis-à-vis 2.3 sur les PC de bureau).
 - l'utilisateur mobile a moins de temps pour naviguer dans la liste des résultats de recherche, généralement seule la première page des résultats de recherche est consultée.
2. **Thèmes ou catégories des requêtes** : les études [122, 123, 54] rapportent que les catégories les plus populaires (après la catégorie adulte) sont les catégories *services locaux, divertissement, internet et telecom, voyages et loisirs et sport*. La figure 2.2 donne une vue plus détaillée de la répartition des requêtes selon les diverses catégories proposées dans [122].
3. **Type du besoin derrière les requêtes** : l'étude [53] a rapporté une classification des requêtes mobiles selon le type du besoin derrière la

¹une session est définie comme dans [221] par "la série de requêtes soumises par un utilisateur unique dans un petit intervalle de temps".

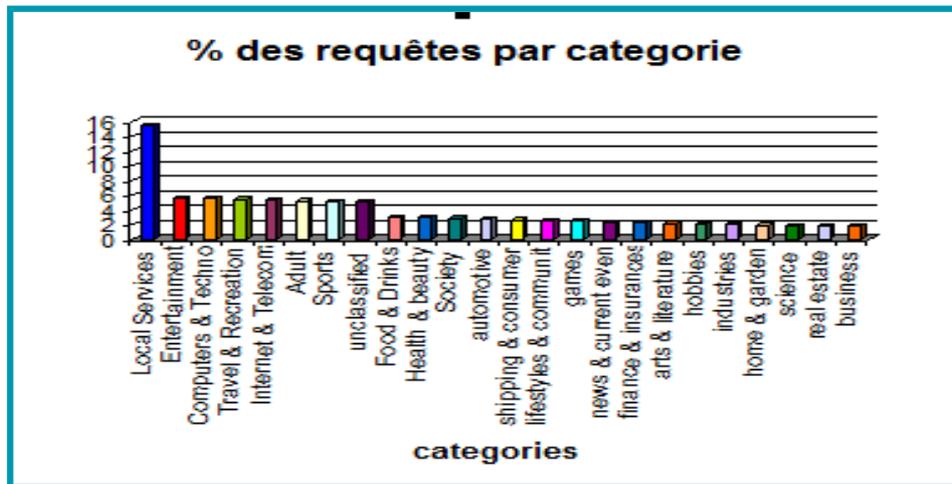


FIG. 2.2 – Pourcentage des requêtes par catégorie selon l'analyse de [122].

requête. Les auteurs ont identifié trois principales classes de besoins : informationnel à 58%, géographique à 31% et personnel à 11%, comme illustré dans la figure 2.3. Les besoins personnels sont liés à la recherche des données personnelles de l'utilisateur sur son mobile (e.g. un contact, un message, une image, etc). Les besoins géographiques sont liés à des informations localisées. Les auteurs soulignent la part importante de cette catégorie qui peut être de plus distinguée en trois sous catégories : locale implicite à 15% (sans référence à une localisation), locale explicite à 13% (avec référence à une localisation) et de directions à 4%.

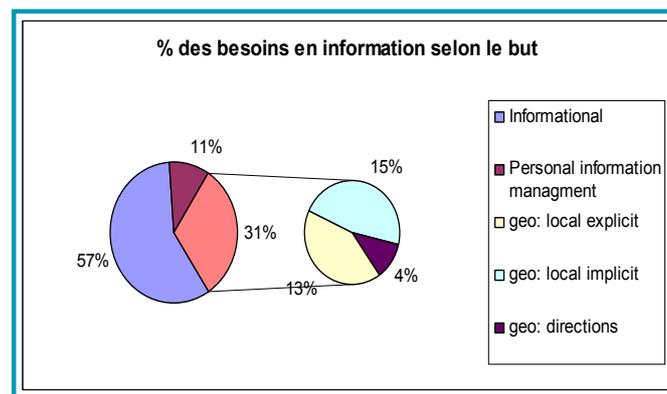


FIG. 2.3 – Pourcentage des requêtes selon le type du besoin informationnel de l'utilisateur mobile.

4. **Sensibilité des requêtes au contexte** : selon l'étude [223], 72% des besoins informationnels des utilisateurs mobiles sont liés à des facteurs contextuels. La figure 2.4 illustre le pourcentage des principaux facteurs contextuels identifiés dans cette étude, notamment la localisation, le temps,

l'activité et la conversation.

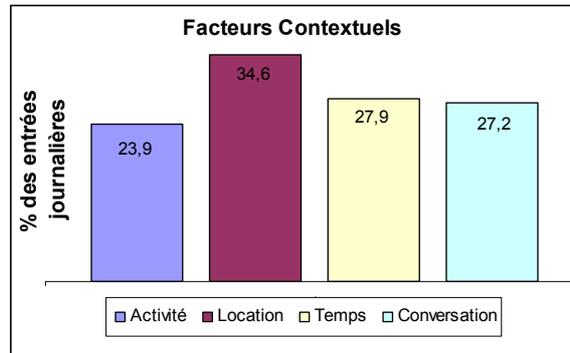


FIG. 2.4 – Pourcentage des différents facteurs contextuels derrière le besoin informationnel de l'utilisateur mobile.

2.2 Notion de contexte dans la RI mobile

Les premières définitions du contexte dans l'informatique mobile remontent à Abowd et al. [12], où : "le contexte est défini par toute information qui peut être utilisée pour caractériser une entité. Une entité peut être une personne, un lieu, un objet, pouvant être considéré comme approprié dans l'interaction homme/application, incluant l'utilisateur et l'application eux-mêmes."

Une autre définition est celle citée dans Chen et al. [49], où le contexte est défini par : "Le contexte est un ensemble d'états et de paramètres qui soit détermine le comportement d'une application ou bien dans lequel un événement d'application se produit et est intéressant pour l'utilisateur."

Après ces définitions intentionnelles et plutôt théoriques, des définitions par extension et orientées application ont été proposées dans la littérature. Le contexte est alors défini comme une liste de dimensions et leurs valeurs associées. Dans le travail de Schilit et al. [212] qui introduit la notion de système "sensible au contexte", le contexte est représenté par la localisation de l'utilisateur et les objets environnants. De la même manière, Brown et al. [39] définissent le contexte par la localisation, la proximité d'autres personnes, la température, le jour, etc. Dans [211], la notion de contexte est divisée en trois catégories : le contexte informatique (réseau, affichage, etc), le contexte de l'utilisateur (profil, les personnes à proximité, etc) et le contexte physique (lumière, bruit, etc.). Chen et al. [49] y ajoutent deux autres catégories : le contexte temporel (jour, mois, etc) et l'historique des activités de l'utilisateur.

La plupart des travaux dans la littérature autour de la RI mobile gèrent la notion de contexte de manière *ad-hoc* en la redéfinissant en fonction des

besoins propres aux applications [192, 35, 94, 85, 161, 20, 90] ce qui présente l'inconvénient de ne pas offrir un cadre formel unificateur. L'étude conduite par Kaenampornpan et al. [121] fournit un état de l'art et une comparaison des divers modèles de contexte basés sur le type d'information qu'ils modélisent.

Dans la liste suivante, une sélection de quelques modèles clés est présentée :

1. **Modèle de contexte de Schmidt et al. [213]** : ce modèle de contexte est défini selon deux catégories principales : facteurs humains et facteurs environnementaux physiques.
 - *Facteurs humains* : qui regroupent les sous catégories d'information utilisateur, de l'environnement social de l'utilisateur et de la tâche de recherche de l'utilisateur.
 - *Facteurs environnementaux physiques* : peuvent être divisés en les sous catégories de la localisation de l'utilisateur, l'infrastructure disponible et les conditions physiques entourant l'utilisateur.

2. **Modèle de contexte de Myrhaug [166]** : ce modèle divise le spectre contextuel en cinq catégories : contexte environnemental, contexte personnel, contexte de la tâche de recherche, contexte social et contexte spatio/temporel.
 - *Le contexte environnemental* : capture les entités entourant l'utilisateur.
 - *Le contexte personnel* : modélise les attributs de l'utilisateur, il est en outre subdivisé en deux sous catégories de contexte physiologique et de contexte mentale.
 - *Le contexte de la tâche de recherche* : représente les attributs qui décrivent des aspects liés au type du besoin en information de l'utilisateur ou son intention de recherche.
 - *Le contexte social* : modélise les relations sociales d'un utilisateur.
 - *Le contexte spatio-temporel* : décrit la localisation de l'utilisateur et les aspects temporels incluant des mouvements potentiels en relation à d'autres objets ou d'entités dans l'environnement de l'utilisateur (eg. des bâtiments, des véhicules).

3. **Modèle de contexte de Arias et al. [20]** : ce modèle propose une classification des différents types de contexte exploités dans la RI mobile en cinq familles : Profil utilisateur, dispositif et navigateur, contexte géomatique, contexte environnemental et contexte temporel.
 - *Profil utilisateur* : contient toutes les informations implicites ou explicites relatives à l'utilisateur (eg., langue, âge, centre d'intérêt).

- *Dispositif et navigateur* : regroupe les propriétés décrivant les caractéristiques de l'appareil mobile de l'utilisateur et le navigateur web utilisé.
- *Contexte géomatique* : regroupe les aspects relatifs à la localisation de l'utilisateur non seulement en termes de coordonnées géographiques, mais aussi en termes de type de place (au bord de la mer, à la campagne, à l'aéroport, etc).
- *Contexte environnemental* : regroupe des propriétés relatives aux conditions environnementales de l'utilisateur (eg., la luminosité, le bruit, le climat, etc).
- *Contexte temporel* : regroupe les aspects relatifs à la date, l'heure système et l'historique des actions de l'utilisateur.

2.3 Problématique de la RI contextuelle et mobile

Toutes ces particularités à la fois techniques et pratiques de la recherche mobile conjuguées avec la croissance du volume d'information sur le web, engendrent des problématiques pour la communauté RI dans ce domaine. Il apparaît clair, que la mise en œuvre d'un SRI mobile doit nécessairement prendre en compte à la fois le contexte du dispositif mobile mais aussi le contexte de l'utilisateur et de son environnement.

On peut identifier plusieurs aspects à promouvoir dans les SRI contextuels mobiles :

- L'amélioration des interfaces de recherche qui passe par la proposition de techniques pour : faciliter la saisie de requêtes, adapter l'affichage des résultats de recherche aux écrans des appareils mobiles et préserver l'énergie et la bande passante qui sont limitées dans cet environnement.
- L'exploitation du contexte courant de recherche de l'utilisateur pour améliorer la précision des résultats retournés. Ceci pose alors des problèmes liés à la définition, la formalisation, puis à la déclinaison du modèle de contexte dans le processus d'accès à l'information.
- La combinaison de plusieurs dimensions du contexte pour assurer une meilleure adaptation. Cela nécessite de réfléchir le modèle de combinaison adéquat pour prendre en compte différentes fonctions de *ranking*.
- L'évaluation de la RI mobile en tenant compte des facteurs contextuels influant la perception de l'utilisateur de la pertinence et l'utilité de l'information retournée par le moteur de recherche. Cela se fera par la proposition de cadres adéquats basés sur des utilisateurs et des situations de recherche spécifiques, de nouvelles métriques d'évaluation intégrant

différents niveaux de pertinence, de nouvelles collections de tests décrites par des méta-données descriptives des contextes de recherche.

Les verrous qui sont posés par la contextualisation de la recherche dans un cadre mobile peuvent se projeter sur trois principaux niveaux. Le premier niveau concerne la modélisation du contexte utilisateur en RI mobile. Le second niveau se rapporte aux approches de d'adaptation qui exploitent les modèles de contexte dans le processus de RI. Le dernier niveau traite des démarches d'évaluation adoptées dans ce cadre. Nous abordons chacun de ces éléments dans la suite de ce chapitre.

3 Modélisation du contexte en RI mobile

La modélisation du contexte vise à permettre une meilleure compréhension des informations contextuelles. Elle est nécessaire pour une exploitation plus efficace du contexte dans les SRI. La nature et la portée du modèle dépendent des dimensions du contexte considérées. Plusieurs modèles du contexte ont été proposés dans le domaine de la RI mobile, ils peuvent être caractérisés par deux éléments clés : les techniques de construction du contexte et les techniques de représentation du contexte adoptées.

Nous présentons dans cette section, les principales techniques utilisées pour la construction puis la représentation du contexte dans le cadre de la RI contextuelle et mobile.

3.1 Construction du contexte mobile

La construction du contexte de l'utilisateur mobile repose sur des outils et des algorithmes permettant de collecter les sources d'information pertinentes servant à la dérivation du modèle de contexte de l'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 sources d'information contextuelles utilisées et leurs stratégies de collecte et les techniques d'acquisition du modèle contextuel.

3.1.1 Les sources d'information et leurs stratégies de collecte

Cette section examine les différentes stratégies et technologies appliquées pour l'acquisition des données pour l'instanciation des modèles contextuels. Selon la dimension du contexte considérée, nous pouvons distinguer les sources d'information pour la dimension : du dispositif mobile, du profil utilisateur, du contexte spatial, du contexte temporel et du contexte social.

1. **Les sources d'information pour le dispositif mobile** : les travaux de la littérature ([44, 244]), se réfèrent à des spécifications standards tels que : WURFL², UAProf³, CC/PP⁴ et le Dépôt de Description de Dispositif (*Device Description Repository-DDR*)⁵, qui incluent des informations relatives à un appareil mobile donné, telles que : les dimensions d'écran, les mécanismes d'entrée, les couleurs supportées, les limites connues, les capacités spécifiques, etc. Différents en-têtes de la requête http mobile (eg. "x-wap-profile", "19-profile", "wap-profile" ou d'autres en-têtes similaires) contiennent des références (URL) vers le fichier de description du dispositif mobile qui a émis la requête.

Le mode d'acquisition est implicite et échappe à la connaissance de l'utilisateur.

2. **Les sources d'information pour le profil utilisateur** : on distingue deux modes d'acquisition des données utilisées dans la construction du profil utilisateur : acquisition explicite et acquisition implicite.

- *L'acquisition explicite* : ce mode permet à l'utilisateur de saisir manuellement ses domaines d'intérêt ou alors juger de façon explicite les documents renvoyés pour une requête servant à la construction de son profil. Dans le système GUIDE [50], les utilisateurs sont invités à saisir leur centres d'intérêts et leur préférences de langue à la première utilisation du système. Dans [29] les auteurs développent un questionnaire spécifique pour collecter les intérêts des utilisateurs sur leur téléphones mobiles. Dans [174], l'utilisateur exprime explicitement son degré d'intérêt (haut, moyen, bas, ou null) en navigant dans une hiérarchie conceptuelle décrivant les sujets de journaux, via une interface utilisateur dédiée. Dans [172] les utilisateurs peuvent améliorer l'initialisation de leurs profils en exprimant leurs préférences de catégories et de types de services comme des poids sur une échelle de valeurs entre -1 et 100. Dans le système MobileIQ [184], l'utilisateur juge explicitement les pages web retournées par le système en assignant

²<http://wurfl.sourceforge.net/>

³http://w3development.de/rdf/uaprof_repository/

⁴www.w3.org/Mobile/CCPP/

⁵<http://www.w3.org/2005/MWI/DDWG/drafts/api/070806.html>

des valeurs dans l'intervalle 1 à 10. Dans [155], une interface graphique est utilisée pour récolter d'une façon interactive les appréciations de l'utilisateur sur des articles recommandés sur une échelle de trois valeurs : pertinent, moyen pertinent, ou pas pertinent.

L'avantage majeur des techniques explicites est que la construction du profil utilisateur est plus précise. Cependant, elles présentent l'inconvénient de nécessiter un effort supplémentaire de l'utilisateur pour spécifier explicitement ses besoins. D'après [183] de nombreux utilisateurs ne sont pas disposés à remplir des formulaires sur leurs préférences, ceci sans oublier les restrictions liées à la difficulté de la saisie des informations sur les claviers des appareils mobiles. Ces limitations ont orienté les travaux vers des techniques d'acquisition implicite des données utilisateur dans le processus de construction de son profil.

- *L'acquisition implicite* : consiste à collecter les données de l'utilisateur en observant ses comportements et interactions avec le système durant les activités de recherche. Les auteurs dans [87, 252, 172] construisent le profil en analysant les pages web visitées par l'utilisateur lors de sa recherche. Les approches de construction des profils dans [52, 148] sont basées sur l'utilisation de l'historique des requêtes et des pages visitées par les utilisateurs. D'autres sources d'information sont également exploitées telles que les applications ouvertes dans [252, 266], l'agenda électronique et les mails envoyés/reçus dans [27], l'historique des places visitées dans [29, 270].

La faiblesse des approches implicites est qu'il faut une certaine quantité d'événements où l'utilisateur aurait manifesté ses intérêts pour des informations retournées avant qu'il puisse faire une première adaptation. Pour compenser cet inconvénient, certains travaux combinent les deux modes implicite et explicite [172, 174, 20, 251].

3. **Les sources d'information pour le contexte spatial** : pour l'acquisition de la localisation des utilisateurs, un large éventail de différentes technologies de positionnement sont utilisées. Ces technologies dépendent selon que l'application de RI fonctionne à l'intérieur ou à l'extérieur.

Pour un positionnement extérieur, le système de positionnement mondial (*Global Positioning System*-GPS) ou son amélioration le GPS différentiel sont largement utilisés. Le GPS est un système de navigation par satellites disponible au niveau mondial qui permet aux périphériques compatibles de déterminer leurs positions, la vitesse et la direction du mouvement de leurs utilisateurs. Les systèmes Cyberguide [12], Magitti [27], WebPark [165], Crumpet [183] ou Compass [251] tous se basent sur ce type de positionnement pour déterminer la localisation de l'utilisateur mobile.

Toutefois, dans des espaces d'intérieur le GPS manque de fiabilité et d'exactitude. Son signal de faible intensité est facilement bloqué par la plupart des bâtiments et en outre perturbé par les réflexions. Pour cette

raison, différentes technologies sont apparues pour faire face au problème du positionnement à l'intérieur. Le système Cyberguide [12] exploite la technologie de positionnement par infrarouge pour couvrir les situations d'utilisation à l'intérieur. [29] exploite les identificateurs des antennes cellulaires (ou station de base) de téléphonie mobile pour identifier la localisation de l'utilisateur. Le système GUIDE [50] utilise les réseaux locaux sans fil pour déterminer le positionnement des visiteurs de la ville de Lancaster. Le système Jimminy [192] combine des balises de localisation radio et des badges infrarouge actifs pour déterminer la localisation de l'utilisateur dans un laboratoire. La technologie Bluetooth a été utilisée pour la détection de la localisation des utilisateurs mobiles dans le guide touristique décrit dans [166].

Le mode d'acquisition des sources d'information pour le contexte spatial est implicite. Cependant, il n'échappe pas à la connaissance de l'utilisateur qui peut autoriser ou bloquer l'accès au service de localisation.

4. **Les sources d'information pour le contexte temporel** : le contexte temporel est le plus facile à obtenir en utilisant l'horloge intégrée au système qui est habituellement disponible dans tous les appareils mobiles et fournit l'heure exacte [52, 28, 139].
Le mode d'acquisition des sources d'information pour ce type de contexte est implicite et automatique.

5. **Les sources d'information pour le contexte de la tâche de recherche** : pour analyser ou identifier les types des besoins derrière les requêtes des utilisateurs mobiles, les *logs* des requêtes sont généralement exploités de façon implicite [54], [269]. Dans [223] les utilisateurs interviennent de façon explicite dans la détermination du type de besoin derrière leurs requêtes en répondant à des questionnaires.

6. **Les sources d'information pour le contexte social** : l'appareil mobile mais aussi d'autres outils propres à la localisation dans les réseaux mobiles (tels que les RFID) constituent une mine de ressources pour ce contexte. Le calendrier électronique et la liste des adresses de l'utilisateur sont utilisés dans [79]. Le système Taptu [239] exploite la liste des contacts mail de l'utilisateur. Les auteurs dans [51, 99] exploitent la liste des amis de l'utilisateur. Dans [244] la liste des amis de l'utilisateur est exploitée combinée avec des données de capteurs de localisation qui déterminent les amis géographiquement proches de l'utilisateur. Le système Jimminy [192] détermine les personnes à côté de l'utilisateur à l'aide de badges infrarouge actifs.

L'acquisition des sources d'information pour le contexte social est implicite.

3.1.2 Techniques d'acquisition du modèle contextuel

Les techniques utilisées pour l'acquisition du modèle contextuel sont principalement des techniques d'apprentissage statistiques. Mais peuvent inclure d'autres disciplines telles que : théorie de probabilités, *data mining*, les systèmes à base de règle, etc. Dans la suite, nous présentons les techniques les plus communes aux travaux de la littérature.

1. **Les techniques d'apprentissage automatique** : l'apprentissage automatique (machine-learning en anglais), un des champs d'étude de l'intelligence artificielle, est la discipline scientifique concernée par l'analyse, le développement, et l'implémentation de méthodes automatisables qui permettent à une machine (au sens large) d'évoluer grâce à un processus d'apprentissage. Plusieurs techniques et algorithmes d'apprentissage supervisé ou non supervisé sont utilisés pour la modélisation, l'apprentissage ou l'évolution d'un modèle de contexte, parmi lesquels on trouve :
 - *Ranking SVM* : algorithme introduit dans [114] dont l'idée de base est de formaliser le problème de l'apprentissage de l'ordonnancement (en anglais machine-learned ranking) [58] comme un problème de classification binaire sur des paires d'instances, et ensuite de résoudre ce problème en utilisant un SVM. Le point clé de la construction d'un RSVM est de construire l'ensemble d'apprentissage sur la base des préférences des utilisateurs et d'obtenir les relations de contraintes entre les instances candidates (des documents). Dans [148], des ontologies de contenus et de localisation, ainsi que les données des clics de l'utilisateur, sont utilisées dans l'apprentissage d'un RSVM pour obtenir un vecteur de préférences de contenu pondérées et un vecteur de préférences de places pondérées utilisées pour le ré-ordonnement des résultats de la recherche pour l'utilisateur.
 - *Corrélation et régression linéaire* : le calcul de corrélation entre des variables statistiques permet d'étudier les relations d'influence entre ces variables généralement afin de prédire une des variables en fonction des autres variables. Un modèle de régression linéaire multiple a été utilisé pour la prédiction de l'utilité de l'information en fonction du contexte spatial, temporel et des centres d'intérêts de l'utilisateur tel que décrit dans [28]. Dans [271] le calcul de coefficient de corrélation entre les variables contextuelles permet de prédire les préférences de l'utilisateur. Dans [124] un modèle linéaire est utilisé pour combiner plusieurs modèles de probabilité de prédiction de mots.
 - *Les réseaux de neurones artificiels-(ANN)* : un réseau de neurones est

constitué d'un groupe inter-connecté de neurones artificiels, et il traite l'information en utilisant une approche connexionniste de calcul. Les ANN sont des systèmes adaptatifs qui modifient leur structures sur la base d'informations externes ou internes qui traversent le réseau au cours de la phase d'apprentissage. Dans le système SmartCon [15], un ANN en trois couches utilisant un algorithme de rétro-propagation, est exploité pour la sélection de services web à recommander à l'utilisateur mobile sur la base d'un ensemble de facteurs contextuels liés aux caractéristiques techniques de l'appareil mobile, aux préférences de l'utilisateur et aux caractéristiques du service web. Dans [90] un réseau de neurones dynamique est utilisé pour apprendre les poids des contextes qui s'adaptent à la situation courante de l'utilisateur. La règle d'apprentissage est basée sur l'équation suivante :

$$\Delta W_k(t) = |\Delta C(t) + \mu \Delta F(t)|_k \cdot W_{k,reference}(t) \quad (2.1)$$

où $\Delta C(t)$ est donnée par : $\Delta C(t) = \frac{|C(t)-C(t-1)|}{C(t-1)}$, $C(t)$ est la valeur actuelle du contexte C et $C(t-1)$ est sa valeur précédente. Si $\Delta C(t)$ est nulle, alors $\Delta C(t) = -|C(t) - C_{average}|$, $C_{average}$ est la moyenne des valeurs du contexte de $\{C(1), C(2), \dots, C(t-1)\}$, μ est un coefficient spécifique à l'architecture du réseau, $\Delta F(t)$ est la différence du nombre d'entités contextuelles entre deux époques $\Delta F(t) = F(t) - F(t-1)$, $W_{k,reference}$ est la moyenne des valeurs du poids de l'entité du contexte k sur les époques passées $\{1, 2, \dots, t-1\}$. Finalement le poids de l'entité k du contexte à l'époque t est mis à jour comme suit :

$$W_k(t) = W_k(t-1) + \Delta W_k(t) \quad (2.2)$$

- *Les réseaux bayésiens* : ce sont des graphes orientés sans cycle issus du mariage entre la théorie des graphes et la théorie des probabilités. Un graphe est appréhendé selon un aspect qualitatif et un aspect quantitatif. L'aspect qualitatif est l'ensemble des nœuds du graphe représentant les variables du domaine traité ainsi que les relations de dépendance entre ces variables. L'aspect quantitatif permet d'évaluer les arcs reliant toute paire de nœuds au moyen d'un calcul de probabilités. Cette technique est exploitée par le système MoBe [173], où les nœuds du graphe représentent des informations contextuels. Le réseau bayésien commence à partir de contextes concrets (e.g. "sonorité 60dB") pour en inférer des contextes plus abstraits (e.g. "écouter de la musique"). Dans [46] les réseaux bayésiens sont exploités pour prédire le type des places visitées par l'utilisateur mobile sur la base de la trace temporelle de ses visites. Un exemple de réseaux bayésiens exploités est illustré dans la figure 2.5, il est composé de 4 nœuds. Le nœud "weekend" représente une variable booléenne qui désigne si une observation donnée a eu lieu en *weekend* ou non. Cette information représente la variabi-

lité de l'activité des utilisateurs mobiles entre les jours de travail et les *weekends*. Le nœud “*Hour*” est une variable discrète qui peut prendre 24-valeurs représentant le temps de la journée. Le nœud “*kind of place*” représente une valeur discrète de type de place (e.g. *home, work, restaurant, pub, disco*). Le nœud “*happens*” est une variable booléenne qui exprime si l'utilisateur visite cet endroit à ce moment.

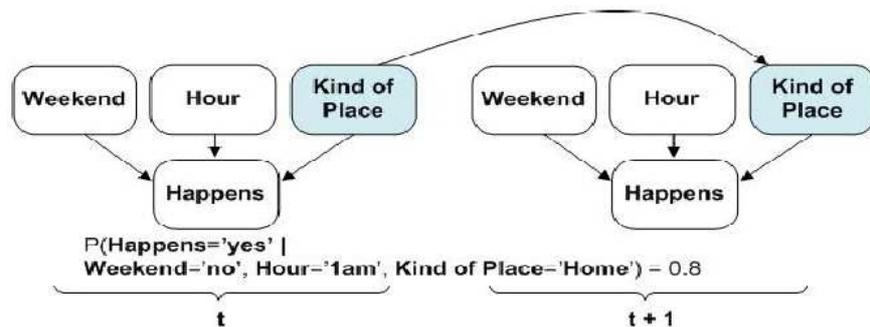


FIG. 2.5 – Réseau bayésien pour inférer le type de place visitée à partir de la trace temporelle (les nœuds blancs sont fournis en entrée du réseau).

- *Les techniques de classification (clustering)* : ces techniques permettent de partitionner les entités au sein d'une collection de données, en classes similaires, sur la base des informations qu'ils contiennent. Les *clusters* ne sont pas définies à l'avance, mais sont basées sur la proximité des entités les unes aux autres dans un hyperespace multidimensionnel. Les algorithmes de *clustering* sont utilisés dans [266] pour regrouper des situations similaires de recherche de l'utilisateur. Le système Hyponym [21] exploite l'algorithme CEM (Classification EM) dans le cadre d'un modèle de mélange de probabilité, pour une classification automatique des requêtes et des pages web. Le système MobileIQ [184] utilise un algorithme de *clustering* flou sur une matrice de préférences de recherche des utilisateurs pour apprendre des *clusters* d'utilisateurs ayant des profils similaires. Dans [87, 252, 174], différents algorithmes de classification sont utilisés pour une classification supervisée des documents selon des catégories de l'ontologie ODP [87, 252] et des catégories d'une taxonomie hiérarchique de News dans [174]. Des algorithmes de *clustering* de données de Cell ID sont utilisés dans [29] pour identifier les régions fréquemment visitées par les utilisateurs mobiles. Le système Hapori [139] utilise les algorithmes de classification KNN et LMNN pour l'apprentissage d'un modèle de préférences pour chaque catégorie de POI. De plus les utilisateurs sont classés en groupes en exploitant une mesure de similarité basée sur deux mesures de distance : la distance de Hellinger et la distance d'information de Fisher, avec l'algorithme de *multidimensional scaling* pour réduire la dimension de la

matrice de similarité.

2. **Les systèmes à base de règles** : les techniques à base de règles sont fondées sur l'apprentissage d'un ensemble de règles qui sont en mesure de classer les observations à partir de leurs attributs associés. Un tel système fondé sur les règles exprime alors une représentation explicite de connaissances - habituellement présentées comme des arbres de décision ou des probabilités. Cette technique est adaptée par les systèmes MoBe [173] et Compass [251], pour l'inférence de contexte abstrait à partir de contexte de bas niveau et l'agrégation de contextes. D'autres travaux qui exploitent cette technique sont présentés dans [126, 20] qui proposent d'utiliser des *règles sémantiques* pour la mise en correspondance des valeurs contextuelles de bas niveau (eg. les coordonnées GPS) avec des concepts abstraits d'ontologie et au même temps permettre l'interprétation de ces valeurs de bas niveau en tenant compte des préférences de l'utilisateur. Le langage de règles du web sémantique (SWRL⁶) est employé avec une extension par des règles de la logique du premier ordre. Un exemple de règles d'inférence est le suivant :

$$\text{Person}(?p) \wedge \text{hasAge}(?p, ?a) \wedge \text{swrlb:lessThan}(?a, 18) \rightarrow \text{Minor}(?p)$$

Le côté gauche et droit d'une règle représentent deux valeurs contextuelles et la règle suggère un lien entre elles.

3. **Les modèles probabilistes** : ces modèles permettent de calculer la probabilité d'un évènement sur la base des observations de données généralement issues de l'historique des activités de l'utilisateur ou de groupes d'utilisateurs. Dans [94] des probabilités de co-occurrence de mots contextuels sont calculées pour reformuler les requêtes initiales des utilisateurs. Dans [124] des probabilités de prédiction de mots sur la base de préfixes saisis par l'utilisateur ou de préfixe et de contexte local ou temporel sont calculées pour la complétude des requêtes des utilisateurs mobiles. Le système Magitti [27] se base sur le modèle probabiliste pour prédire l'activité de l'utilisateur en se basant sur la combinaison de modèles d'activité observés dans la population démographique de l'utilisateur et son modèle de comportement individuel, comme illustré dans la figure 2.6.
4. **Les techniques de fouille de données-*data mining*** : ces techniques permettent l'extraction de modèles à partir d'un grand nombre de données. Des techniques de fouille de l'usage du web (*Weblog mining techniques*) sont appliquées dans [94] pour extraire des activités typiques (en terme de noms d'objet et de noms d'action) liées à un nom de place géographique de l'utilisateur. Dans [270] un algorithme de

⁶<http://www.w3.org/Submission/SWRL/>

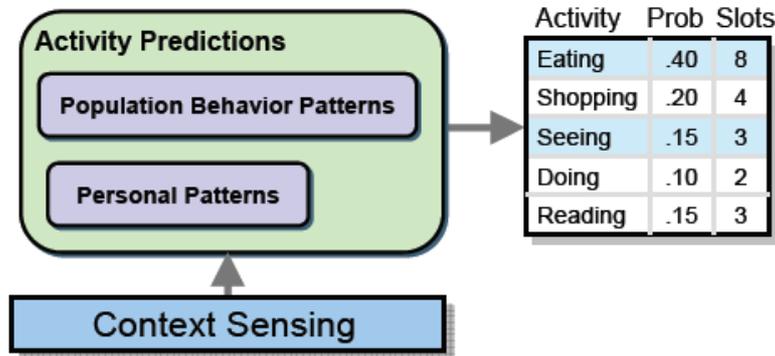


FIG. 2.6 – Processus de prédiction du type d’activité de l’utilisateur mobile.

fouille de données séquentielles ”*Prefix-Span*” est appliqué sur l’ensemble des trajectoires sémantiques de l’utilisateur pour extraire les trajectoires sémantiques les plus fréquentes et les plus représentatives de son profil. Dans [148], les auteurs définissent une formule de support, inspirée du problème bien connu de recherche d’items fréquents en *data mining*, pour mesurer l’intérêt et la pertinence d’un mot/phrase par rapport à une requête. Le formule de support est définie comme suit :

$$support(c_i) = \frac{sf(c_i)}{n} \cdot |c_i| \quad (2.3)$$

où c_i désigne un mot/phrase extrait à partir des *snippets*⁷ associés à la requête q , $sf(c_i)$ est le nombre de *snippets* contenant le mot/phrase c_i , n est le nombre de *snippets* retournés pour la requête q et $|c_i|$ est le nombre de termes dans le mot/phrase c_i .

3.2 Représentation du contexte

La représentation du contexte utilisateur consiste à représenter de manière interne une formalisation des données contextuelles qui le définissent. Dans cette section, nous décrivons les ressources de modélisation utilisées, puis nous présentons les principaux modèles de représentation adoptés pour aboutir à un modèle formalisé du contexte, et nous terminons par une synthèse des modèles de représentation proposés dans ce domaine.

⁷“snippet” dénote le titre, le résumé et l’URL d’une page web retournée par un moteur de recherche web.

3.2.1 Ressources de modélisation

Des ressources, généralement sémantiques, sont parfois exploitées ou même développées pour les besoins des applications pour enrichir les données du modèle contextuel. Selon la portée de l'application de recherche, on peut distinguer les travaux qui utilisent des ressources sémantiques générales (e.g. taxonomies du web), des ontologies de domaine, ou des ressources sémantiques propriétaires.

1. **Exploitation d'une ressource sémantique générale** : ce type de ressources sont généralement conçues pour répertorier le contenu du web, comme les portails en ligne tels que "Yahoo", "Magellan", et "l'ODP". Elles couvrent la plupart sinon tout les domaines d'information, et peuvent être donc utilisées pour représenter les centres d'intérêts des utilisateurs. Les auteurs dans [87] et [252] utilisent l'ontologie générale ODP comme ressource de modélisation des centres d'intérêt de l'utilisateur pour une recherche web mobile. La figure 5.1 illustre l'interface de l'ODP qui montre en gras les concepts du plus haut niveau de l'ontologie. Chaque concept de l'ODP représente un domaine d'intérêt des utilisateurs du web et est associé manuellement par des éditeurs à des pages web dont le contenu correspond au concept. Les concepts sont organisés hiérarchiquement,

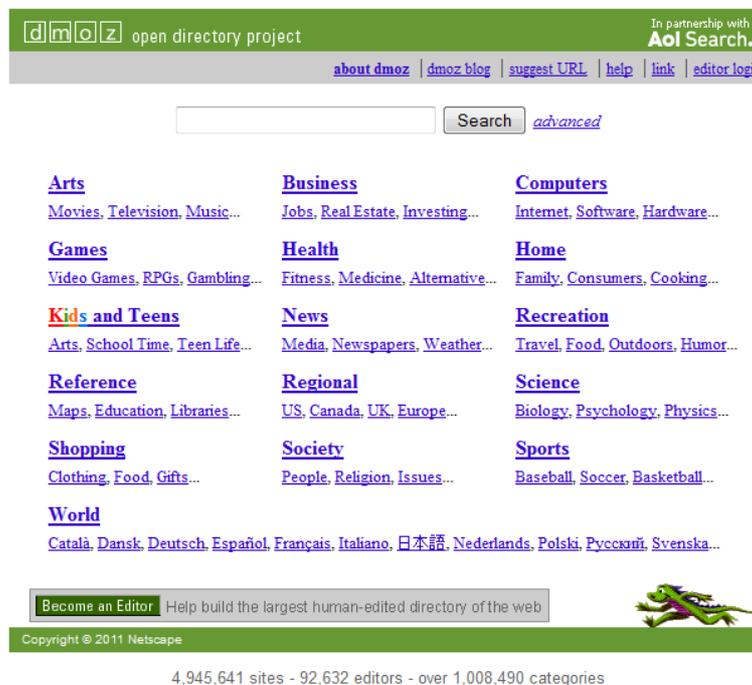


FIG. 2.7 – L'interface de l'ontologie générale ODP.

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 représenté par un titre et une description décrivant en général le contenu des pages web associées. Chaque page web est également associée à un titre et une description décrivant son contenu. Les concepts de l'ontologie sont reliés avec des relations de différents types tels que "is-a", "symbolic" et "related".

2. **Exploitation d'une ontologie de domaine** : ces ressources sont conçues pour couvrir des domaines spécifiques. Des ontologies du domaine touristique sont utilisées dans [183]. La taxonomie de catégories de thèmes de journaux de Reuters est exploitée dans [174], de même la taxonomie de catégories d'événements de Reuters est exploitée dans [28]. Pour modéliser le contexte géographique les travaux exploitent soit des bases de données ou systèmes d'information géographiques (GIS) en local [165, 35, 28, 183, 242, 229, 270, 148], soit des services de systèmes d'information géographiques en ligne via des services web tels que *AddressFinder* [253].
3. **Exploitation d'une ressource propriétaire** : ces ressources sont développées dans certains travaux pour représenter leur propres ontologies du contexte. Des langages de représentation standards sont notamment utilisés tels que OWL⁸ ou RDF⁹. Pour représenter les préférences de contenus (thèmes) des utilisateurs, les auteurs dans [148] construisent une ontologie de concepts thématiques à partir des documents retournés pour chaque requête en utilisant des heuristiques de co-occurrence de mots dans les documents résultats de la recherche pour la requête. La figure 2.8 présente l'ontologie de contenu extraite pour la requête "apple". Les concepts sont reliés par une relation de similarité illustrée par une flèche double (\leftrightarrow) ou une relation de descendances père-fils illustrée par une flèche simple (\rightarrow).

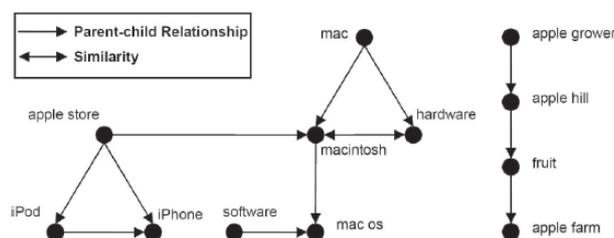


FIG. 2.8 – Exemple d'ontologie extraite pour la requête "apple".

Une ontologie décrivant le domaine du commerce et des finances est développée puis utilisée dans [155] pour modéliser le profil utilisateur. Cette ontologie est composée de deux parties : une partie contenant les

⁸www.w3.org/TR/owl-features/

⁹www.w3.org/RDF/

concepts de l'ontologie du domaine, dénotée O_{BTO} , et une seconde partie contenant les entités du domaine (e.g. entreprises, managers, produits, etc), dénotée O_{CMO} . L'ontologie O_{BTO} est composée de 1500 concepts du domaine du commerce. Elle est construite manuellement et se base sur des relations de spécification/généralisation. Cette ontologie est également enrichie par des liens de références vers des articles de Wikipedia¹⁰ ou des concepts de Wordnet¹¹. La figure 2.9 illustre un exemple partiel de la représentation du concept “*Option Contract*” et ses propriétés. La construction de la partie O_{CMO} est automatique et se base sur des techniques statistiques.

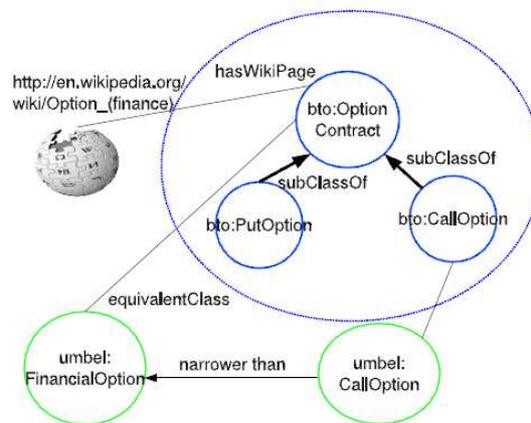


FIG. 2.9 – Représentation partielle du concept “*Option Contract*” et ses propriétés.

Certains travaux [20, 107, 244, 108, 253] proposent de développer des ontologies contextuelles qui peuvent regrouper diverses dimensions contextuelles (profil utilisateur, géographique, matériel, temporel, etc) et peuvent également réutiliser des ontologies existantes. La figure 2.10 présente une vue simplifiée de l'ontologie de contexte développée dans [107] dans une notation type UML.

La partie supérieure de l'image représente les concepts relatifs au domaine médical (e.g. personnes, rôles, document médical, matériel informatique/médical, etc). La partie inférieure représente des concepts pour la position d'une entité dans l'espace et dans le temps. Différentes relations relient les concepts de cette ontologie.

3.2.2 Modèles de représentation du contexte

Ces modèles permettent de formaliser la représentation du contexte en qualité de structure unifiée (partie d'une ontologie, classe de vecteurs de termes,

¹⁰http://en.wikipedia.org/wiki/Main_Page

¹¹<http://wordnet.princeton.edu/>

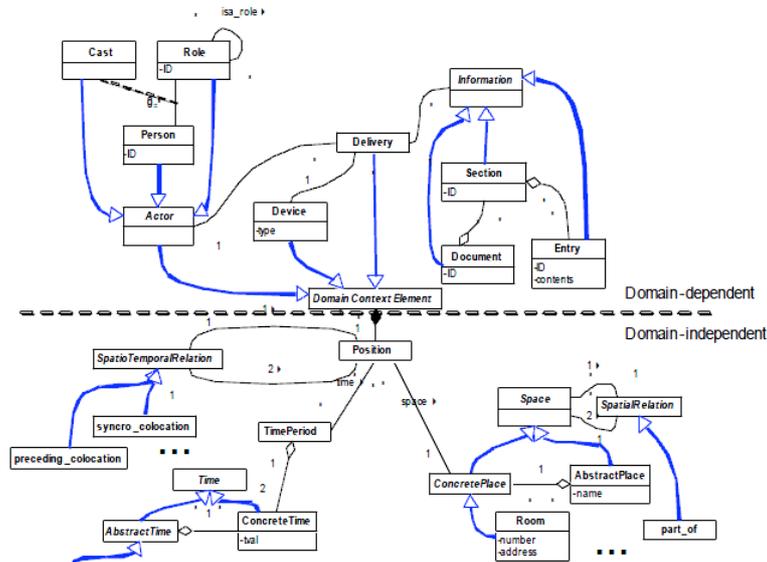


FIG. 2.10 – Exemple d’ontologie propriétaire développée dans [107].

ensemble de concepts ...) ou d’un ensemble d’informations avec des structures différentes et spécifiques, puis de les faire évoluer au cours du temps. Dans la suite de cette section, nous présentons les principaux modèles de la littérature.

3.2.2.1 Modèles de représentation basés sur l’historique de recherche

Ces modèles sont généralement utilisés pour représenter les centres d’intérêt de l’utilisateur. Le système MobileIQ [184] se base sur un ensemble de pages web jugés par l’utilisateur pour représenter son profil.

Dans [37, 43], le contexte de l’utilisateur est représenté par la liste de ses requêtes soumises précédemment, la liste de ses préférences et la liste des documents qu’il a visité pour ses requêtes précédentes.

3.2.2.2 Modèles attribut-valeur

Ces modèles représentent la plus simple structure de données pour la modélisation du contexte. L’attribut représente un nom ou un identifiant de l’élément contextuel considéré. Ces attributs peuvent bien prendre des valeurs scalaires (nom de personne, nom de place, le jour, etc) que vectorielles (position géographique, liste de mots, etc). A chaque attribut contextuel, le système peut de plus associer un poids pour déterminer l’importance de l’élément contextuel considéré et exploite un algorithme à correspondance ou de *matching* exacte sur ces attributs.

L'approche de modélisation *attribut-valeur* est utilisée dans le système Jimminy [192] pour la représentation du contexte de documents ici des notes. Cinq attributs sont utilisés pour modéliser le contexte : la localisation physique de l'utilisateur, les personnes à côté, le temps, le sujet de la note et son contenu textuel. Les attributs sont utilisés comme des *triggers* pour la recommandation de notes.

Dans [47], le modèle de contexte proposé est sous forme d'un 4-uplet d'attributs décrivant le contexte d'un utilisateur mobile. Ce modèle est dénoté : $W4model = (who, what, where, when)$. L'attribut "who" modélise une entité active du contexte de l'utilisateur (objet, personne, source d'information, etc). Cet attribut est représenté par une paire *type : valeur* (e.g. *person : Gabriella*). L'attribut "what" modélise l'activité de l'entité décrite par l'attribut "who". Cet attribut est représenté par une paire *predicate : complement* (e.g. "read :book", "work :pervasive computing group"). L'attribut "where" décrit l'aspect localisation du contexte de l'utilisateur. Cet attribut peut prendre des valeurs de coordonnées géographiques, de nom de places, ou même d'expressions de positionnement (e.g. "here", "within :500m", etc). L'attribut "when" associe l'information temporelle sous forme d'instant, d'intervalle ou d'expression temporelle (e.g. "today", "now", etc) au contexte de l'utilisateur.

Dans [266] la situation de l'utilisateur mobile est modélisée par un n-uplet pouvant contenir différents attributs physiques : localisation, temps, luminosité, etc, ou des attributs d'actions (navigation, lecture, etc). Les centres d'intérêts sont représentés par des listes de mots clés pondérés extraits à partir des documents visités par l'utilisateur dans chaque situation.

3.2.2.3 Modèles orientés objet

Le but des approches de modélisation de contexte basées sur le modèle objet est de tirer profit des principaux avantages de la modélisation objet - à savoir l'encapsulation et la réutilisation. Les détails de l'acquisition et la transformation du contexte sont ainsi encapsulés au niveau de l'objet et donc cachés aux autres composants. L'accès à l'information contextuelle est fourni à travers des interfaces spécifiées.

Une approche dans cette catégorie est le modèle d'*objet actif* décrit dans le système GUIDE [50], où chaque objet de l'espace géographique est décrit par un ensemble de propriétés et des références vers d'autres objets et des liens hypertextes vers des documents les décrivant.

Dans [29] le modèle objet est adopté pour représenter un profil situationnel de l'utilisateur, comme illustré dans la figure 2.11. Le profil situationnel est

principalement représenté par l'entité "LocationType" qui représente la localisation sémantique de l'utilisateur, et l'entité "ActivityType" qui représente les centres d'intérêts de l'utilisateur reliés à chaque localisation (des valeurs prises par ces deux entités sont données dans la figure en guise d'illustrations). L'instanciation de ce modèle est basée sur l'observation et l'analyse des déplacements de l'utilisateur mobile pour renseigner les informations de localisation. Pour les activités de l'utilisateur elles sont renseignées par des questionnaires explicites et optimisés pour l'utilisateur mobile. La figure 2.11 est annotée par un graphe des questions indiquant les entités et les propriétés sur lesquels opèrent ces questionnaires.

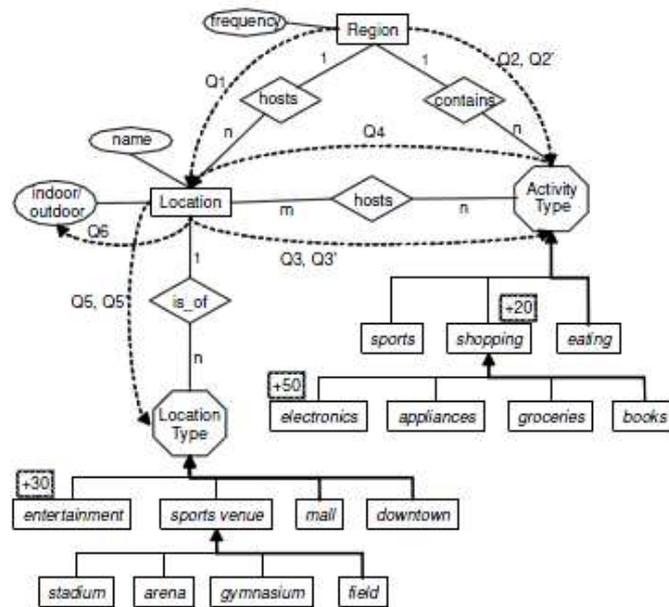


FIG. 2.11 – Un schéma simplifié du modèle objet du profil situationnel.

Une autre approche dans cette catégorie est le modèle de *contexte d'objet* décrit dans [271], où chaque objet possède des propriétés et des relations organiques avec les autres objets de son environnement. La figure 2.12 présente un exemple de ce modèle de contexte d'objet. Les relations entre les objets sont des relations de corrélations pondérées selon des règles de corrélation.

3.2.2.4 Modèles hiérarchiques et de balisages

Ces modèles se basent sur une structure hiérarchique de données généralement annotées par des balises avec des attributs et du contenu. En particulier, le contenu des balises (ou tags) est généralement défini récursivement par d'autres balises.

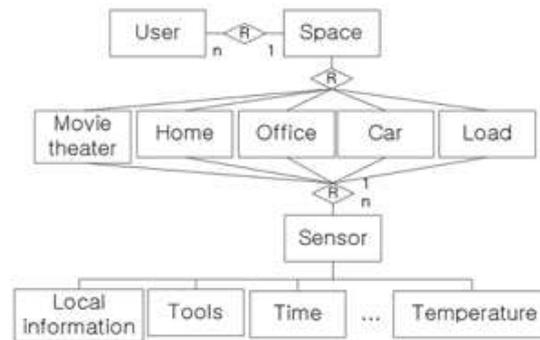


FIG. 2.12 – Exemple du modèle de contexte d'objet.

Le système Stick-e Note [176] est un premier exemple de ces approches, il est basé sur la norme du langage de balisage générique (SGML). Les contextes sont modélisés comme des balises avec des attributs. La note de la balise `<body>` sera automatiquement déclenchée lorsque les contraintes contextuelles dans la balise `<require>` sont remplies.

Le système AmbieSense [166] a choisi de mettre en œuvre une hiérarchie de classes de contextes en se basant sur la technologie XML intégrée avec des classes du Langage Java.

Un autre exemple est le "modèle de contexte en couche" décrit dans [20]. Ce modèle est composé de trois couches. Une première couche responsable de l'acquisition des informations contextuelles directement des capteurs physiques de bas niveau (GPS, horloge système, microphone, etc) ou des capteurs virtuels (activité de l'utilisateur par analyse des mouvements de la souris, des interactions claviers ou tactiles, etc). Une deuxième couche d'abstraction qui est chargée de dériver ou d'inférer des propriétés de plus haut niveau notamment en utilisant des ontologies et des taxonomies sémantiques. Une dernière couche, dite de niveau applicatif, permet de redéfinir et d'entendre les propriétés du contexte selon le besoin propre de l'application.

Dans le système mPERSONA [172], le profil utilisateur est représenté en se basant sur le langage de balisage XML. Des balises sont utilisées pour décrire les préférences des utilisateurs sur les contenus des documents. De plus, des balises temporelles sont utilisées pour décliner les préférences de l'utilisateur sur le contenu en fonction de zones temporelles prédéfinies. La figure 2.13 illustre la représentation d'une partie du profil utilisateur selon ce modèle.

```

<restaurantsTypeByTime categoryWeight="50">
  <timeZone time="0-3">
    <restaurantType name="Fast Food" weight="90"/>
  </timeZone>
  <timeZone time="3-6"></timeZone>
  <timeZone time="6-9"></timeZone>
  <timeZone time="9-12">
    <restaurantType name="Fast Food" weight="90"/>
    <restaurantType name="Cypriot" weight="60"/>
    <restaurantType name="Kebab House" weight="95"/>
  </timeZone>
  <timeZone time="12-15">
    <restaurantType name="Pizzarias" weight="85"/>
  </timeZone>
  <timeZone time="15-18"></timeZone>
  <timeZone time="18-21">
    <restaurantType name="Chinese" weight="90"/>
    <restaurantType name="Mexican" weight="75"/>
  </timeZone>
  <timeZone time="21-24"></timeZone>
</restaurantsTypeByTime>

```

FIG. 2.13 – Exemple de modèle utilisateur avec des balises de préférences de contenu et des balises de zones temporelles.

3.2.2.5 Modèles sémantiques et conceptuels

Les ontologies sont un outil prometteur pour spécifier les concepts et leurs interrelations. L'avantage de ce type de modélisation est sa capacité de partage des connaissances, la possibilité d'inférence logique et de réutilisation de connaissances.

Une première approche qui propose la modélisation du contexte en se basant sur une ontologie est décrite dans [244]. L'ontologie du contexte contient des concepts pour décrire le temps, le lieu, les aspects sociaux, et les caractéristiques du dispositif mobile.

Dans [107] le contexte médical est représenté par une ontologie couvrant le domaine médical en se basant sur la norme HL7 pour le modèle d'information de référence (RIM) et de l'architecture du document clinique (CDA)¹², et une extension pour couvrir le contexte spatio-temporel.

Dans [252, 87] les centres d'intérêt de l'utilisateur sont représentés comme des vecteurs de concepts pondérés. Les concepts utilisés appartiennent au premier niveau de l'ontologie ODP dans [252] et aux quatre premiers niveaux de l'ODP dans [87]. Les poids des concepts dans le profil sont résultants de la classification de l'information jugée pertinente par l'utilisateur dans ce concept. La classification est basée sur l'apprentissage d'un classificateur de l'état de l'art à partir des documents initialement classés sous les concepts de l'ODP.

Dans [155], le profil utilisateur est modélisé comme une collection de sous-profil par thèmes d'intérêts. Chaque thème d'intérêt T est lui même modélisé

¹²www.hl7.org

par un sous-ensemble de concepts à partir de l'ontologie C du domaine du commerce, i.e $T = (c_1, c_2, \dots, c_n) \subset C$. Chaque sous-profil utilisateur $P(U, T)$ est représenté par :

$$P(U, T) = (c, w_{U,T}(c)), c \in C \quad (2.4)$$

où les poids des concepts $w_{U,T}(c)$ sont appris sur la base des jugements de pertinence explicites de l'utilisateur.

Dans [148], le profil utilisateur et le profil de la requête sont représentés par des concepts pondérés combinant des préférences de contenus et de places, appelé OMF- *ontology-based multi-facet profile*. Les concepts dans le profil de la requête et le profil utilisateur sont issus de deux ontologies spatiale et thématique. Le calcul d'une entropie de contenu et de localisation permet de spécifier la nature de la requête ou la préférence de l'utilisateur pour les concepts de contenus ou de localisation. Les entropies de contenus $H_C(q)$ et de localisation $H_L(q)$ de la requête q sont calculées comme suit :

$$H_C(q) = - \sum_{i=1}^k p(c_i) \log p(c_i) \quad H_L(q) = - \sum_{i=1}^m p(l_i) \log p(l_i) \quad (2.5)$$

où k est le nombre de concepts de contenus extraits $C = \{c_1, c_2, \dots, c_k\}$ pour la requête q , $|c_i|$ est le nombre de résultats de la recherche contenant le concept de contenu c_i , $|C| = |c_1| + |c_2| + \dots + |c_k|$, $p(c_i) = \frac{|c_i|}{|C|}$, m est le nombre de concepts de localisation $L = \{l_1, l_2, \dots, l_m\}$ appris pour la requête q , $|l_i|$ est le nombre de résultats de la recherche contenant le concept de localisation l_i , $|L| = |l_1| + |l_2| + \dots + |l_m|$, et $p(l_i) = \frac{|l_i|}{|L|}$. De même les entropies de contenus cliqués $H_{\bar{C}}(q, u)$ et de localisation cliquées $H_{\bar{L}}(q, u)$ par l'utilisateur u pour la requête q sont calculées comme suit :

$$H_{\bar{C}}(q, u) = - \sum_{i=1}^t p(\bar{c}_{iu}) \log p(\bar{c}_{iu}) \quad H_{\bar{L}}(q, u) = - \sum_{i=1}^v p(\bar{l}_{iu}) \log p(\bar{l}_{iu}) \quad (2.6)$$

où t est le nombre de concepts de contenus présents dans les documents cliqués par l'utilisateur u , $\bar{C}_u = \{\bar{c}_{1u}, \bar{c}_{2u}, \dots, \bar{c}_{tu}\}$, $|\bar{c}_{iu}|$ est le nombre de résultats de la recherche contenant le concept de contenu c_i et cliqués par l'utilisateur u , $|\bar{C}_u| = |\bar{c}_{1u}| + |\bar{c}_{2u}| + \dots + |\bar{c}_{tu}|$, $p(\bar{c}_{iu}) = \frac{|\bar{c}_{iu}|}{|\bar{C}_u|}$, v est le nombre de concepts de localisation $\bar{L} = \{\bar{l}_{1u}, \bar{l}_{2u}, \dots, \bar{l}_{mu}\}$ cliqués par l'utilisateur u , $|\bar{l}_{iu}|$ est le nombre de fois que le concept de localisation l_i est cliqué par l'utilisateur u , $|\bar{L}| = |\bar{l}_{1u}| + |\bar{l}_{2u}| + \dots + |\bar{l}_{mu}|$, et $p(\bar{l}_{iu}) = \frac{|\bar{l}_{iu}|}{|\bar{L}|}$. Les entropies ainsi calculées permettent de mesurer l'utilité de la personnalisation selon les deux dimensions de contenu et de localisation selon les formules suivantes :

$$e_C(q, u) = \frac{H_C(q)}{H_{\bar{C}}(q, u)} \quad e_L(q, u) = \frac{H_L(q)}{H_{\bar{L}}(q, u)} \quad (2.7)$$

3.2.2.6 D'autres modèles de structures spécifiques

D'autres modèles propriétaires présentant des structures spécifiques sont aussi développés dans la littérature.

Le système MoBe [161] utilise un modèle de contexte appelé "descripteur de contexte" - *context descriptor*". C'est une structure de données synthétisées à partir des informations contextuelles directement acquises des capteurs mobiles. Le descripteur de contexte contient un identificateur unique, avec des informations contextuelles pondérées selon un degré d'importance et filtrées sur la base de leurs probabilités de vraisemblance. La figure 2.14 illustre le descripteur de contexte construit, certains des attributs sont textuels et d'autres sont structurés.

<p>Name: a context unique identifier.</p> <p>Description: composed by</p> <ul style="list-style-type: none"> • <i>description base</i>: the explicit description of the context, in free text; • <i>description domain</i>: mainly related to the user location: for example, when the user enters the university the description domain of the university is sent to his mobile device. <p>Location: user's spatial position. It contains both low level data, like GPS coordinates, and semantic enriched information, like "office, 2nd floor".</p> <p>Time: date and time referred to the current context. It can contain both numerical values like "11:00pm" and temporal labels like "morning".</p> <p>Activity: composed by</p> <ul style="list-style-type: none"> • <i>activity</i>: the activity description; • <i>activity-type</i>: the activity categorization. <p>Posture: the user posture, like "lying down", "standing", etc. This information is important in connection with usability concepts: on the basis of the posture a user can manage different cognitive loads and so the application should be different. For example an application to take notes with a lot of functions could be useful if the user is sitting, but could be inconvenient if the user is walking.</p> <p>Probability: likelihood of the context (needed because some contexts are inferred with uncertainty).</p> <p>Privacy: a numeric value (predefined by default and modifyable by the user) modeling the privacy of the contextual information. Contexts with privacy value higher than the privacy threshold defined by the user will not be sent outside the mobile device.</p> <p>Importance: a numeric value (predefined by default and modifyable by the user) that allows a context to be sent to remote servers even if its probability is lower than the filtering threshold. The idea is that important contexts, like for example "car is going to break", should be used in the retrieval process even if they have a low probability value.</p>
--

FIG. 2.14 – Descripteur de contexte.

Un autre exemple est le "modèle d'arbre de contexte dimensionnel" décrit dans [108]. C'est un modèle de descripteur de contexte amélioré avec le concept de dimension. Chaque dimension représente un aspect du contexte ou du profil utilisateur. Les dimensions définissent ainsi un espace multidimensionnel où chaque point représente un utilisateur dans un contexte potentiel. Chaque contexte dans l'arbre de contexte est exprimé en termes d'un ensemble de valeurs, une valeur pour chaque sous-dimension. Des contraintes peuvent être définies pour éviter des sous-espaces inutiles correspondants à des instanciations de dimensions non-compatibles. La Figure 2.15 montre un exemple d'arbre de contexte dimensionnel, qui modélise les contextes possibles pour une application de guide touristique pour la visite d'un site archéologique.

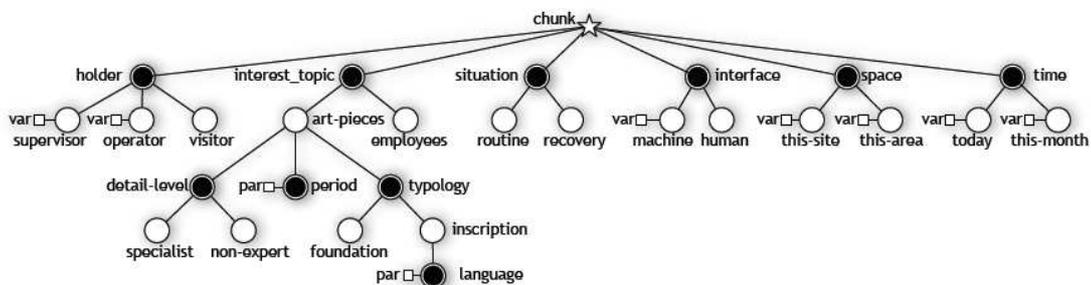


FIG. 2.15 – Modèle d'arbre de contexte dimensionnel.

Bierig [28] propose de modéliser le contexte en utilisant des "schémas de causalité" inspirés de la "théorie d'attribution" [128]. Ce modèle permet de représenter la relation entre les dimensions du contexte, considérées comme des causes, et leurs effets sur les besoins informationnels des utilisateurs mobiles. Le contexte est alors représenté comme un cube en 3 dimensions de temps, de localisation et de centre d'intérêt de l'utilisateur (comme illustré dans la figure 2.16). Ce cube, appelé "schéma de causalité", est interprété comme une forme simplifiée d'une analyse de régression linéaire multiple pour prédire une relation de causalité basée sur des corrélations supposées entre les dimensions modélisées du contexte.

Ying et al [270] propose de modéliser le profil utilisateur par des "séquences sémantiques des trajectoires" de l'utilisateur mobile. Une trajectoire sémantique consiste en une séquence des localisations visitées par l'utilisateur annotées par des balises (*tags*) sémantiques. La figure 2.17 illustre un exemple des trajectoires sémantiques construites. Les balises sémantiques servent à capturer les points d'intérêt¹³ visités par l'utilisateur mobile. Les trajectoires sémantiques

¹³En topographie, un point d'intérêt (POI), en anglais *Point Of Interest*, représente un site utile, un point digne d'intérêt. Ce terme est utilisé par différents logiciels de navigation et appareils GPS ; il désigne un endroit ou destination potentiellement intéressant(e), (extrait de Wikipédia)

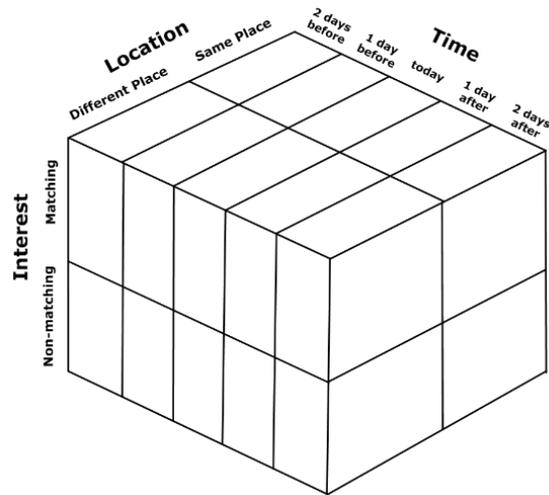


FIG. 2.16 – Schéma de causalité modélisant les facteurs de temps, de centres d'intérêts et de localisation.

ainsi construites sont à la base d'un calcul de similarité d'utilisateurs mobiles pour la recommandation de places à visiter.

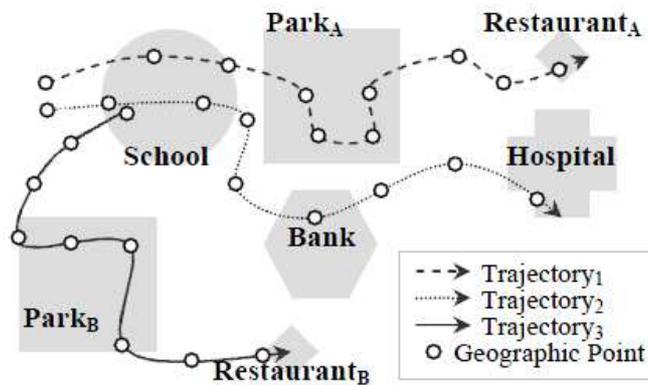


FIG. 2.17 – Exemple de trajectoires sémantiques.

3.3 Synthèse des approches de modélisation du contexte mobile

Nous présentons dans cette section une synthèse des approches de construction et de représentation du contexte de l'utilisateur mobile abordées précédemment. Ces approches sont groupées dans le tableau 2.1 selon le modèle de représentation du contexte, la technique de construction du contexte, ainsi que des exemples de systèmes correspondants.

Représentation	Mode de construction	Exemple
----------------	----------------------	---------

Historique de recherche	accumulation des pages Web sélectionnées au cours des sessions de recherche	le système MobileIQ [184]
	la liste de requêtes de l'utilisateur soumises précédemment, la liste de ses préférences et la liste des documents qu'il a visitée	[37, 43]
Attribut-valeur	classification des situations de recherche de l'utilisateur	Yau et al. [266]
Orientés-objet	clustering des données des Cell ID pour inférer les régions fréquemment visitées par l'utilisateur	Bila et al. [29]
	analyse de corrélations entre les variables contextuelles	Yoon et al. [271]
Hiérarchiques et de balisage	techniques à base de règles pour inférer des contextes abstraits à base des données contextuelles de bas niveaux	Arias et al. [20]
	apprentissage des préférences de l'utilisateur par zones temporelles de la routine journalière de l'utilisateur	le système mPERSONA [172]
Sémantiques et conceptuels	apprentissage d'un RSVM pour obtenir un vecteur de préférences de contenu pondérées et un vecteur de préférences de places pondérées	Leung et al. [148]
	classification supervisée des documents dans des ontologies de domaines prédéfinies	[252, 87]

TAB. 2.1: Synthèse des différents modèles de représentation du contexte en RI mobile

Selon l'évaluation des modèles de représentation de contexte réalisée dans [226], les modèles de représentation basés ontologie sont les modèles les plus expressives, flexibles et extensibles. Plusieurs outils sont disponibles en ligne pour définir, partager, ou réutiliser des ontologies. Notamment des ressources du World Wide Web Consortium, e.g., le langage de description de ressources sémantiques-RDF (e.g. utilisé pour la description de l'ontologie OPD) et le langage du web pour la description d'ontologie-OWL (e.g. utilisé par [171] pour la description d'une ontologie temporelle).

Dans nos travaux de modélisation du contexte nous nous référons à ce modèle de représentation. Plus spécifiquement, nous utilisons l'ontologie ODP pour la modélisation des centres d'intérêts de l'utilisateur mobile. De même nous exploitons des ontologies spatiale et temporelle pour une modélisation sémantique de la situation spatio-temporelle de l'utilisateur mobile.

4 Approches et modèles d'accès contextuel dans un environnement mobile

La phase suivante au processus de modélisation du contexte consiste en son intégration dans le processus d'accès à l'information. Un système d'accès contextuel à l'information, dans un environnement mobile, est donc un système qui intègre une ou plusieurs dimensions du contexte (eg. l'utilisateur, son environnement, les caractéristiques de son dispositif mobile de recherche, etc) dans la tâche d'accès à l'information. Dans la littérature nous distinguons différents systèmes d'accès contextuel à l'information pour la recherche d'information mobile selon deux principaux points de vue : la tâche d'accès à l'information et le type de contexte exploité.

Dans cette section nous présentons en premier lieu les principales tâches d'accès contextuel à l'information dans un environnement mobile, puis nous donnons un aperçu des modèles d'accès par type de contexte exploité, nous terminons par une synthèse de ces modèles.

4.1 Tâches d'accès à l'information en RI mobile et contextuelle

L'intégration du contexte dans le processus d'accès à l'information dans l'environnement mobile a permis de développer différents tâches ou techniques d'accès en plus de la tâche traditionnelle de recherche interactive, tels que : des systèmes de recommandation, des systèmes de filtrage d'information, et des services-basés localisation.

On peut classer ces systèmes selon deux tâches principales : la tâche d'accès interactive et la tâche d'accès proactive.

1. **La tâche d'accès interactive** : la tâche de recherche est toujours initiée par la requête utilisateur comme dans la RI classique. Deux façons ensuite sont utilisées pour intégrer le contexte au processus de recherche.

Soit la requête est combinée au contexte courant pour dériver une nouvelle requête et lancer ensuite la recherche sur la collection de documents [266, 94, 35]. Soit la requête est lancée au SRI puis les résultats retournés sont ré-ordonnés en tenant compte d'un score contextuel [252, 87] ou filtrés selon un seuil contextuel [165].

2. **La tâche d'accès proactive** : différents travaux [40, 118, 145] ont soutenu la thèse qu'une démarche de recherche proactive guidée ou initiée par le contexte serait prometteuse dans l'environnement mobile. Deux approches sont identifiées dans [40] : approche axée-utilisateur et approche axée-document. Dans la première approche, sur la base du contexte utilisateur une requête sera formulée et envoyée automatiquement au SRI (e.g. [161]). Dans la seconde approche, chaque document dans la collection a un contexte déclencheur (*trigger*) ; s'il correspond au contexte courant de l'utilisateur alors le document sera envoyé à l'utilisateur (e.g. [176, 192]).

Certains travaux proposent d'offrir une interface de recherche à tâche multiples, offrant à l'utilisateur mobile la possibilité de choisir entre la tâche de recherche interactive ou proactive [50, 86].

4.2 Les modèles d'accès par type de contexte

Selon la dimension du contexte modélisé, nous pouvons distinguer plusieurs approches : techniques d'adaptation au dispositif mobile, à la localisation de l'utilisateur, au contexte social, et les techniques d'adaptation multidimensionnelle. Nous présentons dans ce qui suit quelques approches et techniques représentatives des différents types d'accès contextuels à l'information en RI mobile. Certains travaux peuvent appartenir à plusieurs approches.

4.2.1 Techniques d'adaptation au dispositif mobile

Cette catégorie de travaux a abordé les questions liées à l'adaptation de la recherche aux contraintes imposées par les fonctionnalités limitées des appareils mobiles. Le but principal de ces travaux est de faciliter la saisie des requêtes, d'adapter l'affichage des pages web sur les dispositifs mobiles de capacités très variables et d'optimiser l'utilisation des plateformes de connexions et de couverture des réseaux sans fils. Des approches sont proposées pour optimiser l'accès à l'information sur les réseaux mobiles, adapter la visualisation de la liste des résultats de recherche et pour faciliter la saisie des requêtes.

4.2.1.1 Optimiser l'accès à l'information sur les réseaux mobiles

Cette catégorie de travaux a abordé les questions liées à l'optimisation de l'exploitation des bandes passantes mobiles et aux problèmes de la perte de connexion liée à la mobilité de l'utilisateur et les ruptures de connexion.

Dans un scénario de recherche d'information proactive, les auteurs dans [133] exploitent les informations contenues dans le calendrier de l'utilisateur et l'historique du *feedback* utilisateur pour chercher en amont les pages web susceptibles d'intéresser l'utilisateur. Ces pages Web sont ensuite chargées a priori (*pre-cached*) sur l'appareil mobile de l'utilisateur à chaque nouvelle connexion. Dans le but d'améliorer les temps de réponse et de réduire les délais de transmission sur les réseaux mobiles, le système MobileIQ [184] effectue le pré-chargement des pages web (en suivant leurs URLs) sur le proxy en se basant sur les évaluations faites par les utilisateurs les ayant déjà visitées. Si une URL possède une évaluation faible, alors seul son contenu de premier niveau est chargé, alors que les URLs qui sont très bien notées sont téléchargées jusqu'à des niveaux de n , où n est un paramètre système configurable. Le contenu pré-chargé est sauvegardé dans le cache du serveur proxy où l'utilisateur est actuellement connecté.

4.2.1.2 Adapter la visualisation de la liste des résultats de recherche

Cette catégorie de travaux porte sur l'adaptation de l'affichage de la liste des résultats de la recherche sur les dispositifs mobiles de taille réduite afin d'éviter à l'utilisateur d'avoir à naviguer dans de longues listes. Les principales approches proposées se basent sur des techniques de résumé, de regroupement des résultats de recherche en des catégories thématiques, ou encore les listes des résultats sont remplacées par une projection des documents dans des interfaces de cartes géographiques.

Dans [48], une méthode de résumé des titres des résultats de recherche (des communiqués de presse de Microsoft) est proposée. La méthode se base sur la combinaison d'une technique d'abréviation et d'une technique de suppression de mots insignifiants. L'abréviation se base sur le remplacement des mots du titre par des abréviations ou des acronymes issus d'un dictionnaire du domaine. La suppression se base sur une mesure de signification d'un terme dans le titre. Elle est exprimée par sa fréquence d'apparition dans le titre et le nombre de paragraphes dans lesquels il apparaît dans le document.

Dans [228], une méthode de *résumé hiérarchique biaisé-requête* est utilisée pour la présentation d'articles de journaux pour des utilisateurs de téléphones mobiles. La méthode se base sur la combinaison de plusieurs techniques d'extraction de phrases qui exploitent des informations extraites de la collection des documents et des requêtes passées.

Le système Hyponym [21], propose une technique pour l'organisation et l'affi-

chage des résultats sur les appareils mobiles de zones d'affichage limitées. En se basant sur un algorithme de *clustering*, chaque requête utilisateur est classée en des groupes de requêtes. Les URLs retournées par le SRI sont aussi classées pour chaque groupe de requêtes. L'affichage des résultats de la recherche est organisé en deux écrans : dans un premier temps les groupes de la requête sont affichés dans l'ordre décroissant de l'appartenance de la requête à chaque groupe. Une fois l'utilisateur aura cliqué sur un groupe, la liste des URLs classées sous ce groupe est affichée dans l'ordre décroissant de la probabilité d'appartenance des URLs à ce groupe.

Une autre méthode de présentation des résultats de recherche sous des classes est présentée dans le système Findex [96]. La présentation des résultats de recherche est divisée sur deux écrans : un écran représentant les classes des résultats et un écran représentant les *snippet* des documents classés sous la catégorie sélectionnée par l'utilisateur. La spécificité de cette méthode, est que les noms des classes des documents sont représentés par des mots simples (extraits des *snippets*) et non par des noms de catégories sémantiques qui peuvent être peu significatifs pour l'utilisateur.

Dans le système HyCon [35] et dans [51], l'affichage des résultats de la recherche est présenté sous forme de points sur une carte géographique centrée autour de la localisation de l'utilisateur. Pour voir les contenus des points, l'utilisateur peut passer la main sur un point pour afficher le *snippet* décrivant ce résultat. Si l'utilisateur juge le *snippet* intéressant, il peut cliquer pour afficher tout le document. Cette méthode a le double avantage de permettre à l'utilisateur d'avoir une première appréciation des résultats de la recherche par rapport à leur proximité géographique et d'éviter le défilement de la liste des résultats, mais suppose des interfaces tactiles.

4.2.1.3 Faciliter la saisie des requêtes

Cette catégorie de travaux vise à l'optimisation des interfaces de saisie de requêtes par des techniques de complétude de mots, des zones de sélection graphiques, proposition de mots via des thésaurus, ou sont complètement remplacées par un service de formulation de requêtes orales.

Kamvar et al. [124] proposent des suggestions de mots ou de plusieurs mots pour faciliter la saisie de la requête à l'utilisateur. Plusieurs modèles de suggestions sont proposés et sont combinés linéairement pour améliorer la qualité des prédictions. Les modèles de suggestion sont basés sur des probabilités de co-occurrence de préfixes de mots dans un corpus de requêtes mobiles et de mots contextuels représentant la localisation de l'utilisateur.

Tezuka et al. [242], propose de faciliter la saisie des requêtes aux utilisateurs mobiles en offrant à l'utilisateur une interface sophistiquée. En effet, en plus de la forme traditionnelle de saisie de mots, l'utilisateur peut spécifier une région par un simple clic sur une carte graphique, et peut de plus indiquer le temps

par simple interaction avec une métaphore représentant une horloge. Cette interface soulage ainsi l'utilisateur du fardeau de saisir plus de mots pour affiner sa recherche, notamment des mots relatifs au lieu et au temps.

Une interface basée facettes est proposée dans le système FaThumb [125] pour faciliter la saisie des requêtes et la navigation aux utilisateurs sur les téléphones mobiles à clavier lors de la recherche et l'accès à l'information en utilisant une collection de pages jaunes. L'interface de recherche offre à l'utilisateur une zone de navigation à 8 facettes : catégorie, place, distance, prix, évaluation, heures, raccourcis et favoris. Chaque facette est elle même organisée en hiérarchie à plusieurs niveaux. Chaque choix de l'interface est lié à une touche du clavier. L'approche dans [52] propose également une interface sophistiquée à l'utilisateur pour lui éviter le plus possible de saisir des requêtes. L'interface de recherche intègre les requêtes et les résultats de recherche passés des utilisateurs et les présente en les situant sur une carte centrée sur la localisation courante de l'utilisateur. L'utilisateur peut facilement interagir avec l'interface en cliquant sur une requête pour la relancer ou pour visualiser les documents visités par d'autres utilisateurs. Deux barres horizontales permettent de plus à l'utilisateur de naviguer dans l'historique de recherche dans le temps, ou de choisir des résultats plus spécifiques ou plus généraux en relation à deux catégories de requêtes : tourisme ou divertissement.

Arias et al. [20] proposent d'utiliser un thésaurus et une ontologie de domaine pour assister l'utilisateur lors de la saisie de sa requête. Les entrées du thésaurus sont utilisées pour recommander des suggestions de mots à fur et à mesure de la saisie de la requête. Les entrées de l'ontologie sont utilisées pour affiner et étendre la requête de l'utilisateur. Si l'utilisateur accepte de sélectionner une entrée sémantique de l'ontologie, un petit formulaire simple apparaît sur son interface lui permettant de choisir ses préférences par rapport aux attributs liés à l'entrée sélectionnée.

Le système DSR [37] contourne le problème de la difficulté de saisie des requêtes sur les mobiles en proposant une interrogation par des requêtes orales (parlées) formulées dans le langage naturel. Plusieurs modèles acoustiques sont entraînés sur une base de données contenant des échantillons de données téléphoniques. Pour la modélisation du modèle de langue des requêtes orales, le système utilise une grammaire et des règles qui définissent un ensemble réduit de structures de phrases.

4.2.2 Techniques d'adaptation au contexte utilisateur

Le but principal de ces travaux est d'utiliser le contexte utilisateur pour personnaliser les résultats de la recherche en tenant compte des centres d'intérêt de l'utilisateur et de ses préférences. On distingue trois approches principales de l'utilisation du profil utilisateur : re-formulation des requêtes, appariement documents-profil ou ré-ordonnement des résultats de la recherche.

4.2.2.1 Modèles de reformulation de la requête

Dans ce type d'approches, les profils utilisateurs sont utilisés pour étendre les requêtes initiales soumises au SRI.

Dans Yau et al. [266], le profil utilisateur est utilisé pour reformuler la requête initiale de l'utilisateur. La personnalisation se fait en deux étapes : d'abord identifier la situation courante de l'utilisateur en utilisant un algorithme de classification, puis utiliser le vecteur de termes représentant les centres d'intérêts dans le profil utilisateur, lié à la situation identifiée, pour étendre la requête initiale de l'utilisateur. La figure 2.18 illustre un exemple de profils construits et leur utilisation pour l'extension des requêtes des utilisateurs.

User	Interests	Profile	Situation	Initial Request	Modified Request
Alice	1) Research information on <i>Marble Mountain Terrane</i> , which is Alice's major focus.	1)<InField, DayTime> => keywords (<i>Ultramafic rocks, Marble Mountain Terrane</i>).	<InField, DayTime>	Marble	Marble Mountain Terrane <and> Ultramafic rocks
	2) <i>Marketing information</i> about <i>Marble</i> since her laboratory need to buy some marble for experiments.	2)<AtNight> => Keywords (<i>Marble, Marketing information</i>).	<AtNight>	Marble	Marble <and> Marketing information
Bob	1) Research information on <i>Precambrian Marble</i>	1)<RegionX> =>Keywords (<i>Ultramafic rocks, Precambrian Marble</i>)	<RegionX>	Marble	Precambrian Marble <and> Ultramafic rocks
	2) Research information on <i>Cambrian Marble</i> .	2)<RegionX> =>Keywords (<i>Ultramafic rocks, Cambrian Marble</i>)	<RegionY>	Marble	Cambrian Marble <and> Ultramafic rocks

FIG. 2.18 – Exemple de profils construits dans [266] et leurs exploitation pour la reformulation des requêtes.

Dans [43], le profil utilisateur est utilisé dans le processus de formulation/reformulation de requêtes orales. Dans la phase de formulation de la requête orale, le profil à long terme, fourni explicitement par l'utilisateur et noté P_{LT} , est utilisé pour sélectionner parmi les résultats du processus de reconnaissance automatique de la parole (ASR), celui qui est le plus probable. Pour cela chaque sortie du ASR est comparée au profil P_{LT} en utilisant une fonction de similarité floue. Les poids des termes dans le profil P_{LT} sont fixés à 1. La sélection de la sortie du ASR est faite sur la base d'une combinaison (par produit) de sa probabilité retournée par le ASR et son score de similarité avec le profil P_{LT} . Le profil P_{LT} est ensuite utilisé pour l'extension de la requête Q de l'utilisateur avant de lancer la recherche avec la nouvelle requête $Q + P_{LT}$. Dans la phase de reformulation, les n top résultats de la recherche, considérés comme un profil à court terme (noté P_{ST}), sont utilisés dans un processus de *feedback* automatique pour étendre la requête qui devient alors $Q + P_{LT} + P_{ST}$. Le poids S d'un terme t_i dans le profil P_{ST} est calculé selon la fonction suivante :

$$S_{Term}(t_i) = Occ_i + \sum_{j=1, j \neq i, s_{ij} > \delta}^k s_{ij} \quad (2.8)$$

où Occ_i est le nombre d'occurrences du terme t_i dans les n top résultats de la recherche, s_{ij} et le nombre de relations sémantiques qu'a le terme t_i avec

les autres termes dans les n top résultats de la recherche et δ est un seuil fixé expérimentalement. Les termes dans le profil P_{ST} sont de plus associés à un indicateur temporel qui fixe leur durée de vie dans le profil à court terme avant de passer dans le profil à long terme.

4.2.2.2 Modèles d'appariement documents-profiles

Dans ce type d'approches, les profils utilisateurs sont construits au sein même du processus de recherche, et sont utilisés pour ordonner ou filtrer les documents à renvoyer ou à recommander à l'utilisateur selon ses centres d'intérêts. Dans le système Magitti [27], le profil utilisateur représenté par une liste d'activités et leur probabilités sont utilisées pour recommander à l'utilisateur des documents pertinents à chaque type d'activité. Pour cela, les documents de la collection sont annotés par les types d'activités qu'ils couvrent.

Dans [174], le profil utilisateur est utilisé pour la recommandation d'information d'actualité (des *News*) à des utilisateurs mobiles. L'adaptation des *News* se fait d'abord en amont du côté serveur, où les informations sont filtrées par appariement des *News* avec les préférences de l'utilisateur exprimées dans son profil à long terme. L'appariement se base sur la mesure de similarité cosinus entre le vecteur de catégories des *News* (résultat d'un algorithme de classification), et le vecteur de préférences de l'utilisateur exprimées sur ces mêmes catégories. Seuls les 100% (resp. les top 50%, les top 30%) des *News* ayant le degré de préférence haut (resp. moyen, faible) sont envoyés au dispositif mobile de l'utilisateur. Le profil utilisateur à court terme est ensuite utilisé pour l'ordonnement des *News* après cette première phase de filtrage du côté serveur. L'adaptation du côté client se base sur une mesure de similarité cosinus entre les représentations plus fines des *News* et du profil utilisateur à court terme, comme des vecteurs de noms et d'entités nommées pondérés par une formule TF*IDF adaptée. Les *News* sont ainsi retournés à l'utilisateur classés par ordre décroissant de leurs similarité avec son profil à court terme.

Dans [155], un utilisateur est représenté par un ensemble de profils par thèmes. A chaque connexion de l'utilisateur, celui-ci spécifie un thème d'intérêt actuel, ensuite le profil correspondant dénoté $P(U, T)$ est utilisé pour sélectionner les documents (ici des articles de journaux) à lui recommander. Les documents de la collection sont également représentés par des profils conceptuel. La sélection se base sur une mesure de similarité entre le profil du document $P(D)$ et le profil utilisateur désignée par "linear topic-specific document-value function" et donnée par la formule suivante :

$$V(D, T) = \sum_{c \in T} w_D(c) \cdot w_{U, T}(c) \quad (2.9)$$

où $w_D(c)$ et $w_{U, T}(c)$ sont les poids du concept c respectivement dans le profil du document $P(D)$ et dans le profil utilisateur $P(U, T)$.

Le système MobileIQ [184] se base sur un profil utilisateur pour personnaliser les résultats de la recherche. Chaque requête de l'utilisateur est issue d'une catégorie sémantique. Les pages web sont aussi classées par catégorie, et ont une évaluation de pertinence par chacun des utilisateurs les ayant précédemment visitées. Le système MobileIQ retourne une liste d'URLs en réponse à une requête d'un utilisateur en se basant en premier lieu sur les évaluations de pertinence issues de son profil. Si l'utilisateur n'a pas jugé assez d'URLs dans la catégorie de la requête courante, alors d'autres pages web seront retournées sur la base des évaluations numériques dans le profil du *cluster* de l'utilisateur. S'il n'y a pas assez de résultats préalablement jugés pour la requête de l'utilisateur, alors sa requête sera soumise au moteur de recherche Google, et les résultats retournés par Google viendront compléter la liste initiale des résultats de recherche issue du profil utilisateur et/ou de son *cluster*.

Dans Panayiotou et al. [172], le profil utilisateur est utilisé pour personnaliser un portail de services web pour l'utilisateur mobile. Après que l'utilisateur choisit une catégorie générale de services, son profil est comparé aux documents représentant les descriptions des services web contenues dans cette catégorie. Chaque document est représenté par une structure XML, où chaque nœud décrit un concept et ses caractéristiques. Le profil utilisateur contient pour chaque concept et chacune de ses caractéristiques un poids appris pour chaque zone temporelle de la journée. Les documents sont finalement ordonnés sur la base de la moyenne des préférences de l'utilisateur pour tous les nœuds du document pour la zone temporelle courante. Si le score d'un document est inférieur à un seuil de sélection spécifié par l'utilisateur, il sera omis de la liste des résultats qui lui seront retournés.

Dans Ying et al. [270], le profil utilisateur est exploité pour lui recommander des amis potentiels dans un réseau social. Dans ce contexte, le profil d'un utilisateur donné est assimilé à une requête, les profils des autres utilisateurs dans le réseau social sont assimilés à des documents. L'appariement entre la requête et les documents ainsi considérés, se fait sur la base d'une fonction de similarité de trajectoires (*MSTP Similarity-Maximal Semantic Trajectory Pattern Similarity*). La fonction de similarité entre deux utilisateurs est finalement donnée comme la moyenne pondérée de la similarité entre chaque deux trajectoires de l'ensemble des trajectoires contenues dans les profils des deux utilisateurs. Les résultats de cette fonction de similarité sont utilisés pour recommander à un utilisateur cible la liste des top k plus similaires utilisateurs dans le réseau social.

4.2.2.3 Modèles de ré-ordonnement des résultats de la recherche

Ce type d'approches, récupère la liste de documents retournés à partir d'un système standard de recherche- non personnalisé, et personnalisent cette liste en aval en tenant compte des centres d'intérêt de l'utilisateur.

Dans Varma et al. [252] et Kapil et al. [87], le profil utilisateur est utilisé pour ré-ordonner les résultats de la recherche. Ils proposent de combiner linéairement le rang initial d'un document retourné par un SRI classique avec un score de personnalisation calculé entre le document et le profil de l'utilisateur, comme suit :

$$\text{match}(user_i, result_j) = a \cdot \text{score}(user_i, result_j) + (1 - a) \cdot \text{initialRank}(result_j) \quad (2.10)$$

où $a \in [0, 1]$. Le score de personnalisation $\text{score}(user_i, result_j)$ est calculé pour chaque document avec le profil de l'utilisateur en se basant sur une mesure de similarité entre le vecteur de concepts pondérés du document et du profil utilisateur également représenté comme vecteur de concepts pondérés de l'ODP, donné par :

$$\text{score}(user_i, result_j) = \sum_{k=1}^N wp_{ik} \cdot wd_{jk} \quad (2.11)$$

où wp_{ik} est le poids du concept k dans le profil utilisateur i , wd_{jk} est le poids du concept k dans le résultat j , N est le nombre de concepts considérés. Pour cela, les documents sont préalablement classés en utilisant les concepts de premiers niveaux de l'ontologie ODP.

Dans [148], une fonction de ré-ordonnement linéaire est également adoptée pour adapter les résultats de la recherche à l'utilisateur. Le score initial du document d est multiplié par un score de personnalisation comme suit :

$$f(q, d) = \overrightarrow{w_{q,u}} \cdot \phi(q, d) \quad (2.12)$$

où q est une requête, $\phi(q, d)$ est une fonction d'ordonnement classique, $\overrightarrow{w_{q,u}}$ est un poids représentant le score de personnalisation. Il est lui même une combinaison linéaire sous forme de somme pondérée des préférences de contenus et de localisation de l'utilisateur.

4.2.3 Techniques d'adaptation au contexte spatial

Le but principal de ces techniques est l'adaptation des résultats de la recherche au contexte spatial de l'utilisateur. Cette adaptation consiste à retourner des documents qui décrivent la place ou l'endroit où est situé l'utilisateur, ou qui se rapportent à des objets, des services qui se situent dans un périmètre proche de la localisation actuelle ou future de l'utilisateur. Toute adaptation à la localisation de l'utilisateur suppose une indexation spatiale ou du moins une annotation géographique de la collection de documents. L'appariement entre la localisation de l'utilisateur et la localisation associée à un document se fait soit par une simple correspondance, généralement dans le cas d'une annotation textuelle, ou par un calcul d'un score géographique des documents dans le cas d'une indexation géographique. Dans le second cas, des méthodes et des

techniques du domaine de la RI géographique sont alors adoptées. La majorité des travaux sur la RI mobile ne donne pas de précision concernant le processus d'indexation spatiale utilisée ou les fonctions d'ordonnancement spatial adoptées. Les moteurs de recherche commerciaux sur le web tel Google, Yahoo, Bing ont introduit une version mobile de leurs services de recherche locale. Ces moteurs de recherche locaux se focalisent principalement sur une collection de services locaux, e.g. des restaurants, des hôtels. Cependant, les algorithmes utilisés ne sont pas publiés. Peu de travaux [44, 59, 45], ont développé leurs propres modèles d'indexation géographique. L'adaptation au contexte géographique de l'utilisateur est généralement utilisée selon deux approches : re-formulation des requêtes ou appariement documents-requêtes en incluant la dimension localisation.

4.2.3.1 Modèles de reformulation de la requête

Dans ce type d'approches, le contexte spatial est utilisé pour étendre les requêtes initiales soumises au SRI.

Le système HyCon [35], utilise la localisation de l'utilisateur pour reformuler sa requête initiale avant de l'envoyer au moteur de recherche Google. La re-formulation se fait par simple ajout de l'adresse à laquelle se trouve l'utilisateur aux mots de sa requête initiale. Cela permet de renvoyer des résultats de recherche centrés autour de la localisation courante de l'utilisateur mobile. Dans [94], une autre méthode de re-formulation de la requête en tenant compte de la localisation de l'utilisateur est proposée. Des termes relatifs à la localisation de l'utilisateur (e.g. noms de places, d'activités liés à cette place) sont ajoutés à sa requête initiale avec un opérateur logique "AND". Les termes ajoutés sont pondérés en tenant compte de leurs probabilités de co-occurrence avec les mots des documents retournés par la requête initiale et leurs probabilité dans l'ensemble des documents du corpus cible.

4.2.3.2 Modèles d'appariement documents-localisation

Dans ce type d'approches, la localisation de l'utilisateur est utilisée comme un facteur de sélection ou d'ordonnancement des documents à retourner à l'utilisateur.

Dans les systèmes Jimminy [192] et Stick-e [176] des notes sont filtrées sur la base de la localisation de l'utilisateur et sont recommandées à l'utilisateur.

Dans le système MobileIQ [184] l'adaptation à la localisation de l'utilisateur est effectuée uniquement sur certaines catégories de recherche. Ces catégories de recherche concernent des services web tels que : la météo, le cinéma, etc. Le système restreint alors les résultats de ces catégories de recherche à ceux couvrants la localisation actuelle de l'utilisateur.

Dans [165], la localisation des utilisateurs (ici visiteurs de parc) est prise en compte pour la recherche d'information liées au parc visité. Les documents de la collection (liés au parc) sont manuellement annotés par les coordonnées géographiques de la région géographique du parc qu'ils décrivent. L'adaptation à la localisation de l'utilisateur se fait en exploitant un mécanisme de filtres sur les résultats retournés par un SRI standard pour la requête initiale de l'utilisateur. Deux filtres sont implémentés, ils se basent sur la proximité géographique du résultat : en terme de proximité spatiale (distance euclidienne) de la localisation actuelle de l'utilisateur, ou en terme de proximité à une destination prochaine de l'utilisateur exprimée en terme de temps nécessaire à l'atteindre. L'utilisateur spécifie lui même la distance et le temps d'atteinte servant de filtres aux documents retournés par le SRI.

Dans les guides touristiques CRUMPET [183], GUIDE [50], COMPASS [251], Cyberguide [12] l'adaptation à la localisation de l'utilisateur se fait par filtrage des services ou des informations à recommander à l'utilisateur selon sa localisation.

Dans le système de recherche basé localisation décrit dans [45], des documents (plans et informations sur des pièces dans des bâtiments) sont ordonnés selon leur distance euclidienne de la position actuelle de l'utilisateur et l'orientation spécifiée dans sa requête.

Dans [229], Kokono un système de recherche basée localisation est présenté. Dans ce système, la localisation de l'utilisateur est le seul critère d'ordonnement des résultats de la recherche. Pour cela, les pages web sont indexées par des coordonnées géographiques exprimées en degré de latitude longitude. De même la localisation de l'utilisateur est aussi représentée par des coordonnées géographiques. Les documents sont alors ordonnés selon un degré de recouvrement avec la localisation de l'utilisateur.

Dans [59], des entités géographiques (notée O) sont ordonnées par une fonction de combinaison linéaire d'un score géographique de leur localisation (notée $O.loc$) par rapport à la localisation de l'utilisateur (exprimée dans sa requête et notée $Q.loc$) et d'un score textuel du document la décrivant (noté $O.doc$) par rapport aux mots clés de la requête (noté $Q.keywords$).

La fonction d'ordonnement pour une entité géographique O par rapport à une requête Q est donnée par :

$$D_{ST}(Q, O) = \alpha \frac{D_\epsilon(Q.loc, O.loc)}{maxD} + (1 - \alpha) \left(1 - \frac{P(Q.keywords|O.doc)}{maxP} \right) \quad (2.13)$$

où $\alpha \in [0 1]$ est un paramètre utilisé pour équilibrer la pertinence spatiale et textuelle, la distance euclidienne $D_\epsilon(Q.loc, O.loc)$ entre la requête Q et l'entité géographique O est normalisée par le facteur $maxD$ exprimé par la distance maximale entre les entités géographiques, la probabilité $P(Q.keywords|O.doc)$ est calculée en utilisant les modèles de langue, et $maxP$ est utilisé pour normaliser le score de probabilité dans l'intervalle $[0 1]$.

4.2.4 Techniques d'adaptation au contexte social

Le but principal de ces techniques est l'adaptation des résultats de la recherche au contexte social de l'utilisateur. La particularité de ces travaux réside dans la richesse des ressources pouvant être exploitées pour l'acquisition du contexte social (e.g. à partir de l'appareil mobile via la liste des contacts de l'utilisateur et des réseaux sociaux auxquels il est abonné, ou via des dispositifs matériels de localisation dont on peut inférer les personnes à proximité ou similaires à l'utilisateur).

Une première utilisation simpliste du contexte social est présenté dans le système Jimminy [192]. L'utilisation du contexte social de l'utilisateur se résume à des recommandations de notes ayant été produites dans la présence des mêmes personnes.

Dans le moteur de recherche multimédia et mobile Zurfer¹⁴ [99] le contexte social de l'utilisateur est utilisé pour filtrer les photos créées par les amis (issus du site Flickr) d'un utilisateur, ou les commentaires laissés par ses amis sur ses propres photos.

Le moteur de recherche mobile Taptu¹⁵ [239] utilise une technologie de "recherche communautaire assistée" - *social-assisted search*. En effet, les résultats de la recherche sont générés algorithmiquement, mais ils sont classés et triés sur la base d'opinions ou d'évaluations sociales. Les opinions ou évaluations sociales sont apprises à travers des indices et méta-données extraites des réseaux sociaux, tels que : le nombre de téléchargements, nombre de partages, nombre de *bookmarks*, nombre de visites, des évaluations explicites par des utilisateurs (*ratings*). L'évaluation sociale peut intervenir sur deux niveaux : 1) une pertinence générale "World thinking" issue de l'agrégation des évaluations de toute la communauté dans un réseau social, et 2) une évaluation de groupe issue des évaluations du groupe ou des amis de l'utilisateur. La recherche sociale ou communautaire se base sur la combinaison en temps réel des résultats de recherche issus de différents réseaux communautaires : de Twitter à Facebook en passant par MySpace, Flickr ou Digg. Une extension de la recherche sociale consiste à offrir à l'utilisateur la possibilité de partager les résultats de sa recherche (qu'il juge pertinents) avec un ou plusieurs amis qu'il aurait importé de sa boîte email ou d'un réseau social tel que Twitter.

Dans le système SSB [51], le contexte social de l'utilisateur, représenté par son groupe d'amis sur le site Facebook, est intégré dans le processus de RI locale à plusieurs niveaux : 1) les requêtes issues de son groupe d'amis sont marquées et visualisées sur l'interface de recherche, 2) les réponses émises pour la requête par un ou plusieurs amis de l'utilisateur sont marquées et intégrées à la liste des résultats de recherche locale retournés par le service de recherche locale Google¹⁶ et les résultats de recherche d'événements retournés par Eventful

¹⁴<http://zurfer.research.yahoo.com>

¹⁵<http://taptu.mobi/home>

¹⁶<http://code.google.com/apis/ajaxsearch/local.html>

API¹⁷, et 3) un filtre d'amis permet de filtrer les résultats de la recherche selon un degré d'amitié estimé par le nombre de notes (*posts*) et de commentaires échangés entre les deux amis sur Facebook.

Dans [27] et dans [187], le contexte social de l'utilisateur est utilisé pour la recommandation d'événements sociaux et d'informations pour les utilisateurs mobiles pour répondre au problème de "profil vide". Dans [27], la recommandation d'informations exploite des modèles d'activités des utilisateurs issus de la même population démographique et des modèles de filtrage collaboratifs basés sur des évaluations explicites des utilisateurs pour mesurer leurs similarités. Les informations sont alors classées sur la base des évaluations d'autres utilisateurs similaires. Dans [187], la recommandation d'événements est basée sur une mesure de popularité de l'événement au sein du voisinage de l'utilisateur.

4.2.5 Techniques d'adaptation multidimensionnelle

Plusieurs travaux de la littérature (y compris certains travaux parmi les travaux cités plus haut), ont proposé une adaptation multidimensionnelle. On distingue deux principales approches d'adaptation : avec fusion de types de contexte ou avec distinction de types de contexte.

Les approches par fusions des types de contexte intègrent toutes les données contextuelles sans différenciation de leur type. Cela se base sur un modèle unifié des contextes généralement des ontologies contextuelles. Les mots, valeurs ou concepts représentant des éléments contextuels sont ensuite utilisés pour étendre la requête de l'utilisateur sans distinction de leurs types, avec ou sans techniques de pondération [20, 90].

Les approches par distinction de types de contexte proposent des adaptations par contexte individuel puis éventuellement combinent (e.g. linéairement) les scores contextuels individuels lorsque on peut définir un score contextuel du document relativement aux contextes considérés [27, 44, 28, 139]. Dans ces approches, un principal problème qui se pose alors est : comment combiner les pertinences individuelles d'un document issues de plusieurs dimensions contextuelles pour définir un seul score de pertinence du document ?.

Deux catégories principales de combinaisons sont proposées dans la littérature : la combinaison visuelle et la combinaison des scores. Ci-dessous nous donnons un aperçu des principaux travaux dans chacune de ces deux catégories.

¹⁷<http://eventful.com/>

4.2.5.1 Combinaison visuelle

Dans les travaux proposant la combinaison visuelle, la pertinence des résultats de la recherche est représentée par des éléments graphiques généralement des métaphores représentant plusieurs dimensions contextuelles.

Dans le système Zurfer [99], les documents, ici des images, sont affichés dans plusieurs canaux contextuels (comme illustré dans la figure 2.19) selon leurs pertinences suivant les dimensions : de localisation, sociale, personnelle et thématique. A l'intérieur de chaque canal, les images sont généralement triées par la dimension temps, de sorte que les photos les plus récentes apparaissent en premier.

Dans [52, 51], les résultats de la recherche sont affichés sur une carte graphique centrée sur la localisation de l'utilisateur, sur sa communauté et sont la réponse aux requêtes les plus récentes dans le temps (comme illustrée dans la figure 2.20). L'utilisateur peut interagir avec la carte pour changer la localisation spatiale. De plus, deux filtres (métaphores) représentant une échelle temporelle et une échelle de spécificité/généralité selon la communauté de l'utilisateur, permettent à l'utilisateur de voir d'autres résultats de recherche en manipulant ces métaphores. Un autre filtre social [51] permet de filtrer les résultats de la recherche selon le degré d'amitié des utilisateurs.



FIG. 2.19 – Un aperçu de l'interface d'affichage par canaux contextuels selon [52].

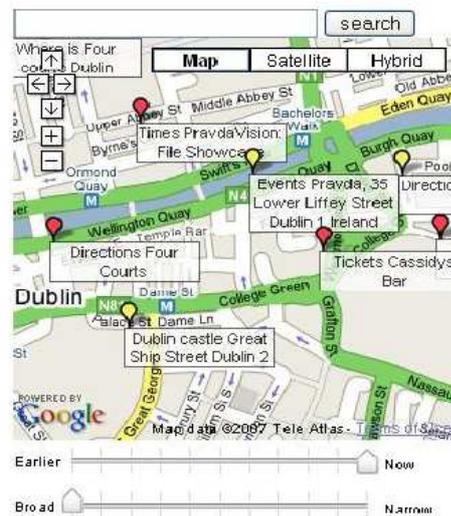


FIG. 2.20 – Un snapshot de l'interface de recherche intégrant les dimensions du contexte selon [52].

4.2.5.2 Combinaison des scores

Dans les travaux proposant la combinaison des scores, la pertinence finale des documents est calculée en combinant des scores individuels des documents selon les diverses dimensions contextuelles considérées.

Dans [244] les messages à délivrer à l'utilisateur sont filtrés sur la base d'une combinaison logique de scores contextuels en utilisant les opérateurs booléens ET, OU et NON.

Dans la système Magitti [27], le score final des informations à recommander à l'utilisateur est calculé sur la base d'un ensemble de règles de combinaison spécifiées, de plusieurs modèles contextuels : des modèles de filtrage collaboratif, de distance, de préférence implicites et/ou explicites de l'utilisateur. etc.

Dans [28], le score final des documents est calculé par combinaison linéaire d'un score contextuel et d'un score retourné par le moteur de recherche. Le score contextuel d'un document est lui même une combinaison de scores contextuels individuels. Le modèle de combinaison des scores contextuels individuels est appris sur un modèle de régression linéaire qui représente un schéma de causalité décrivant l'utilité d'un document (un événement) sur la base du temps (par rapport à la date d'occurrence de l'événement), de sa localisation (par rapport à la localisation de l'utilisateur) et des centres d'intérêts de l'utilisateur.

Dans le système Hapori [139] le problème d'ordonnancement multidimensionnel des documents (ici des POIs) est abordé comme un problème de classification. La requête de l'utilisateur présentée par un choix d'une catégorie de POI, est étendue par le contexte de localisation, temps, météo, popularité et groupe communautaire de l'utilisateur. Elle est ensuite considérée comme un nouveau point à classer dans l'espace des POIs présenté dans les mêmes dimensions contextuelles, appelé *modèle de préférences de POI*. Le modèle de classification KNN est utilisé de manière itérative pour ordonner tous les POIs dans la catégorie sélectionnée par l'utilisateur. La classification se base sur une mesure de distance apprise par l'algorithme de classification LMNN.

Le système Mymose [44] propose d'utiliser le modèle MCI - *multiplicative competitive interaction* [167] comme modèle de combinaison des scores contextuels individuels en une seule expression de l'utilité d'un document à l'utilisateur mobile. L'expression générale de l'utilité d'un document selon le modèle MCI est :

$$U_{ij} = \prod_{k=1}^q \left(X_{kij}^{\beta_k} \right) \quad (2.14)$$

Où X_{kij} est le score du k^{th} attribut décrivant une dimension du contexte du document j pour l'utilisateur i , β est un paramètre qui mesure la sensibilité ou la pertinence de l'attribut k . Les attributs contextuels considérés sont principalement la localisation et le contexte du dispositif mobile qui sont combinés avec le score textuel d'un document. La valeur du paramètre de combinaison β est calculée expérimentalement. Les documents sont finalement ordonnés par

la formule suivante :

$$\pi_{ij} = \frac{U_{ij}}{\sum_{t=1}^n U_{it}} \quad (2.15)$$

Le score π_{ij} est directement proportionnel à l'utilité relative d'un document i par rapport à la somme de l'utilité des n documents dans la collection.

4.3 Synthèse des modèles d'accès contextuel à l'information en RI mobile

La revue des travaux de l'état de l'art sur les modèles d'accès contextuel à l'information abordés précédemment, nous a permis de distinguer plusieurs modèles d'accès selon : le type de la tâche de recherche, la phase du processus d'accès choisie pour intégrer le contexte, le contexte exploité et la technique d'adaptation adoptée.

Différentes tâches d'accès contextuel à l'information sont proposées dans l'environnement mobile, en plus des systèmes de recherche *ad hoc*, des systèmes de recherche proactive sont proposés, ils varient des systèmes de recommandation, aux services basés localisation en passant par les systèmes de filtrage. Certains travaux proposent d'offrir des interfaces de recherche combinant différentes tâches de recherche.

Les modèles d'accès par type de contexte en environnement mobile varient des modèles exploitant un seul contexte à ceux combinant plusieurs contextes proposant des modèles d'adaptation multi-dimensionnels. L'adaptation multi-dimensionnelle présente une problématique spécifique liée au choix ou la proposition d'une technique de combinaison adaptée. Nous avons remarqué une diversité des contextes exploités comparativement aux modèles d'accès contextuels développés pour l'accès classique avec les PC de bureau, qui sont centrés particulièrement sur le profil utilisateur.

Tous ces éléments sont à la base de la différence de performance des SRI contextuel dans cet environnement. Compte tenu de ces éléments, l'évaluation de l'efficacité de la recherche contextuel en RI mobile est très importante. Nous présentons dans la section suivante l'émergence des cadres d'évaluation orientés-contexte dédiés à l'évaluation des SRI contextuel dans cet environnement.

5 Évaluation de l'efficacité de la RI contextuelle dans un environnement mobile

L'évaluation de l'efficacité du processus de la RI a toujours été au cœur de tout processus de développement d'un SRI. L'émergence de la RI contextuelle et mobile a remis en cause la viabilité du modèle d'évaluation de type laboratoire pour évaluer les systèmes d'accès contextuel à l'information. En effet, l'absence du contexte de l'utilisateur et son environnement de recherche dans le protocole de l'évaluation des SRI représente les limites de ce modèle d'évaluation [127, 143].

Ces limites ont encouragé la communauté en RI à promouvoir des méthodologies d'évaluation intégrant l'utilisateur et son environnement de recherche dans le processus d'évaluation. Nous présentons dans cette section l'émergence de l'évaluation orientée contexte des SRI mobiles. Nous présentons en premier lieu les limites de l'évaluation orientée-système, puis nous ferons un tour d'horizon des différentes approches et mesures d'évaluation permettant d'intégrer le contexte de recherche dans le processus d'évaluation, et nous terminerons par une discussion autour des approches d'évaluation proposées.

5.1 Limites de l'évaluation orientée-laboratoire

Les approches classiques de l'évaluation des SRI se basent sur les requêtes de test comme la seule ressource qui traduit le besoin en information de l'utilisateur. Avec l'émergence de la RI contextuelle, ce type d'évaluation présente des limitations. L'étude menée par Tamine et al. [235], sur les protocoles d'évaluation des SRI en présence du contexte, détaille les limitations de ces protocoles. On peut résumer les principales limites des protocoles d'évaluation orientée-laboratoire dans les points suivants :

1. **Les collections de test sont peu adéquates pour l'évaluation de la recherche d'information en contexte**

- *Du point de vue des requêtes* : le protocole d'évaluation étant en mode *batch*, les requêtes sont supposées ainsi représenter à elles seules l'utilisateur. Les utilisateurs directs ayant émis ces requêtes, leurs centres d'intérêt, leurs situations, leurs dispositifs d'accès à l'information et interactions avec le SRI ne font pas partie intégrante du cadre d'évaluation.
- *Du point de vue de la pertinence* : la pertinence considérée des documents est thématique, indépendante du contexte de recherche et centres d'intérêt des utilisateurs. Or, il a été bien montré ([160, 62]) que la no-

tion de pertinence est plus complexe, couvrant des niveaux divers liés à la situation de recherche en cours : pertinence cognitive, pertinence affective, pertinence situationnelle, etc. Dans [62] les auteurs présentent une notion de "w-relevance" et souligne la nécessité de déplacer la notion de pertinence dans le "monde réel/physique" de sorte qu'elle soit le plus proche possible de ce que les utilisateurs veulent et attendent.

2. Les mesures d'évaluation ne sont pas exhaustives

- *Portée* : les mesures de précision ne permettent pas d'évaluer des SRI opérationnels. En effet d'autres critères tels que le degré de couverture du besoin en information et la valeur ajoutée des résultats (cas de documents déjà vus par exemple) ne sont pas corrélés aux mesures de précision [153]. De plus, il n'a pas été prouvé que les différences statistiques entre le rappel et la précision de différents systèmes sont significatives dans des situations réelles de recherche d'information.
- *Utilisation de la pertinence binaire* : les mesures de rappel et de précision se basent sur une pertinence binaire dans l'évaluation de performance des SRI. Dans [32], les auteurs estiment qu'une pertinence graduelle est plus adaptée pour l'évaluation de la performance des SRI en fonction de facteurs liés à l'environnement de recherche et à la disposition d'esprit de la personne ayant réellement besoin de l'information.

La limite des méthodologies d'évaluation de type laboratoire a motivé les réflexions autour de méthodologies d'évaluation adaptées à une recherche d'information contextuelle. C'est particulièrement l'objet de *workshops* très récents tels que CIRSE (2009 et 2010), "future of IR evaluation" dans le cadre de SIGIR'2010, et d'autres *workshops* dans ECIR'2011. Dans la suite de cette section, nous donnons un aperçu de ce qui se fait dans la littérature.

5.2 Approches d'évaluation orientées-contexte

A ce jour, il n'existe pas de cadre standard pour l'évaluation de l'efficacité des modèles et techniques d'accès contextuel à l'information. Les travaux de validation en RI contextuelle et mobile reposent sur des méthodologies et cadres spécifiques en terme de choix de : collections, jugements de pertinence, contextes, requêtes, etc., liés à la technologie mise en œuvre. Nous pouvons distinguer trois principales approches : approches basées sur la simulation du contexte, approches basées sur les fichiers *logs* de recherche et approches basées sur des études utilisateurs.

5.2.1 Approches basées simulation du contexte

Les cadres de simulation contextuelle permettent aux SRI d'être évalués, selon une vue formative, sans égard pour les contraintes liées aux technologies utilisées pour la détection des contextes ou les fonctionnalités de bas niveau du système. Ce type d'évaluation simule les utilisateurs, leurs contextes et leurs interactions avec le système à travers des scénarios de recherche (hypothèses) [164, 202].

Les démarches d'évaluation par simulation de contexte consistent à simuler des scénarios de recherche par une ou plusieurs requêtes liées à des centres d'intérêt de l'utilisateur [266, 219, 155], à des contextes physiques tels que la localisation, le temps et l'activité [90, 161, 126], ou des interactions avec le SRI à travers des interfaces dédiées comme celles proposées dans [125].

Le but principal de ce type d'évaluation, consiste à évaluer l'impact de l'addition du contexte de recherche dans le processus de recherche d'information, sur l'efficacité du SRI dans ce type d'approches. Cette évaluation est généralement basée sur la comparaison des deux listes de résultats obtenues selon les deux scénarios de recherche classique et personnalisée par rapport à des jugements de pertinence hypothétiques [90, 126] ou préalablement donnés [161, 155].

Généralement, des mesures dérivées du rappel-précision sont utilisées pour mesurer l'évaluation de l'efficacité du modèle. Cette approche présente deux avantages majeurs : (1) n'est pas couteuse en temps puisqu'elle n'implique pas des utilisateurs réels (2) permet d'effectuer une évaluation comparative.

5.2.2 Approches basées sur les fichiers *logs* de recherche

Les cadres d'évaluation basés sur l'exploitation des données enregistrées dans les *logs* de recherche sont initialement proposés comme alternative pour la pratique de recueil des jugements de pertinence complets [116]. Ils sont par ailleurs exploités dans [73, 75, 225, 186, 227] pour évaluer et analyser plusieurs techniques et stratégies de personnalisation des résultats de la recherche. Dans ces travaux les interactions de l'utilisateur (*click-through data*), avec les documents retournés par le moteur de recherche sont exploitées par différentes manières pour construire le profil utilisateur. Dans le cadre de la RI mobile, on trouve le travail [139], où un *log* de recherche locale plus riche, étant annoté de plus par des données de temps et de localisation, est exploité pour l'apprentissage d'un modèle d'ordonnement multidimensionnel de points d'intérêts (POIs) pour une recherche locale personnalisée. Les interactions des utilisateurs avec les résultats de la recherche sont de plus utilisées comme source d'évidence

pour les jugements de pertinence.

Par ailleurs des *logs* de connexion à des antennes téléphoniques cellulaires (ou station de base) [29], ou des fichiers de trace GPS [46, 270] sont aussi exploités pour l'évaluation des approches d'adaptation basées localisation.

Pour l'évaluation des performances des systèmes de RI contextuels, des mesures de performance se basant sur le rang des documents cliqués dans les *logs* sont utilisées. On cite la mesure *average rank* (AvgRank) [225, 186, 73, 139], la mesure *rank scoring* [227, 73] et la mesure *R-Score* [75].

L'avantage majeure de l'évaluation basée sur les *logs* de recherche est la richesse de ces derniers en terme de nombre d'utilisateurs, de la variété de leurs préférences, des besoins en information et des contextes réels de recherche. Toutefois, les *logs* de recherche sont généralement mis en place par des moteurs de recherche propriétaires (individuels ou commerciaux), ce qui rend leur accès à la communauté des chercheurs universitaires difficile. En outre, les journaux de recherche ne contiennent pas toutes les informations contextuelles, et doivent être complétés par d'autres sources d'information : *logs* de capteurs externes au SRI (température, localisation, mouvement, personne à côté, etc) ou de services web tels que : les services de météo, géographique, etc.

5.2.3 Approches basées sur des études utilisateurs

Les approches d'évaluation par des études utilisateurs sont menées avec de vrais utilisateurs, appelés participants, pour tester les performances du système à travers des interactions réelles des utilisateurs avec le système. L'objectif de l'évaluation est de mesurer l'efficacité de la recherche en tenant compte de la nature dynamique du besoin en information et du jugement de pertinence ainsi que de l'interaction utilisateur-SRI dans des situations de recherche réelles. Pour évaluer les performances de la recherche contextuelle, chaque participant est tenu de délivrer un certain nombre de requêtes de test et de déterminer si chaque résultat est pertinent dans son contexte. Il existe deux types d'études utilisateurs adoptées dans le domaine : tâches contrôlées et tâches incontrôlées.

1. **Tâches contrôlées** : ce type de démarche d'évaluation est basé sur le cadre d'évaluation spécifié dans [31] dit "*simulated work task situations*". Cela consiste à placer l'utilisateur dans des situations (prédéfinies) les plus proches possible de l'utilisation réelle du système, et de lui recommander de suivre des scénarios d'utilisation construits afin de vérifier des hypothèses identifiées précédemment. Ces scénarios correspondent généralement à des tâches typiques de l'utilisateur. Dans [86], les auteurs ont adopté ce type d'évaluation, et proposent les principales lignes directrices suivantes pour bien mener son évaluation : (1) étudier l'utilisateur

dans son contexte en assurant l'adéquation entre le cadre de l'étude et l'utilisation du système, (2) impliquer des participants qui sont soit des utilisateurs expérimentés ou de futurs utilisateurs du système, (3) choisir des situations qui intègrent les différents aspects de l'environnement, et (4) définir des tâches pertinentes qui incitent l'utilisateur à rechercher des informations conformes aux situations qui ont été identifiées.

Dans [148], 50 utilisateurs se sont vu attribuer la tâche de recherche de documents qui sont pertinents à leurs centres d'intérêts. Chacun des 50 utilisateurs participants à l'étude, s'est vu assigner 5 requêtes prédéfinies sélectionnées de façon aléatoire dans un ensemble prédéfini de 15 catégories thématiques.

Les études utilisateurs contrôlées sont toutefois critiquées parce qu'elles reposent encore sur des besoins informationnels artificiels qui peuvent être faussés par des effets d'inter-sujets et d'ordre des tâches.

2. **Tâches incontrôlées** : aussi dites évaluation par des essais sur le terrain ou "*field trial*". Dans ce type d'évaluation, les participants sont invités à utiliser le système à leur gré quand et comme ils le souhaitent, plutôt que de leur demander d'effectuer une liste de tâches prédéfinies. Plusieurs travaux implémentent leurs systèmes [27, 165, 87] et les évaluent par des essais sur le terrain. Des aspects de l'environnement de l'utilisateur (localisation, temps, activité, personnes à côté) sont ainsi évalués dans des contextes réels associés à des besoins informationnels spontanés des utilisateurs mobiles. Les profils utilisateurs sont manuellement spécifiés par les participants eux mêmes ou automatiquement appris le long des activités de recherche.

Pour l'évaluation des performances du système, en plus des mesures classiques de rappel et de précision et leurs dérivées, d'autres mesures orientées-rang telles que la DCG (*Discounted Cumulative Gain*) [270] et la précision au rang n [86, 174] sont également utilisées et se basent sur la position des documents pertinents dans la liste des résultats de la recherche. Dans [174] une mesure de corrélation de listes ordonnées *Spearman's rank correlation coefficient* [147] est utilisée pour comparer la liste des top 20 résultats retournée par le système et la liste jugée et ordonnée par l'utilisateur. D'autres mesures de pertinence relatives à la vitesse de la recherche et la durée d'une tâche de recherche sont aussi utilisées dans [96, 125].

L'avantage des évaluations basées études utilisateurs est que les jugements de pertinence peuvent être explicitement spécifiés par les utilisateurs selon leurs perception de la pertinence de l'information retournée par le SRI dans des situations de recherche réelles. Cependant, cette approche pose un problème lié à la reproductibilité des résultats et par conséquent à l'inconsistance de l'évaluation comparative. De plus elle est très coûteuse en terme de temps.

5.3 Discussion

L'évaluation par simulation du contexte et l'évaluation par études utilisateurs sont les deux extrêmes du spectre des méthodes d'évaluation de la RI contextuelle et mobile [235]. Les deux approches présentent des mérites qui peuvent être exploités à différents stades du processus d'évaluation d'un SRI. L'approche d'évaluation basée simulation de contexte est bénéfique car elle est moins longue et moins coûteuse que les évaluations avec de vrais utilisateurs dans de vrais contextes. Cependant, cette approche a encore des zones d'incertitude, par exemple le choix des hypothèses sous-jacentes aux scénarios majeurs est ouvert à la critique pour son manque de réalisme. L'avantage des études utilisateurs est qu'elles sont conduites avec de vrais utilisateurs et donc les jugements de pertinence sont explicitement spécifiés par eux. Les limitations majeures de cette approche sont liées au coût en temps, à la reproductibilité des résultats et par conséquent à l'inconsistance de l'évaluation comparative, et au fait qu'elle nécessitent que le système soit complètement développé.

Entre ces deux approches principales, il existe quantité de méthodologies d'évaluation développées dans la littérature de la RI contextuelle et mobile, proposant des collections de test, des protocoles et des mesures d'évaluation. Cependant, tous ces cadres d'évaluation restent propriétaires ou très spécifiques aux applications développées. On déplore à ce jour, l'absence de protocoles standards permettant d'être réutilisés à des contextes d'utilisation différents. Ceci reste un réel défi et une piste de recherche que nous inscrivons dans nos perspectives.

Confrontés à ce défi dans le cadre de nos activités de validation, nous avons alors proposé des méthodologies d'évaluation pour vérifier l'efficacité des modèles d'accès à l'information que nous avons proposés. A cet effet, nous avons adopté en premier lieu une approche basée sur des contextes simulés. Notre choix est justifié par la faisabilité de la validation à moindre coût et dans un temps court. Dans un second lieu, nous avons proposé une approche d'évaluation basée "étude journalière". Notre cadre d'évaluation hérite de l'avantage majeur des études utilisateurs permettant une évaluation avec des utilisateurs et des contextes réels et soulage son exigence que le système soit pleinement développé, en permettant l'évaluation de nos travaux en début de développement. Dans les chapitres suivants, nous spécifions les cadres d'évaluation que nous avons développés et effectivement exploités pour valider nos travaux.

6 Conclusion

Nous avons présenté au cours de ce chapitre l'application de la RI contextuelle dans un environnement mobile en abordant tout d'abord l'émergence de la RI mobile et ses limitations, mais aussi la liaison étroite de la RI mobile aux différents types de contextes. Nous avons présenté une panoplie de travaux portant sur l'accès contextuel à l'information guidé par divers types de contextes allant de l'appareil mobile, au contexte social en passant par le contexte spatio-temporel et le contexte personnel de l'utilisateur. Pour ces divers types de travaux, nous avons développé les techniques d'acquisition, de représentation et d'exploitation du contexte dans le processus d'accès à l'information.

Nous avons constaté qu'il n'existe pas de modèle unique ou universel du contexte, chacun le redéfinit ou l'exploite selon le type de l'application (RI ad hoc, RI web, guide touristique, recherche locale, etc) ou le domaine d'application considéré (système de recherche dédiée informations médicales, des News, des photos, 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.

Comme l'évaluation de l'efficacité de la recherche est une étape nécessaire et importante de tout processus de développement en RI, nous avons présenté les démarches d'évaluation proposées dans ce cadre. Nous avons mentionné l'inadéquation des cadres d'évaluation classiques des SRI basés sur le paradigme d'évaluation de *Cranfiled* et nous avons montré l'évolution vers des approches d'évaluation orientées-contextes. Ces dernières varient des approches basées simulation du contexte aux approches basées études utilisateurs en passant par les approches basées sur les *logs* de recherche.

A ce niveau, nous avons également constaté l'absence d'un cadre d'évaluation standard, qui permettrait une évaluation comparative des différents travaux. En effet, les méthodologies d'évaluation proposées sont dédiées à des technologies bien ciblées et sont dès lors non réutilisables. Ce qui nous motive alors à mener des réflexions pour la proposition de cadres d'évaluation adéquats pour faire valider nos travaux.

Deuxième partie

Contribution à la définition d'un
modèle d'accès contextuel à
l'information dans un
environnement mobile

Introduction

L'état de l'art présenté dans la première partie de cette thèse a donné un aperçu du domaine de la RI contextuelle en environnement mobile, plus particulièrement des méthodes et techniques proposées pour la modélisation et l'exploitation des diverses dimensions du contexte dans le processus de recherche. Les défis majeurs de chacune de ces pistes de recherche sont résumés comme suit :

1. La plupart des approches dans le domaine de la RI contextuelle mobile s'orientent vers la représentation sémantique du contexte de l'utilisateur. Le défi majeur à ce niveau consiste à proposer le modèle du contexte le plus adéquat à la tâche de recherche permettant de répondre à des besoins en information très variables et très dynamiques des utilisateurs.
2. La combinaison de plusieurs dimensions contextuelles. Nécessité de réfléchir et de définir le modèle de combinaison des pertinences individuelles et de personnaliser cette combinaison en tenant compte des préférences des utilisateurs sur chacune des dimensions contextuelles considérées.
3. L'évaluation de l'accès contextuel à l'information. Il n'existe pas un cadre d'évaluation standard pour l'évaluation de l'accès contextuel à l'information. La majorité des travaux dans le domaine s'orientent vers l'évaluation par simulation des contextes, par utilisation des *logs* de recherche ou des études utilisateurs.
 - Le premier type d'évaluation permet de mener des expérimentations reproductibles dans des conditions d'évaluation stables et évaluer l'efficacité de recherche contextuelle dans des situations de recherche hypothétiques contrôlables ou idéales.
 - Le deuxième type d'évaluation permet de mener des expérimentations avec des données très riches en terme de nombre d'utilisateurs, de la variété de leurs besoins en information et des contextes réels de recherche, cependant ils restent des cadres propriétaires non ouverts à la communauté des chercheurs universitaires.
 - Le troisième type d'évaluation permet d'évaluer l'efficacité de la recherche contextuelle dans des contextes de recherche réels mais qui ne peuvent pas être reproductibles dans les mêmes conditions d'évaluation.

Dans cette partie nous présentons notre contribution dans le domaine pour répondre à ces défis. Elle peut être déclinée sous 4 propositions principales :

1. Un modèle d'accès contextuel guidé par un profil sémantique et situationnel de l'utilisateur. En effet, les centres d'intérêts des utilisateurs sont un

critère contextuel très important à considérer pour adapter les résultats de recherche au caractère personnel de la recherche mobile.

2. Un modèle d'accès contextuel guidé par la sensibilité de la requête à la localisation de l'utilisateur. En effet, la mobilité de l'utilisateur suppose des besoins en information liés à l'environnement physique de l'utilisateur.
3. Un modèle d'agrégation prioritaire des dimensions contextuelles. En effet, pour pouvoir combiner les pertinences individuelles de chaque facteur contextuel dans un modèle d'accès multidimensionnel et personnalisé tenant compte des préférences des utilisateurs.
4. Confrontés au problème de l'absence d'un cadre d'évaluation standard des techniques de recherche contextuelles dans l'environnement mobile, nous avons proposé et construit des cadres d'évaluation adéquats pour chacune de nos propositions précédentes. Nous avons exploité principalement des cadres de simulation contextuelle mais nous avons également défini et construit un cadre d'évaluation à partir d'utilisateurs et de données contextuelles réelles.

Cette partie est organisée comme suit. Le chapitre 3 présente notre approche de personnalisation basée sur la construction, évolution et exploitation de profils situationnels des utilisateurs dans un processus d'accès à l'information personnalisé, ainsi que les cadres d'évaluation que nous avons proposé pour son évaluation.

Le chapitre 4 présente notre approche de prédiction de la sensibilité de la requête à la localisation de l'utilisateur et son intégration dans un processus d'adaptation de la recherche mobile au type de la requête, ainsi que son évaluation.

Le chapitre 5 présente notre approche d'agrégation prioritaire de pertinences contextuelles pour une adaptation multidimensionnelle et personnalisée de la recherche mobile ainsi que son évaluation.

Chapitre 3

Modélisation d'un profil flexible pour un utilisateur mobile : proposition et évaluation

1 Introduction

La modélisation des profils utilisateurs se trouve au coeur des systèmes de personnalisation ou d'adaptation de contenu (recherche d'information, système de recommandation, etc.). L'objectif de la RI personnalisée est alors d'intégrer le profil utilisateur dans le processus de RI afin d'adapter les résultats de la recherche aux besoins informationnels spécifiques à chaque utilisateur en tenant compte de ses centres d'intérêts.

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 utilisateur (ou accès personnalisé) dans un environnement mobile. Le défi majeur dans ce cadre est de tenir compte de la dynamique des centres d'intérêt de l'utilisateur liée au changement de son environnement du à sa mobilité. Cela consiste en pratique à modéliser l'environnement de l'utilisateur et ses centres d'intérêt ainsi que leur liaison, et ensuite à personnaliser les résultats de la recherche en identifiant le profil adéquat pour la personnalisation.

Notre contribution dans le domaine de l'accès personnalisé à l'information pour la RI mobile, porte sur la conception d'un modèle d'accès personnalisé à l'information intégrant la construction de profils situationnels des utilisateurs mobiles [10, 2, 7, 1, 6]. Le modèle d'accès personnalisé se base sur la conception d'un profil sémantique de l'utilisateur qui varie selon le changement de sa situation spatio-temporelle. Le profil utilisateur est représenté par un en-

semble de centres d'intérêts par situations de recherche. Chaque situation de recherche est modélisée sur les deux dimensions sémantiques de l'environnement de l'utilisateur notamment le temps et la localisation. Les centres d'intérêts de l'utilisateur sont représentés par des graphes de concepts issus de l'ontologie thématique ODP¹. Une approche de raisonnement à partir de cas (RàpC) est utilisée pour adapter le processus de personnalisation à la flexibilité du profil utilisateur. Chaque cas dans notre base de cas représente une instance du profil sémantique de l'utilisateur.

Pour évaluer l'efficacité de notre approche, nous avons proposé deux protocoles d'évaluation adaptés à l'intégration du profil situationnel de l'utilisateur dans le processus de RI.

Le premier protocole d'évaluation [9] est basé sur une simulation contextuelle de besoins informationnels et de situations de recherche.

Le second protocole [3] est basé sur de vrais besoins informationnels et situations de recherche extraites à partir d'une étude journalière des besoins informationnels d'un ensemble d'utilisateurs réels en situation de mobilité.

Ce chapitre est organisé comme suit. La section 2 présente les problématiques et motivations qui ont suscité notre proposition d'un modèle de représentation de la situation spatio-temporelle de recherche ainsi que le mode de dérivation des centres d'intérêts de l'utilisateur. La section 3 présente une vue générale de notre approche d'accès personnalisé à l'information ainsi que la terminologie et les notations utilisées. La section 5 présente le modèle de représentation des situations et des centres d'intérêts de l'utilisateur. La section 6 se focalise sur le processus de sélection puis d'exploitation du profil de l'utilisateur dans un processus de ré-ordonnancement des résultats de la recherche. La section 7 présente nos travaux d'évaluation. La dernière section 6 conclut le chapitre.

2 Problématiques et motivations

Dans un environnement mobile en changement continu, les besoins informationnels des utilisateurs peuvent changer à tout moment en raison du changement de sa situation environnante (lieu, heure, personnes à côté, etc). Comme simple scénario, imaginons un utilisateur dans un musée, qui soumet la requête "*Water lilies*". L'utilisateur peut à la fois être intéressé par "l'art et les plantes". Sachant, que l'utilisateur est actuellement dans un "musée", nous pouvons améliorer les performances des résultats de la recherche en tenant compte de son centre d'intérêt actuel "art" plutôt que "plantes" étant donné qu'il est au musée et non dans un "parc ou jardin".

¹Open Directory Project(ODP) : <http://www.dmoz.org>

C'est pourquoi les travaux présentés dans le chapitre 2 [266, 29, 172, 52, 27] se sont orientés vers l'exploitation d'autres ressources en plus des interactions de recherche pour construire un modèle utilisateur plus dynamique qui évolue avec le contexte dans lequel se trouve l'utilisateur mobile, affinant ainsi la notion de centre d'intérêt de l'utilisateur vers des besoins situationnels. Les défis majeurs soulevés dans ce domaine sont principalement : 1) Comment modéliser la situation de l'utilisateur et les centres d'intérêts de l'utilisateur ? 2) Comment représenter le lien entre une situation de recherche et ses centres d'intérêts rattachés ? Et 3) Comment personnaliser les résultats de la recherche en tenant compte de la diversité et multiplicité des situations de recherche que peut avoir un utilisateur mobile ?

Notre approche s'inscrit précisément dans ce domaine ; nous proposons la construction d'un profil sensible au contexte qui peut être utilisé pour adapter les résultats de recherche selon les besoins informationnels de l'utilisateur dans une situation/ un contexte donné.

D'une manière plus spécifique, nos objectifs dans le cadre de notre approche de personnalisation s'articulent autour de la définition d'un profil utilisateur ayant une représentation sémantique tant sur le contexte situationnel que sur le contenu thématique, la construction et l'évolution de ce profil tout au long des activités de recherche en vue de l'exploiter pour répondre à des besoins en information situationnels de l'utilisateur mobile. Nos motivations concernant les différents éléments de notre contribution s'articulent autour des points suivants :

- Représenter l'utilisateur par un paquet de situations sémantiques auxquelles sont rattachés des ensembles de concepts sémantiquement reliés traduisant ainsi un centre d'intérêt spécifique à chaque situation. L'objectif est de dégager une représentation unique du besoin en information relié à une situation de recherche spécifique.
- Construire et faire évoluer le profil au sein d'une même situation de recherche selon des mécanismes de collecte de l'information pertinente issue de l'évaluation des requêtes traitant du même besoin en information et le déploiement de ces informations au niveau spécifique du besoin de l'utilisateur dans une situation de recherche donnée.
- Exploiter le modèle de l'utilisateur 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 relatif à la situation la plus similaire identifiée. La similarité des situations de recherche est à la base de l'identification du centre d'intérêt le plus adéquat à une tâche de recherche initiée par une nouvelle requête.

Plus précisément, nos contributions portent principalement sur la proposition d'un modèle sémantique pour la représentation du contexte spatio-temporel de l'activité de recherche de l'utilisateur mobile, d'un modèle de représentation sémantique des centres d'intérêts de l'utilisateur, et l'utilisation de la technique de raisonnement à partir de cas pour l'identification d'un profil utilisateur adéquat à chaque nouvelle situation de recherche. Nous résumons notre orientation générale et la spécificité de notre contribution comparativement aux approches développées dans le domaine selon les aspects suivants :

- Une modélisation de la situation de l'utilisateur utilisant à la fois des concepts sémantiques extraits d'ontologies spatiale et temporelle. La motivation derrière notre choix est triple : d'abord, notre intuition est que les besoins informationnels des utilisateurs mobiles sont liés à leurs activités courantes. Or toute activité humaine est généralement dépendante ou évolue selon la situation temporelle et spatiale de l'utilisateur, nous supposons que ces derniers sont suffisants pour notre tâche de décrire la relation entre les centres d'intérêts de l'utilisateur mobile et son environnement. La seconde raison est que les centres d'intérêts des utilisateurs ne sont pas liés à des aspects physiques de l'environnement (comme les coordonnées GPS), mais plutôt à la sémantique que l'on peut rattacher à un lieu ou à un temps donné. La troisième raison est typiquement technique, elle concerne le fait que ces éléments contextuels peuvent être facilement acquis de façon automatique. Tandis que [172] exploitent uniquement le contexte temporel, [29] exploitent uniquement la localisation, [52] utilise une représentation en valeurs physiques.
- L'exploitation du modèle de l'utilisateur dans une approche de raisonnement à partir de cas (RàpC) afin de sélectionner de manière flexible le profil adéquat à la personnalisation. Le principal avantage de cette approche est la flexibilité qu'elle offre pour modéliser les situations et les centres d'intérêt à partir d'ontologies externes et l'adaptation du processus de personnalisation.
- Une acquisition implicite et automatique du modèle de l'utilisateur ; aucun effort n'est demandé à l'utilisateur, tandis que dans [172, 29] l'utilisateur est sollicité dans le processus de construction de son profil.
- Une modélisation générale d'un utilisateur mobile dans le sens où elle ne porte pas sur des situations et des populations spécifiques, tandis que dans [27], l'approche proposée est consacrée à des situations de loisirs et à des populations jeunes.

3 Architecture générale de notre approche d'accès personnalisé à l'information

L'architecture générale de notre approche d'accès personnalisé à l'information est illustrée dans la figure 3.1. Cette architecture est basée sur la modélisation et l'exploitation d'un profil situationnel de l'utilisateur décrivant ses centres d'intérêts à partir du *feedback* de l'utilisateur et de sa situation spatio-temporelle.

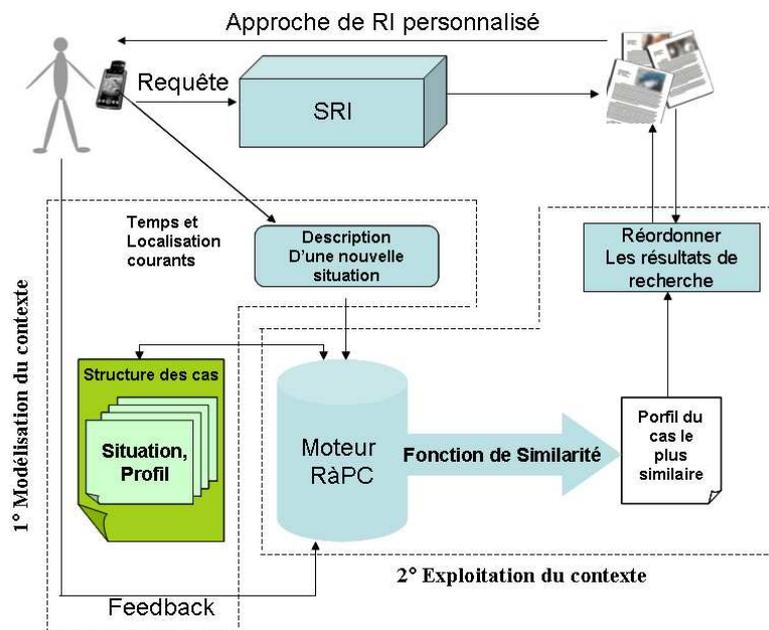


FIG. 3.1 – Architecture générale de notre approche de RI personnalisée

Nous identifions dans cette architecture les composantes principales suivantes : un module comprenant le modèle contextuel de l'utilisateur et un module exploitant ce modèle pour personnaliser la recherche. Nous tentons de présenter notre approche d'accès personnalisé à l'information par la désignation des modèles qui le supportent : modèle du contexte et modèle d'accès contextuel à l'information.

3.1 La modélisation du contexte

Le contexte considéré dans notre approche comprend deux dimensions : une dimension représentant la situation spatio-temporelle de l'utilisateur et une dimension représentant les centres d'intérêts de l'utilisateur liés à la dimension spatio-temporelle. Nous définissons notre modèle pour la représentation de ce

contexte par instanciation de chacun de ses éléments : les sources d'information, la stratégie de collecte des informations, les ressources de modélisation et les modèles de représentation et/ou d'évolution du contexte. Nous décrivons ci-dessous nos choix par rapport à chacun de ces éléments.

1. Les sources d'information

Comme sources d'évidence pour la modélisation de notre profil flexible, nous nous basons sur l'exploitation de deux sources d'information :

- d'un côté les documents cliqués et/ou visités par l'utilisateur comme source d'évidence représentant l'intérêt de l'utilisateur pour ces documents en réponse à sa requête,
- et d'un autre côté les coordonnées géographiques et temporelles encadrant les activités de recherches de l'utilisateur mobile comme source d'évidence représentant le contexte environnemental du besoin en information de l'utilisateur.

2. Les stratégies de collecte des informations

Notre stratégie de collecte des informations contextuelles est complètement implicite, pour éviter tout effort supplémentaire ou surcharge de l'utilisateur mobile qui est généralement engagé dans d'autres activités. Pour les centres d'intérêts de l'utilisateur, la collecte des informations se base sur la construction de *logs* de recherche, et pour son contexte situationnel, elle se base sur des *logs* ou des traces de ses déplacements. Bien sûr, pour des raisons de sécurité des données personnelles de l'utilisateur, cela nécessiterait en réalité l'autorisation préalable de l'utilisateur de l'exploitation de ses données en vue de la personnalisation de ses recherches.

3. Les ressources de modélisation

Comme ressource de modélisation nous avons opté pour l'utilisation de ressources sémantiques externes. Une ontologie thématique générale est utilisée pour la modélisation des centres d'intérêts. De même, une ontologie spatiale pour la localisation et une ontologie temporelle pour la modélisation du temps.

4. Les modèles de représentation et/ou d'évolution

Notre modèle de représentation de l'utilisateur se base sur une représentation en base de cas. Chaque cas est composé d'une partie prémisses et d'une partie valeur. La partie prémisses est modélisée sous forme de vecteur à 4 dimensions temporelles et spatiales. La partie valeur est modélisée par un

graphe conceptuel représentant le centre d'intérêt de l'utilisateur associé à cette situation.

3.2 L'accès contextuel à l'information

L'exploitation du modèle de l'utilisateur pour réaliser l'accès contextuel à l'information passe par une phase de sélection du profil adéquat en se basant sur une mesure de similarité, puis vient la phase de personnalisation proprement dite.

Notre stratégie adoptée pour la personnalisation de la recherche est basée sur une approche de ré-ordonnement des résultats initiaux de la recherche retournés par un SRI standard. Les algorithmes de ré-ordonnement appliquent généralement des fonctions sur le rang ou le score d'appariement requête-document ou alors sur l'ordre des rangs des résultats restitués par le SRI. Notre fonction de ré-ordonnement est basée sur la combinaison des scores d'appariement original et personnalisé du document.

4 Terminologie, définitions et notations

Dans cette section nous donnons les définitions autour de quelques notions constituant les briques de base de notre approche de personnalisation. Plus spécifiquement, nous nous intéressons à la définition de la notion d'une activité de recherche, d'une situation de recherche, d'un profil de l'utilisateur, d'un modèle de l'utilisateur, ainsi que les relations pouvant lier ces différentes définitions.

1. Activité de recherche

Une activité de recherche peut être représentée par la succession d'un ensemble d'événements permettant d'accomplir une tâche de recherche. Ces événements sont les suivants :

- L'utilisateur u formule puis soumet une requête, que l'on dénote par q à un SRI,
- Le SRI retourne une liste de résultats de recherche dénoté D correspondant à la requête q ,
- L'utilisateur exprime ses préférences sur les documents qui l'intéressent parmi ceux retournés par le SRI. On dénote par D_r la sous-liste des résultats jugés pertinents implicitement par l'utilisateur. Un document est considéré comme pertinent s'il a généré des comportements observables de l'utilisateur tels que sauvegarde et/ou impression du docu-

ment ou s'il satisfait des mesures telles que le taux de *clics*, le temps de lecture, etc.

2. Situation de recherche

Elle constitue le cadre environnemental de l'exécution de l'activité de recherche. Plus exactement, elle encapsule le contexte spatio-temporel d'une activité de recherche. Une situation de recherche est dénotée S^i , elle est représentée par un vecteur de 4-dimensions sémantiques issues des ontologies spatiales et temporelles.

3. Profil situationnel

Le profil situationnel d'un utilisateur mobile se réfère au centre d'intérêt de l'utilisateur appris sur une situation S^i de recherche donnée. Le profil situationnel est dénoté G^i , est représenté comme un graphe de concepts sémantiquement reliés dans une ontologie thématique générale.

4. Modèle utilisateur

Représente l'ensemble des situations de recherche de l'utilisateur et leur profils associés. Le modèle de l'utilisateur est dénoté par $U = \{(S^i, G^i)\}_{i=1,n}$. Où S^i est une situation et G^i son profil correspondant.

5 Modélisation et apprentissage du profil flexible d'un utilisateur mobile

Notre approche générale pour la personnalisation de la recherche repose sur la construction d'un modèle dynamique de l'utilisateur qui évolue en fonction de son environnement. Plus spécifiquement, le modèle de l'utilisateur noté $U = \{(S^i, G^i)\}_{i=1,n}$ comprend un ensemble de situations de recherche S^i à lesquelles sont associés des centres d'intérêt G^i de l'utilisateur.

Dans notre contexte, une situation de recherche ne se caractérise pas par des pures coordonnées géographiques et des points temporels. Notre défi pour la construction de ces profils sensibles au contexte est d'utiliser des données brutes récupérées par des capteurs attachés à l'appareil mobile, pour obtenir des informations contextuelles de haut niveau permettant d'identifier une situation de l'utilisateur. C'est pour cela que nous proposons d'associer les informations du contexte récupéré par les capteurs du mobile (horloge système, coordonnées GPS) à des concepts sémantiques, extraits d'ontologies temporelles et spatiales.

Pour illustrer cette proposition, prenons l'exemple d'un utilisateur qui se trouve aux coordonnées GPS "48.7818034, 2.2183314" & au moment "Mardi Août 3 12 :10 :00 CEST 2009". Avec une représentation sémantique de sa situation, on peut déduire que cet utilisateur se trouve "à la plage, en été, en vacances, à midi".

Pour modéliser les centres d'intérêts de l'utilisateur nous nous sommes reposés sur une approche de représentation des centres d'intérêts de l'utilisateur basée-concept développée dans un travail précédent dans notre équipe [70]. Cette approche permet de construire une représentation sémantique des centres d'intérêts de l'utilisateur à partir de son activité de recherche et en exploitant l'ontologie générale ODP.

Dans la suite de cette section, nous décrivons en un premier temps notre approche pour la modélisation de la situation de l'utilisateur. Puis, nous donnons le principe de l'approche de modélisation des centres d'intérêts.

5.1 Modélisation de la situation de l'utilisateur

Notre modèle de situation de l'utilisateur comprend la représentation des dimensions du contexte temporel et spatial de l'utilisateur. Les informations de temps et de localisation sont complexes et hiérarchiques, et peuvent être représentées à plusieurs échelles de granularité. Dans la suite de cette section, nous présentons notre approche pour la représentation de la localisation, du temps et leur agrégation dans un modèle de situation.

5.1.1 Modélisation de la localisation

La localisation de l'utilisateur mobile est généralement considérée comme un facteur d'une grande importance pour les applications mobiles. Souvent, la localisation de l'utilisateur est considérée comme un concept seulement caractérisé par les coordonnées physiques par rapport à un système de coordonnées géographiques. Cela permet de développer des applications du type *location based services* et *Nearest-X services* qui permettent de recommander à l'utilisateur mobile des informations et des services sur la base de leurs proximité de sa localisation physique.

Cependant, les coordonnées physiques, même étant pertinentes pour caractériser une localisation, ne sont pas les seules caractéristiques que nous pouvons considérer lorsqu'on définit une localisation. En effet, tel que discuté dans [72], il existe différentes façons plausibles et correctes pour caractériser la loca-

lisation de l'utilisateur mobile :

- position absolue dans un système de coordonnées géographiques,
- relative (à côté de, ...),
- nom de place (eg. "Champs de Mars"),
- classe nommée qui représente le type de la place, eg. école (démographique), plage (environnement), monument (historique), domicile (personnel) ou magasin (commercial), ...
- ...

Nous avons retenu la représentation en "classe nommée" ou "type de la place" pour caractériser la dimension spatiale de la situation de l'utilisateur. En effet, ce type de classification reflète généralement une caractérisation fonctionnelle de la place (comme les catégories de place des "pages jaunes").

"Le type de la place" peut être récupéré à partir de classifications manuelles de places qui sont pratiquées par la majorité des systèmes d'information géographiques, comme *the Alexandria Digital Library*² et *GeoNames servers*³, qui donnent à chaque coordonnée géographique (adresse ou nom de place) une classification standard de catégorie de type de localisation.

Pour assurer la connexion entre les coordonnées de la localisation et sa classification sémantique, un modèle conceptuel est nécessaire pour représenter et raisonner sur la localisation. Nous proposons de nous appuyer sur la base de données spatiale téléchargeable à partir du service géographique en ligne *GeoNames servers* pour représenter et raisonner sur les données géographiques. La figure 3.2 présente le modèle simplifié pour la représentation des informations spatiales relatives à la situation de l'utilisateur.

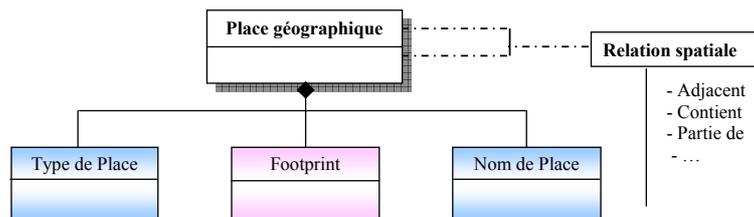


FIG. 3.2 – Modèle simplifié de l'ontologie spatiale.

Dans ce modèle, chaque localisation est caractérisée à la fois par sa représentation concrète en termes de coordonnées géographiques par la classe "footprint", et sa représentation sémantique : nom de la place et son type. Par exemple, l'entité géographique "Louvre", a pour coordonnées "48.861073°, 2.335784°" et pour type "musée".

Dans notre travail, nous nous basons sur cette double représentation pour passer automatiquement de la localisation physique de l'utilisateur à sa représentation sémantique, cette opération est généralement appelée : "reverse geo-coding". Ceci nous permet en particulier de regrouper des lieux de même type, par

²<http://www.alexandria.ucsb.edu/>

³<http://www.geonames.org/>

exemple si l'utilisateur se trouve à une autre coordonnée géographique donnée par "40.779447 °, -73.96311 °", nous pouvons en déduire que l'utilisateur est dans une situation similaire "musée", bien qu'il n'est pas dans le même lieu physique.

De plus les places géographiques sont reliées par diverses relations spatiales comme adjacence, partie de, contenir, etc, qui permettent d'inférer des liens entre les places géographiques. Par exemple, la place géographique "Tour Eiffel" est à côté de place géographique "Champ de Mars", et les deux places sont parties de la place géographique "Paris"

5.1.2 Modélisation du temps

L'information temporelle est une information complexe, elle est continue et peut être représentée à divers niveaux de granularité. Plusieurs travaux dans le domaine de la RI géographique ([144, 185]) se sont intéressés à des questions relatives à l'extraction, la représentation et l'exploitation des aspects temporels pour le calcul d'une pertinence temporelle des documents et les requêtes utilisateurs.

Dans notre travail, nous considérons que le contexte temporel pourrait avoir une influence sur l'activité de recherche de l'utilisateur mobile. Par exemple, imaginons un utilisateur qui émet la requête "restaurant", selon le moment de la journée, on peut décliner des préférences de type de restaurants : "bar café" en matinée, "restaurant rapide" à midi et "restaurant gastronomique" le soir. Nous tentons donc d'exploiter l'information temporelle en vue de décliner les centres d'intérêts des utilisateurs mobiles selon cette dimension du contexte.

Pour définir les aspects temporels caractérisant la situation de recherche de l'utilisateur, nous proposons d'abstraire l'instant temporel dans certaines périodes spécifiques et significatives que nous estimons ayant un effet sur le comportement de l'utilisateur (e.g. matin, weekend, été). Nous avons choisi de représenter le temps sur trois dimensions : journée, semaine et saison. En effet, nous avons estimé que l'activité de l'utilisateur mobile peut être altérée par la variation du temps selon chacune de ces dimensions. Les valeurs prises par chacune de ces dimensions sont représentées par les classes sémantiques suivantes :

- **Moment de la journée** : C'est la classe représentant la dimension de la journée. Si nous étudions la routine quotidienne des utilisateurs, on peut distinguer cinq périodes principales à savoir matin, midi, après midi, soir et nuit, selon l'activité de l'utilisateur pour chaque période.
- **Jour de la semaine** : Cette classe représente la dimension de la semaine. On peut différencier les jours de la semaine entre les jours de

travail (parmi lundi, mardi, mercredi, jeudi et vendredi) et les jours de repos parmi fin de semaine (généralement samedi et dimanche), les jours fériés (tels que la fête des mères, la fête de Noël, ...) et les jours de vacances.

- **Les saisons** : Représente la dimension de la saison. Elle comprend l'une des saisons de l'année parmi automne, hiver, printemps, et été.

Pour permettre une bonne représentation de l'information temporelle et sa manipulation, la tendance est actuellement vers des approches sémantiques avec des ontologies temporelles. L'ontologie OWL-Time [171] est aujourd'hui une référence pour la représentation et le raisonnement sur le temps. Nous proposons de baser notre travail sur cette ontologie et de l'étendre avec nos classes sémantiques du temps. La figure 3.3 illustre notre modèle simplifié du temps où sont représentées les deux couches de temps concret et abstrait. La couche temps abstrait représente notre extension de l'ontologie de base décrite dans [171] par nos classes sémantique du temps.

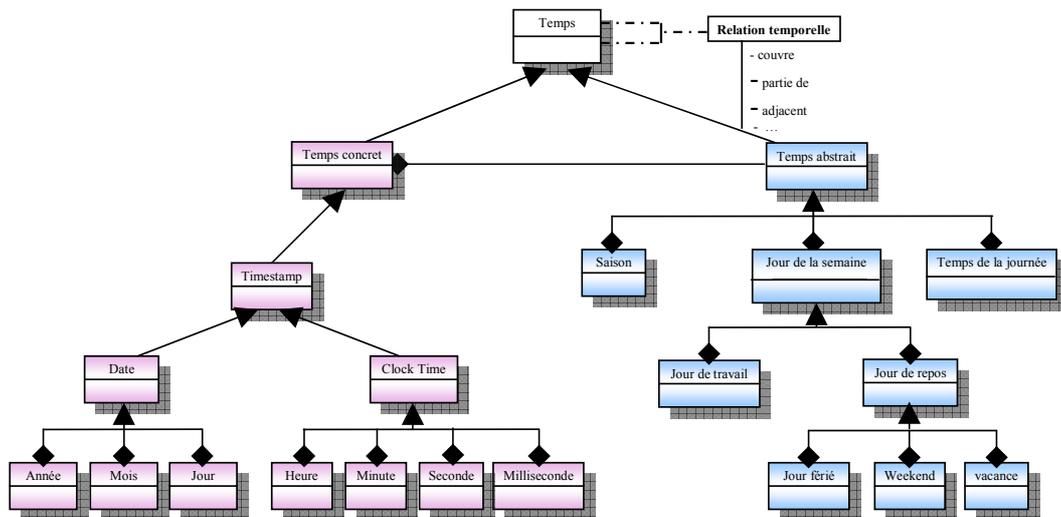


FIG. 3.3 – Modèle simplifié de l'ontologie temporelle.

Le lien entre les deux représentations abstraite et concrète est mis en œuvre par des axiomes et des règles simples de l'ontologie qui permettent d'inférer d'un point temporel des intervalles de temps correspondants à nos périodes sémantiques du temps qui sont plus significatives dans la vie quotidienne des utilisateurs. En plus, des relations sont définies entre les classes temporelles : couvre, est partie de, adjacent, ..., qui permettent de lier les différentes dimensions du temps.

5.1.3 Agrégation dans un modèle de situation

Nous avons choisi d'utiliser le modèle vectoriel pour agréger nos contextes de localisation et de temps en un modèle de la situation de l'utilisateur. Chaque situation de l'utilisateur sera vue comme un point dans l'espace défini par les dimensions de localisation et de temps. Une situation sera donc représentée par l'agrégation des quatre dimensions sémantiques suivantes : type de la localisation, la saison, le jour de la semaine, et le moment de la journée, comme illustré dans le figure 3.4. La combinaison des deux dimensions de temps et

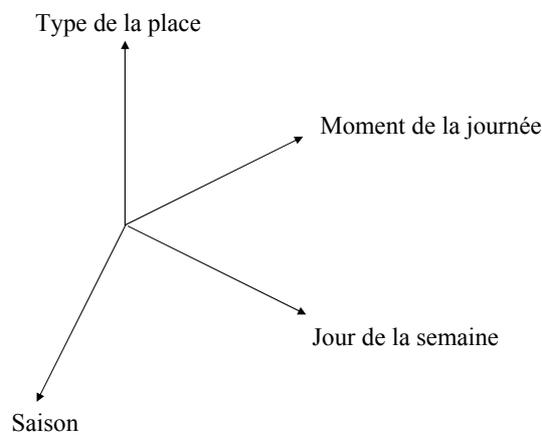


FIG. 3.4 – Dimensions agrégées dans le modèle de situation.

de localisation permet de définir des points dans l'espace spatio-temporel, par exemple : "Nuit, weekend, hiver, à la maison". Plus spécifiquement, une situation S est représentée comme un vecteur dont les caractéristiques sont les valeurs X attribuées à chaque dimension :

$$S = (X_l, X_u, X_v, X_w) \quad (3.1)$$

où X_l et X_u, X_v, X_w est la valeur de la dimension type de la localisation respectivement de la saison, du jour de la semaine et du moment de la journée.

Des fonctions de similarité permettant la manipulation et la comparaison de situations sur les dimensions de temps et de localisation seront définies plus loin dans la section 6.2.

5.2 Modélisation des centres d'intérêt

Nous donnons ci-dessous, un aperçu de l'approche de représentation des centres d'intérêt comme graphe de concepts, développée dans un précédent travail au sein de l'équipe [70]. Cette approche peut être caractérisée par :

- la représentation sémantique des centres d'intérêt de l'utilisateur selon un ensemble de concepts sémantiquement liés via l'ontologie de référence ODP.
- un procédé d'évolution du profil utilisateur à court terme qui repose sur l'augmentation de l'importance des concepts récurrents dans la session de recherche. Une session de recherche est définie par une séquence de requêtes liées à un même besoin en information.
- Une mesure de délimitation conceptuelle des sessions de recherche qui permet de scruter le changement des centres d'intérêt au cours du temps.

Nous avons adapté cette approche pour construire des profils par situations de recherche au lieu de profil par session de recherche. Nous avons proposé notre propre mesure de comparaison des situations de recherche.

L'approche générale de dérivation du profil situationnel est basée sur trois principales étapes :

1. initialisation du contexte de la requête comme un ensemble de mots clés extraits à partir des documents jugés pertinents par l'utilisateur dans une activité de recherche, et projection du profil basé mots clés sur l'ontologie de référence ODP, pour construire un ensemble initial de concepts pondérés.
2. désambiguïsation du profil de la requête par extension de l'ensemble de concepts initial construit dans la phase précédente par une stratégie de propagation de scores pour construire un graphe de concepts représentant le profil utilisateur,
3. évolution du profil au sein de la même situation de recherche en appliquant une stratégie de combinaison des scores des concepts à travers des activités de recherche similaires.

Ces principales phases sont détaillées ci-dessous.

5.2.1 Initialisation du contexte de la requête

Formellement, chaque nouvelle requête q^* est associée à un ensemble de documents D retournés par le système et un ensemble de documents pertinents D_r jugés implicitement par l'utilisateur. La construction du profil de la requête est basée 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 extraction des concepts les plus similaires à la requête.

1. **Extraction du vecteur de pertinence de la requête** : nous dérivons tout d'abord le vecteur de pertinence de la requête, noté K , comme étant un seul vecteur de termes les plus représentatifs dans les documents pertinents associés D_p . Le poids d'un terme dans le vecteur K est calculé selon la formule suivante :

$$K(t) = \frac{1}{|D_p|} \sum_{d \in D_p} w_{td} \quad (3.2)$$

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 .

2. **Projection sur l'ontologie** : pour améliorer cette représentation simpliste, le centre d'intérêt basé mots clé K est comparé aux concepts de l'ontologie ODP, en vue de choisir un ensemble de concepts qui reflètent d'une façon sémantique les centres d'intérêt de l'utilisateur. Chaque concept de l'ontologie ODP est relié à des sous-concepts par la relation "est-un" et il est associé à un ensemble d'URLs de pages Web qui sont classifiées sous ce concept. Chaque concept c_j sera représenté par un seul vecteur de termes pondérés extraits de toutes les pages Web individuelles classifiées sous ce concept (représentées par une description et un titre) et sous tous ses sous-concepts. Le poids d'un terme w_i dans un concept \vec{c}_j est calculé en utilisant un schéma de pondération $tf * idf$. La table 3.1 montre la liste des 10 premiers termes représentant le concept "*Business and Economy/Restaurants and Bars*" de l'ODP, selon cette approche.

Terme	poids dans le concept
restaur	0.922452
guid	0.817915
informat	0.811328
locat	0.761535
menu	0.752425
pub	0.680363
list	0.680363
review	0.446800
dine	0.417205
cuisin	0.385304

TAB. 3.1 – La liste des "10" premiers termes et leurs poids représentant le concept "*Business and Economy/Restaurants and Bars*" de l'ontologie ODP.

Pour un concept c_j de l'ODP, représenté par le vecteur de termes \vec{c}_j , son score de similarité $score(c_j)$ avec \vec{K} est calculé selon la mesure de similarité vectorielle *cosinus* comme suit :

$$score(c_j) = \cos(\vec{c}_j, \vec{K}) \quad (3.3)$$

Nous obtenons ainsi un ensemble $\theta^* = \{(c_1, score(c_1)), \dots, (c_i, score(c_i)), \dots\}$ de concepts pondérés associés à la requête q^* .

5.2.2 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. Le principe de cette étape est d'extraire un graphe de concepts sémantiquement liés à partir de l'ensemble des concepts θ^* associés à la requête tout en respectant la topologie du graphe de l'ontologie de l'ODP. A cet effet, une distinction est faite entre 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^* . 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, la pondération des arcs adoptée est celle utilisée dans [154]. 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 $e_{jk} \in R$.

L'algorithme 1 décrit ce processus de désambiguïisation basé 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 θ^* . Les principales étapes de cet algorithme sont les suivantes :

- **Augmentation des concepts par activation des liens de référence** : chaque concept c_i de θ^* 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 :

$$score(c_k) = \begin{cases} score(c_k) + \sum_{c_j} (\alpha_R * score(c_j)) / (\alpha_R + 1) & \text{si } e_{jk} \in R \\ score(c_k) + \sum_{c_j} (\alpha_S * score(c_j)) / (\alpha_S + 1) & \text{si } e_{jk} \in S \end{cases}$$

Algorithme 1 Algorithme de propagation des scores des concepts

Entrée : θ^* est l'ensemble initial de concepts activés par l'information agrégée issue de l'évaluation de q^*

Sortie : $G_{q^*} = (V_{q^*}, E_{q^*})$ le graphe sémantique résultat
 $\theta^* = \{c_1, c_2, \dots, c_n\}$, $ListGraphs = \emptyset$

pour chaque concept $c_i \in \theta^*$ **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 $c_k \in \theta^*$ **alors**

$\theta^* = \theta^* - \{c_k\}$

$Queue.PushElement(c_k)$

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

$G_{q^*} = argmax_{ListGraphs(G_i)}(w(G_i));$

- **Extraction des graphes de concepts** : à partir de l'ensemble des concepts initiaux dans θ^* 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 basée 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^*} associé à la requête q^* 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 :

$$w(G_i) = \sum_{c_k \in G_i} score(c_k)$$

5.2.3 Évolution du profil utilisateur au sein d'une même situation de recherche

L'évolution du profil de l'utilisateur le long d'une même situation de recherche se fait par combinaison linéaire des listes de concepts composant le profil de la situation et le profil de la nouvelle requête. La combinaison linéaire consiste à privilégier les concepts récurrents dans la situation et issus de l'évaluation de la nouvelle requête.

Selon ce principe, l'évolution du profil d'une situation S^i consiste à définir une fonction linéaire qui permet :

1. d'augmenter le poids des concepts de la requête récurrents dans la situation S^i via son profil,
2. d'atténuer le poids des concepts non récurrents en utilisant un facteur d'amortissement η .

Soient G_u^i et G_q les représentations sémantiques associées respectivement au profil de la situation S^i de l'utilisateur u et de la requête courante q , on calcule le nouveau poids d'un concept c_j dans le profil de l'utilisateur G_u^i comme suit :

$$score_{G_u^i}(c_j) = \begin{cases} \eta * score_{G_q}(c_j) + (1 - \eta) * score_{G_u^i}(c_j) & \text{si } c_j \in G_u^i \\ \eta * score_{G_q}(c_j) & \text{sinon} \end{cases} \quad (3.4)$$

où $score_{G_u^i}(c_j)$ est le poids du concept c_j dans le profil de l'utilisateur G_u^i , $score_{G_q}(c_j)$ est le poids du concept c_j dans le profil de la requête G_q .

6 Approche de raisonnement à partir de cas pour l'accès personnalisé à l'information

Dans cette section nous présentons notre approche d'exploitation des profils situationnels dans le processus d'accès à l'information pour une tâche de recherche *ad hoc*. Nous nous sommes intéressés à un processus d'accès à l'information initié explicitement par la requête d'un utilisateur donné. C'est une démarche de recherche d'information active, contrairement à la recherche d'information proactive mise en œuvre dans les systèmes de recommandation ou de filtrage d'information, visant l'adaptation à un utilisateur unique et non à un groupe d'utilisateurs, comme abordé dans les travaux portant sur le filtrage collaboratif.

A un moment donné, notre modèle utilisateur peut contenir un ou plusieurs profils situationnels. Un seul d'entre eux correspond le mieux à la requête et la situation de recherche courante de l'utilisateur. Cela nécessite de choisir le profil de l'utilisateur le plus adéquat à utiliser pour la personnalisation de la requête courante. La mise en œuvre de notre modèle d'accès personnalisé à l'information implique l'exploitation du modèle dynamique de l'utilisateur dans le processus de RI. Un tel système ne peut s'envisager qu'en étant dynamique, sans cesse en état d'apprentissage.

Notre solution consiste à utiliser la technique du raisonnement à partir de cas comme notre approche générale pour l'exploitation du modèle de l'utilisateur dans un accès personnalisé à l'information (qui sera abrégé dans la suite RàPC). La RàPC est une approche de résolution de problèmes basée sur la réutilisation par analogie d'expériences passées appelées "cas" au cours d'un cycle de raisonnement [11]. Chaque cas est décrit par un couple (prémisse, valeur). La *prémisse* ou problème est la description du cas avec des attributs caractéristiques. Tandis que la *valeur* est le résultat du raisonnement sur la prémisse ou la solution du problème. Les cas sources représentent des expériences passées et sont stockés dans une base de cas. Lorsqu'un nouveau problème est à résoudre, un *cas cible* est construit avec la partie *valeur* inconnue, celle-ci devant être déduite par le raisonnement. Le cycle de raisonnement à partir de cas est illustré dans la figure 5.2(a), il est composé de quatre phases : recherche d'un cas similaire, réutilisation du cas similaire, révision et apprentissage.

Dans la suite, nous explicitons l'application de la technique RàPC pour la résolution de notre problème de personnalisation de la recherche. Plus spécifiquement, nous décrivons la description de nos cas puis le processus de raisonnement appliqué à notre base de cas.

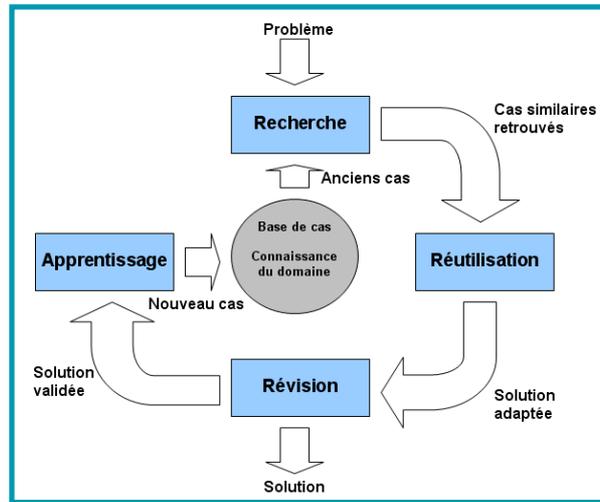


FIG. 3.5 – Le cycle de raisonnement à partir de cas (RàPC).

6.1 Représentation des cas

Dans une technique RàPC, les connaissances sur la façon d'accomplir les buts, sont représentées dans le système par un ensemble de cas sources.

Dans notre approche de personnalisation nous réutilisons des cas principalement composés de :

1. Une partie prémisses (problème) : contient la description de la situation S^i de l'utilisateur au cours d'une activité de recherche. Comme décrit dans la section 5.1, chaque situation sera composée d'un quadruplet (type de localisation, saison, jour de la semaine, temps de la journée). La prémisses sera utilisée pour calculer le degré de similarité entre le cas cible et les cas sources dans le processus de raisonnement.
2. Une partie valeur (solution) : contient le profil utilisateur G^i associé à cette situation S^i , représenté par un graphe de concepts de l'ODP, comme décrit dans la section 5.2.

Notre base de cas représente ainsi le modèle de l'utilisateur U . Chaque cas, de notre base de cas, modélise la liaison entre une situation et son profil utilisateur associé. Un cas est dénoté : $Cas = (S^i, G^i)$, et représente un élément spécifique du modèle de l'utilisateur U .

6.2 Processus de raisonnement

Le processus de raisonnement est lancé à chaque nouvelle activité de recherche de l'utilisateur en vue de sélectionner le profil adéquat à la personnalisation de la requête courante.

Dans cette section nous décrivons les phases du processus de raisonnement appliquées à notre base de cas :

1. La phase de recherche d'un cas similaire : permet de sélectionner le profil correspondant à la requête actuelle de l'utilisateur sur la base d'une mesure de similarité de situations pour la comparaison des situations de recherche précédentes.
2. La phase de réutilisation : permet de ré-ordonner les résultats de recherche en utilisant le profil sélectionné.
3. La phase de révision : permet de récupérer le *feedback* utilisateur.
4. La phase d'apprentissage et mise à jour de la base des cas : assure l'évolution et la mise à jour du modèle de l'utilisateur.

6.2.1 Phase de recherche d'un cas similaire

A chaque nouvelle requête q^* de l'utilisateur, un cas cible est construit avec comme prémisses la situation courante de l'utilisateur. Pour déterminer le profil de l'utilisateur le plus adapté à la requête courante q^* , la base des cas est parcourue pour comparer la situation courante S^* aux situations précédentes. Le cas le plus similaire est renvoyé pour la phase suivante.

Soit $PS = \{S^1, \dots, S^n\}$ l'ensemble des situations dans la base des cas de l'utilisateur. Le raisonnement consiste à sélectionner la situation S^{opt} qui vérifie :

$$S^{opt} = \arg \max_{S^i \in PS} \left(\sum_j \alpha_j \cdot sim_j (X_j^*, X_j^i) \right) \quad (3.5)$$

Où X_j^* (resp. X_j^i) est la valeur de la $j^{\text{ème}}$ dimension du vecteur de situation S^* (resp. S^i) (cf. section 5.1.3, équation 3.1), sim_j est la mesure de similarité pour la $j^{\text{ème}}$ dimension du vecteur de situation et α_j est le poids associé.

Les mesures de similarité, définies sur les dimensions du modèle de situation, sont basées sur la proximité des concepts dans l'ontologie utilisée pour la représentation du temps ou de la localisation. Ci-dessous, nous décrivons les mesures de similarité définies sur chacune des dimensions de localisation et de temps de nos situations de recherche.

6.2.1.1 Similarité pour la dimension de localisation

Pour une situation S la dimension de localisation X_l est donnée par la taxonomie des lieux associée à l'ontologie géographique (comme présenté dans la section 5.1.1. La similarité entre deux types de localisation X_l^* et X_l^i dépend de leur degré de proximité dans la taxonomie. En effet, la similarité entre *musée* et *théâtre* est plus grande que la similarité entre *musée* et *hôpital*.

Nous utilisons la mesure de similarité décrite dans [262] définie par :

$$sim_{location}(X_l^*, X_l^i) = \frac{2 * depth(lcs)}{(depth(X_l^*) + depth(X_l^i))} \quad (3.6)$$

Où lcs est le concept le plus spécifique (*Least Common Subsumer*) qui subsume les deux concepts à comparer X_l^* et X_l^i , et $depth$ est le nombre de nœuds dans le chemin d'un nœud donné à la racine dans la taxonomie.

6.2.1.2 Similarité pour les dimensions de temps

Nous avons proposé une mesure de similarité pour chacune des dimensions de temps de notre modèle de situation (décrites dans la section 5.1.2) pour tenir compte de la particularité de chaque dimension.

- **La similarité pour la dimension de saison** : pour une situation S , la dimension de saison X_u est un élément de la liste $\{automne, hiver, printemps et été\}$. Nous supposons que les situations sont plus similaires pour des saisons successives. Nous définissons la fonction f qui associe à chaque élément de la liste des saisons une valeur dans $\{a_1 = \pi/2, a_2 = 0, a_3 = -\pi/2 \text{ et } a_4 = -\pi\}$. La fonction de similarité entre les saisons X_u^* est alors donnée comme suit :

$$sim_{season}(X_u^*, X_u^i) = 1 - \frac{|f(X_u^*) - f(X_u^i)|}{\max_{(j,k) \in [1..4]} |a_j - a_k|} \quad (3.7)$$

où $f(X_u^*)$ et $f(X_u^i)$ est un nombre réel assigné à l'élément X_u^* et X_u^i de la dimension saison de la situation S^* respectivement S^i , et $\max_{(j,k) \in [1..4]} |a_j - a_k|$ est la différence maximale entre deux valeurs de la liste $\{a_1, a_2, a_3 \text{ et } a_4\}$.

- **La similarité pour la dimension jour de la semaine** : pour une situation S la dimension jour de la semaine X_v est un élément de la liste $\{jour de travail, fin de semaine, jour férié, vacances\}$. Nous supposons que les situations sont plus similaires en jours de repos (jour férié, fin de

semaine, vacances) qu'en jours de travail. La fonction de similarité entre deux jours de la semaine X_v^* et X_v^i est définie comme suit :

$$sim_{week}(X_v^*, X_v^i) = \begin{cases} 1 & \text{if } X_v^* = X_v^i \\ 1/2 & \text{if } X_v^* \text{ et } X_v^i \text{ sont des jours de repos} \\ 0 & \text{sinon} \end{cases} \quad (3.8)$$

- **La similarité pour la dimension moment de la journée** : pour une situation S la dimension moment de la journée X_w est un élément de la liste $\{\text{matin}, \text{midi}, \text{après-midi}, \text{soir}, \text{nuit}\}$. Nous supposons que les situations sont plus similaires dans les périodes de travail (matin et après-midi) que dans les autres périodes de la journée. La similarité entre deux moments de la journée X_w^* et X_w^i est donnée par :

$$sim_{day}(X_w^*, X_w^i) = \begin{cases} 1 & \text{if } X_w^* = X_w^i \\ 1/2 & \text{if } X_w^* \text{ et } X_w^i \text{ sont des temps de travail} \\ 0 & \text{sinon} \end{cases} \quad (3.9)$$

6.2.2 Réutilisation : Ré-ordonner les résultats de recherche

Le profil utilisateur G^{opt} correspondant à la situation la plus similaire S^{opt} sélectionnée à la suite de la phase précédente sera utilisé pour ré-ordonner les résultats de la recherche à retourner à l'utilisateur. Pour assurer une meilleure précision des résultats de la recherche, le processus de ré-ordonnement aura lieu uniquement si la condition suivante est vérifiée :

$$sim(S^*, S^{opt}) \geq \beta \quad \text{où } \beta \text{ est une valeur seuil.}$$

Le profil utilisateur correspondant G^{opt} est alors utilisé pour ré-ordonner les résultats de recherche retournés par le système de RI en réponse à la requête courante q^* .

Les algorithmes de ré-ordonnement appliquent généralement des fonctions sur le rang ou le score d'appariement requête-document ou alors sur l'ordre des rangs des résultats restitués par le système de RI. Notre fonction de ré-ordonnement est basée sur la combinaison des scores d'appariement

original et personnalisé du document. Le score personnalisé du document est calculé selon une mesure de similarité entre son vecteur représentatif d_k et le vecteur conceptuel représentatif du profil utilisateur adéquat G^{opt} .

Le calcul de ce score est donné selon la formule suivante :

$$score_c(d_k, G^{opt}) = \sum_{c_j \in G^{opt}} sw(c_j) * \cos(\vec{d}_k, \vec{c}_j) \quad (3.10)$$

Où $sw(c_j)$ est le poids du concept c_j dans le centre d'intérêt G^{opt} .

Le score final du document sera ainsi calculé par combinaison de son score d'appariement original normalisé et son score personnalisé normalisé selon la formule suivante

$$score_f(d_k) = (1 - \gamma) * score_o(q^*, d_k) + \gamma * score_c(d_k, G^{opt}) \quad (3.11)$$

Où γ varie dans l'intervalle $[0..1]$.

6.2.3 Révision

Après la phase de ré-utilisation, la phase de révision consiste à récupérer le résultat de l'exploitation du cas similaire.

Dans notre cadre d'application de RI, nous récupérons le *feedback* utilisateur qui servira comme retour pour la construction d'un nouveau cas ou la mise à jour d'un cas existant. Le *feedback* utilisateur sera représenté par la liste des résultats jugés implicitement pertinents par l'utilisateur, notamment la liste des documents qu'il a *cliqués* ou visités.

Le *feedback* est exploité pour construire le profil utilisateur associé à la requête courante q^* , comme décrit dans la section 5.2.

6.2.4 Apprentissage et mise à jour de la base des cas

Dans cette phase, la base des cas est mise à jour par l'ajout de nouveaux cas ou par la mise à jour des cas existants en intégrant le profil de la requête G_{q^*} appris pour la requête courante q^* .

Selon la valeur de similarité entre la situation S^* et la situation la plus similaire S^{opt} , on peut distinguer deux scénarios possibles : mise à jour d'un cas existant ou ajout d'un nouveau cas.

6.2.4.1 Ajout d'un nouveau cas

Pour assurer une meilleure précision lors du processus de personnalisation, nous avons préféré la spécification des cas sur la généralisation. En d'autres termes, nous avons établi qu'un nouveau cas sera ajouté à la base dès que les deux situations de recherche courante S^* et la plus similaire S^{opt} ne sont pas égales.

Plus spécifiquement, nous définissons la règle suivante :

si $sim(S^*, S^{opt}) \neq 1$ **alors**

Former un nouveau cas composé de la situation courante S^* et le profil de la requête courante G_{q^*} .

Ajouter le nouveau cas à la base des cas.

finsi

6.2.4.2 Mise à jour d'un cas existant

La mise à jour d'un cas existant se fait si et seulement si la situation courante correspond exactement à la situation similaire. Dans ce cas le profil associé à la situation similaire est mis à jour en utilisant la profil de la requête courante.

Plus spécifiquement, la règle est la suivante :

si $sim(S^*, S^{opt}) = 1$ **alors**

Mettre à jour le profil G^{opt} associé à la situation similaire S^{opt} en utilisant le profil de la requête courante G_{q^*} , comme décrit dans la section 5.2.3.

Remplacer le cas existant dans la base des cas par le cas mis à jour.

finsi

7 Évaluation expérimentale

Dans le but de valider nos propositions, nous avons mené des expérimentations permettant d'évaluer l'impact de l'intégration des centres d'intérêt flexibles de l'utilisateur sur les performances des résultats de recherche.

Dans le chapitre 2 nous avons montré les limites de l'approche de l'évaluation classique à la *Cranfield* en présence du contexte et nous avons présenté les nouvelles démarches d'évaluation adoptées dans le domaine. Cependant, les nouveaux cadres d'évaluation proposés restent propriétaires ou très spécifiques.

En l'absence d'un cadre standard d'évaluation d'un modèle d'accès personnalisé à l'information dans un environnement mobile, nous avons proposé deux cadres d'évaluation :

1. évaluation expérimentale selon un cadre basé sur une simulation contextuelle.
2. évaluation expérimentale basée sur une étude journalière.

Nous explicitons dans ce qui suit la méthodologie menée pour construire chacun des cadres d'évaluation, la stratégie de validation adoptée et les résultats expérimentaux obtenus.

7.1 Évaluation expérimentale selon un cadre d'évaluation par simulation contextuelle

Au début de notre processus d'évaluation, nous nous sommes confrontés au problème de l'absence de données contextuelles dans les collections de test standards. Pour contourner ce problème, et dans le but d'avoir une première évaluation rapide de notre approche, nous avons opté pour une approche d'évaluation par simulation contextuelle. Nous nous sommes donc proposés de simuler des situations de recherche d'utilisateurs mobiles.

7.1.1 Principe

Notre idée générale consiste à simuler quelques scénarios de recherche où l'utilisateur mobile se retrouverait dans différents contextes, d'assigner à ces contextes des besoins informationnels probables, et puis de dérouler une activité de recherche pour ces besoins informationnels. Pour faciliter notre tâche, nous avons représenté les situations de l'utilisateur mobile par une seule dimension, notamment des contextes géographiques.

Nous avons alors défini un ensemble de six (6) scénarios d'utilisation comme partie de notre cadre d'évaluation. Chaque scénario d'utilisation est défini autour d'un contexte géographique donné. Nous avons simulé six (6) contextes géographiques qui sont décrits par un type de classe parmi : *librairie*, *cinéma*, *musée*, *jardin*, *zoo* et *magasin de musique*. Chaque scénario d'utilisation est de plus composé d'un ensemble de requêtes hypothétiques associées à son contexte géographique.

Sur la base de nos scénarios d'utilisation, nous avons élaboré une collection de test, défini un protocole d'évaluation et des mesures de pertinence, et nous avons évalué notre approche de personnalisation. Nous présentons dans la

suite de cette section la spécification de chacune de ces composantes puis nous présentons et discutons les résultats expérimentaux obtenus.

7.1.2 Ressources expérimentales

Nous avons construit une collection de test composée : d'un ensemble de requêtes contextuelles, d'un corpus de documents, de jugements de pertinence des documents par rapport à ces requêtes, ainsi que des profils utilisateurs. Nous décrivons ici, l'ensemble de nos données expérimentales.

1. **Collection de requêtes contextuelles** : l'ensemble des requêtes sont imaginées en gardant à l'esprit le cadre applicatif de notre technique de personnalisation, ici le cadre d'un environnement mobile et les situations géographiques hypothétiques que l'on a définies dans nos scénarios d'utilisation. Nous avons simulé pour chaque scénario d'utilisation cinq (5) requêtes qui peuvent être naturellement exprimées par un utilisateur se trouvant dans le contexte géographique associé. Nous avons créé au total un ensemble de trente (30) requêtes à travers les six (6) scénarios d'utilisation. Vingt et cinq (25) requêtes sont différentes (respectant la consigne dans [254] pour le nombre minimum de requête à respecter pour que l'évaluation soit crédible). La table 3.2 illustre la répartition de l'ensemble de nos requêtes sur les six (6) contextes géographiques.

Scénario	Contexte spatial	Requêtes
1.	Zoo	"shark", "rabbit", "jaguar", "cobra", "fennec"
2.	Magasin de musique	"glorious", "sunrise", "stay positive", "human nature", "fire burning"
3.	Cinema	"2012", "seraphine", "shark", "the road", "harry brown"
4.	Librairie	"dan brown", "the road", "harry potter", "da vinci", "scarlet poppy"
5.	Jardin	"iris", "lotus", "dahlia", "scarlet poppy", "water lilies"
6.	Musée	"da vinci", "sunflowers", "woman with a parasol", "edgar degas", "water lilies"

TAB. 3.2 – Les contextes géographiques et leurs requêtes associées

Comme les requêtes mobiles sont plutôt courtes et par la même peuvent être ambiguës, nos requêtes sont généralement courtes (la taille des requêtes ≤ 3) et certaines d'entre elles sont par conséquent ambiguës (eg. *jaguar*). Certaines requêtes ambiguës sont de plus intégrées dans plus d'un

scénario d'utilisation, par exemple la requête "*water lilies*" est testée sous les deux contextes "*jardin*" et "*musée*". Cinq (5) requêtes sont communes à au moins deux scénarios d'utilisation, elles sont marquées en couleur dans la table 3.2.

Notre but dans cette démarche consiste à vérifier si la prise en compte des centres d'intérêts attachés aux contextes géographiques peut améliorer la performance du système de RI pour répondre à ce type de requêtes ambiguës.

2. **Collection de documents** : nous avons soumis chaque requête de notre ensemble des 25 requêtes de test au web via l'API de recherche yahoo BOSS. Nous avons récupéré pour chaque requête l'ensemble des 150 pages web en tête de la liste des réponses retournées par le moteur web après avoir éliminé les pages vides ou les doublons.

Nous avons aspiré ces pages web pour l'ensemble des 25 requêtes de test pour construire une collection de documents de 3750 pages web. Cette collection a été ensuite indexée en local en utilisant le système de RI *Mercur*⁴ [33].

3. **Jugements de pertinence en contexte** : les jugements de pertinence pour les documents de la collection sont collectés, comme dans les campagnes d'évaluation en RI, via un outil d'assessment (en ligne).

Pour ce faire, nous avons conduit une étude utilisateur, où nous avons invité 6 participants à l'étude, à juger la pertinence des documents renvoyés pour chaque requête. Les participants sont des doctorants de notre laboratoire. Chaque participant s'est vu attribuer un cas d'utilisation et donc 5 requêtes contextuelles. Le jugement de pertinence est effectué en tenant compte de la requête mais aussi de son contexte géographique. Pour cela, nous avons enrichi chaque requête dans les scénarios d'utilisation par une partie narrative décrivant la pertinence d'un document. La Table 3.3 donne un exemple de l'enrichissement du scénario d'utilisation *musée* avec une partie narrative décrivant la pertinence d'un document par rapport à une requête et son contexte géographique. Les participants ont été sollicités en deux étapes pour jugé deux listes de 20 documents par requête, de façon transparente sans aucun indice de l'origine des documents :

- Un premier jugement a été établi pour la liste des résultats initiaux d'un SRI standard, ici les résultats retournés par Yahoo BOSS.
- Un second jugement est effectué sur la liste des résultats personnalisés.

Les jugements de pertinence sont faits sur une échelle graduelle : pertinent, peu pertinent, et non pertinent.

⁴http://www.irit.fr/SIG_RFI/mercure/index.php

Contexte	Termes de la requête	Narrative
Musée	"da vinci"	<i>A document is relevant if it speaks about da Vinci painter and or his paintings</i>
	"sunflowers"	<i>A document is relevant if it speaks about the painting sunflowers and or its painter Van Gogh and or his paintings</i>
	"woman with a parasol"	<i>A document is relevant if it speaks about the painting woman with a parasol and or its painter Claude Monet and or his paintings</i>
	"edgar degas"	<i>A document is relevant if it speaks about painter Edgar Degas and or his paintings</i>
	"water lilies"	<i>A document is relevant if it speaks about the painting water lilies and or its painter Claude Monet and or his paintings</i>

TAB. 3.3 – Enrichissement du scénario d'utilisation "musée" par une partie narrative décrivant la pertinence d'un document dans le contexte.

4. **Profil utilisateur** : nous avons associé à chaque scénario d'utilisation un profil utilisateur. Les profils utilisateurs sont simulés selon un algorithme qui les génère en utilisant des interactions hypothétiques de l'utilisateur simulées pour chaque requête. Les interactions hypothétiques de l'utilisateur sont représentées par les jugements de pertinence manuels des participants qui remplaceraient ainsi les *click-through data* de l'utilisateur dans un scénario d'utilisation réel. L'algorithme de simulation des profils utilisateurs est intégré à la stratégie d'évaluation décrite ci-dessous.

7.1.3 Stratégie d'évaluation

Généralement, l'évaluation de la performance d'une technique de RI consiste à comparer la nouvelle technique par rapport à une *baseline*. Nous avons alors considéré les résultats de la recherche retournés par Yahoo BOSS comme notre *baseline*. Notre démarche expérimentale consiste alors à évaluer l'efficacité de notre approche personnalisée en utilisant le profil de l'utilisateur comparativement à la *baseline*, selon notre séquence de scénarios d'utilisation.

Comme notre approche de personnalisation se base sur l'apprentissage d'un profil utilisateur le long de ses interactions avec le SRI, puis de l'exploiter

pour la personnalisation des requêtes futures de l'utilisateur, notre stratégie de validation se base alors sur deux phases : une phase d'apprentissage du profil utilisateur puis une phase de test de l'efficacité du profil appris.

Avec un petit nombre de requêtes (5) par utilisateur, nous avons jugé utile d'adopter une stratégie de validation croisée, pour pouvoir apprendre avec un maximum de requêtes et tester sur toutes les requêtes. Notre choix de cette stratégie de validation est de plus appuyé par diverses raisons : absence d'un ordre naturel des requêtes, éviter de choisir quelle requête affecter à l'ensemble d'apprentissage ou de test, et par la même éviter tout biais dans les résultats. La validation croisée [159] ou la *k-fold cross validation* est une méthode d'évaluation qui consiste à diviser la collection de test en k sous ensembles de tailles égales (approximativement), d'utiliser $k - 1$ sous ensembles pour l'apprentissage et le $k^{\text{ème}}$ sous ensemble pour le test. Ce processus est réitéré k fois, chaque fois un des k sous-ensembles est utilisé comme l'ensemble de test et les autres $k-1$ sous-ensembles reconstituent l'ensemble d'apprentissage.

La stratégie de validation est illustrée par la figure 3.6 et peut être décrite par l'algorithme ci-dessous :

1. Pour chaque utilisateur associé à un scénario d'utilisation, on construit 5 ensembles d'apprentissage/test, en prenant pour chaque ensemble 4 requêtes différentes pour le sous-ensemble d'apprentissage et la requête restante pour le sous-ensemble de test,
2. Pour chaque sous-ensemble d'apprentissage, on apprend un profil utilisateur G en exploitant un sous-ensemble de documents pertinents listés dans le fichier de jugements de pertinence comme décrit dans la section (5.2.1). Nous avons utilisé les 5 premiers documents selon les jugements de pertinence établis par l'évaluateur.
3. Pour chaque requête de test, on utilise le profil appris sur son sous-ensemble d'apprentissage pour ré-ordonner ses résultats de recherche, comme décrit dans la section (6.2.2). En l'absence d'un score initial des documents retournés par Yahoo boss, la procédure de ré-ordonnement est basée uniquement sur le score personnalisé du document, en mettant le paramètre $\gamma = 0$ dans l'équation 3.11.
4. A la fin, on calcule une moyenne des performances sur l'ensemble des jeux de test.

Dans la suite nous présentons les métriques d'évaluation adoptées pour le calcul des performances.

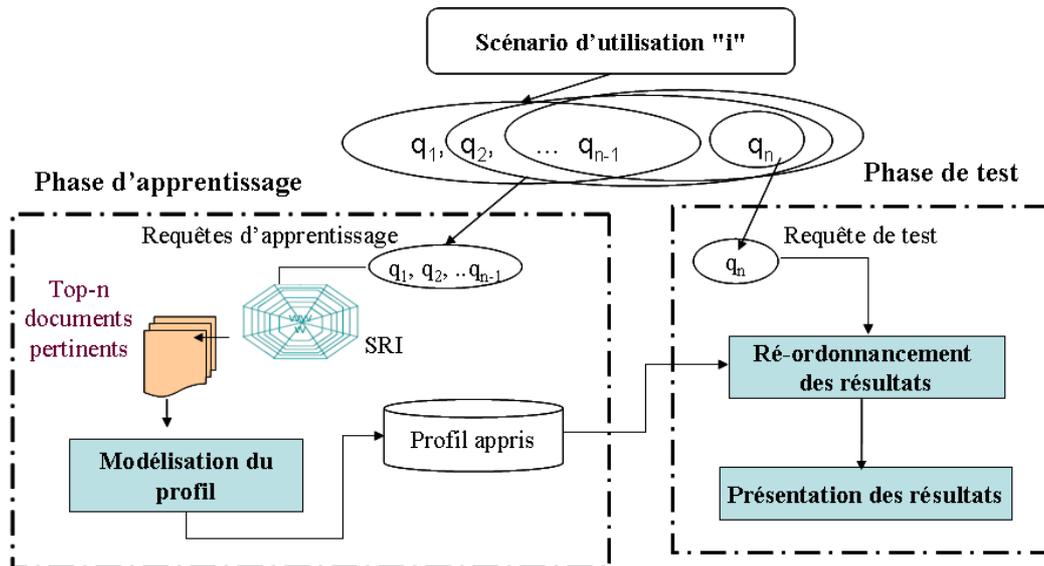


FIG. 3.6 – Stratégie de validation croisée appliqué sur chaque cas d'utilisation.

7.1.4 Métriques d'évaluation

Dans ce cadre d'évaluation, nous avons utilisé la mesure classique de la précision pour les X premiers documents restitués (précision@ X) et une mesure orientée rang DCG "Discounted Cumulative Gain" [111]. En effet, il est très important d'améliorer la qualité des documents au top de la liste de la recherche pour un utilisateur mobile, généralement ayant moins de temps à consulter une longue liste de résultats étant engagé dans d'autres activités. Plus spécifiquement, ces deux métriques permettent d'évaluer les performances de recherche selon deux principaux aspects :

1. *Précision@ X*

C'est la proportion de documents pertinents dans les X premiers documents retrouvés pour chaque requête. 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. On retient les précisions pour les 5, 10, 15 et 20 premiers documents notés respectivement P5, P10, P15 et P20. A la fin, des moyennes de toutes ces précisions@ X (pour chaque $X=5, 10, 15$ et 20) sont calculées sur toutes les requêtes de test.

2. *DGC@ X*

C'est une mesure orientée-rang proposée dans un contexte de pertinence graduelle. Elle traduit deux principes :

- D'une part, un gain croissant en fonction du degré de pertinence des documents. En effet, étant basée sur une pertinence graduelle, les do-

cuments très pertinents sont plus intéressants que les documents peu pertinents.

- D'autre part, un facteur décroissant en fonction du rang X du document. En effet, un document a d'autant moins d'intérêt pour l'utilisateur s'il est loin dans la liste de résultats, car il est d'autant moins probable que l'utilisateur accède à ce document-là.

La mesure DCG à un rang X est alors définie comme :

$$DCG@X = rel_1 + \sum_{i=2}^X \frac{rel_i}{\log_2 i}$$

où rel_i est la pertinence graduelle du document retourné au rang i . La fonction log est utilisée comme un facteur d'atténuation du gain de pertinence apporté par un document en fonction de son rang. Ici, nous retenons les gains cumulatifs aux rangs 10 et 20, notés respectivement DCG10 et DCG20.

7.1.5 Résultats et discussion

L'objectif de nos expérimentations est d'évaluer l'efficacité de la recherche personnalisée qui intègre les centres d'intérêt contextuels de l'utilisateur, plus spécifiquement ici, le contexte géographique.

Nous comparons donc l'efficacité de notre recherche personnalisée par rapport à la *baseline* en utilisant les deux mesures de performance Précision@X et DCG@X. La significativité des tests statistiques calculés pour les différences observées entre les moyennes des deux systèmes, est également rapportée. Les p-valeurs de significativité sont calculées avec le "t-test de Student" pairée avec un seuil de significativité fixé à de 0.05, noté "*" dans les résultats suivants.

Comme première évaluation de la performance, nous rapportons ici la comparaison de l'efficacité de notre recherche personnalisée par rapport à notre *baseline*, représentée par les résultats initiaux retournés par Yahoo BOSS, en utilisant la métrique de la précision@X.

Le tableau 3.4 présente les résultats de comparaison des moyennes de la précision à différents rangs entre les deux systèmes.

Les résultats montrent qu'en général de meilleures performances sont obtenues par notre approche de personnalisation en termes de précision moyenne aux différents points de coupure, atteignant une amélioration significative de 87.50% pour P5, 63.56% pour P10, 53.49% pour P15 et 50.92% pour P20 par comparaison à la *baseline*.

Système/ amélioration	Moyenne des Précision@X à :			
	P5	P10	P15	P20
<i>baseline</i>	0.37	0.39	0.38	0.36
notre approche	0.70	0.64	0.59	0.55
amélioration	87.50% (*)	63.56% (*)	53.49% (*)	50.92% (*)

TAB. 3.4 – Comparaison des moyennes des précisions@X entre notre approche de personnalisation et la *baseline* sur toute les requêtes de test.

Cependant, au niveau local, nous avons observé que l'amélioration de la précision diffère entre les requêtes. La figure 3.7 montre un exemple de la variation de l'amélioration de la précision entre les requêtes du scénario "Musée".

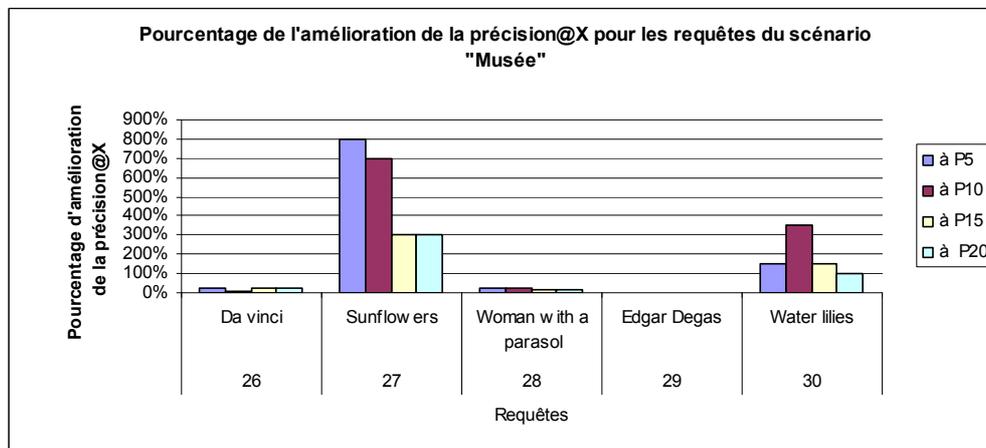
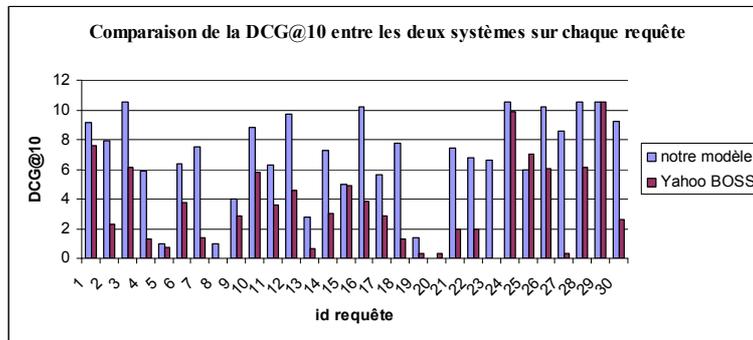


FIG. 3.7 – Cadre d'évaluation basé étude journalière.

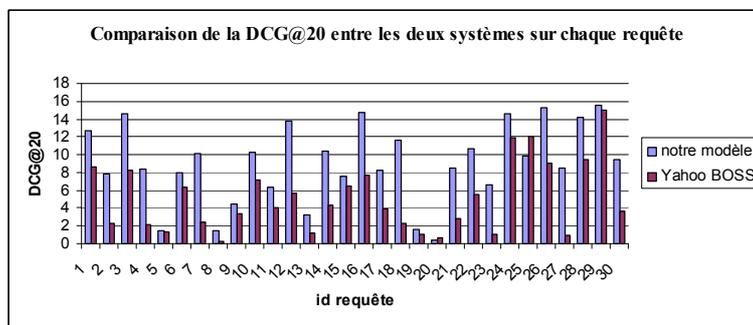
Cela peut être probablement dû à la différence entre le degré d'ambiguïté des requêtes, qui ne peut être expliquée uniquement par la différence de taille des requêtes. En effet, elle dépend également du contenu des documents présents dans la collection, plus les documents de la collection (dans notre cas composée des 150 premiers documents retournés par Yahoo BOSS pour chaque requête), couvrent les différents sens d'une requête, plus on a de chance d'améliorer la précision pour cette requête.

Pour mieux estimer l'efficacité de notre modèle, nous calculons dans un second temps, les valeurs du gain cumulatif reportés aux deux points de coupure 10 et 20. Les figures 3.8(a) et 3.8(b) comparent l'efficacité de la liste initiale des résultats de la recherche retournée par Yahoo BOSS et celle re-ordonnée par notre modèle personnalisé sur chacune des requêtes.

Nous observons qu'en général, notre approche augmente la DCG@10 et la DCG@20 initiales obtenues par la recherche standard et améliore donc la qualité des documents retournés en haut de la liste des résultats de recherche. De même que pour la métrique de Précision@X, nous remarquons ici également une variation de l'amélioration du gain cumulatif entre les différentes requêtes.



(a)



(b)

FIG. 3.8 – Comparaison de l’efficacité entre notre modèle personnalisé et Yahoo BOSS sur chaque requête en termes. (a) de la DCG@10, (b) de la DCG@20.

En conclusion, ces premiers résultats montrent que notre approche de personnalisation améliore les performances de recherche pour les scénarios d’utilisation simulés.

Cela nous a encouragé à investir plus dans la démarche d’évaluation en exploitant de vrais scénarios de recherche avec de vrais utilisateurs et de vraies requêtes exprimées en contexte. Cela nous permettra de renforcer et d’appuyer ces résultats préliminaires pour bien réussir le test de l’efficacité de notre approche dans des contextes de recherche réels.

Dans la suite, nous présentons notre seconde démarche d’évaluation basée sur des scénarios d’utilisation réels construits en utilisant la méthode de l’étude journalière - *diary study*.

7.2 Évaluation expérimentale basée étude journalière

Dans ce second cadre d’évaluation, nous présentons deux contributions : en premier lieu, nous proposons d’exploiter la méthode ”des études journalière” [193] comme outil qui permettra de mener des évaluations avec de vrais utilisateurs et des besoins en information réels exprimés dans de vrais contextes

de recherche. En second lieu, nous proposons d'évaluer notre approche en utilisant deux protocoles d'évaluation possibles et de comparer les résultats obtenus selon les deux protocoles.

L'étude journalière ou *diary study* est une méthode qui a ses origines tant dans le domaine de la recherche en psychologie qu'en anthropologie. Dans sa forme la plus simple, elle consiste à étudier un échantillon représentatif de sujets notons des informations sur leur vie quotidienne *in situ* pour une période donnée. Les données ainsi recueillies peuvent être alors analysées de différentes façons selon leur nature.

Les études journalières sont présentées dans les premiers travaux de Rieman [193] comme un outil orienté "workplace" pour guider les approches d'évaluation basées laboratoires dans le domaine des HCI. Elles sont exploitées dans [223] en vue de l'analyse des besoins en information des utilisateurs mobiles.

Dans notre travail, nous proposons d'entreprendre une étude journalière comme un outil pour l'acquisition, *in situ*, d'une collection de requêtes mobiles avec leurs contextes externes (non captés par les collections classiques d'évaluation des SRI), notamment : le temps, la localisation, les centres d'intérêt de l'utilisateur, etc. Les entrées journalières sont utilisées comme les briques de base qui composent la collection de test de notre cadre d'évaluation.

Nous présentons dans la suite de cette section la méthodologie menée pour la construction de notre cadre d'évaluation, la collection de test, la stratégie d'évaluation définie et les mesures d'évaluation adoptées, ainsi que la mise en œuvre de ce cadre d'évaluation en conduisant notre propre étude journalière.

7.2.1 Méthodologie de construction du cadre d'évaluation

L'objectif de notre cadre est de permettre l'évaluation de l'efficacité de technique de personnalisation contextuelle pour la RI mobile, en premier stade de développement. Ce type de technique implique la prise en compte du contexte de recherche des utilisateurs mobiles à savoir : centres d'intérêts, lieu, temps, etc à la fois dans le processus de développement mais aussi dans le processus d'évaluation. Ce cadre peut être appliqué pour évaluer n'importe quelle approche de personnalisation exploitant un ou plusieurs aspects du contexte d'un utilisateur mobile. La figure 3.9 illustre le processus général du cadre d'évaluation que nous définissons dans cette perspective. Ce processus peut être sommairement décrit par les étapes suivantes :

1. Tout d'abord, mener une étude journalière *in situ*, en invitant un ensemble d'utilisateurs potentiels, appelés participants, à enregistrer leurs requêtes ainsi que pour chaque requête son contexte : où et quand elle s'est produite. Pour bien mener son étude journalière, certaines directives

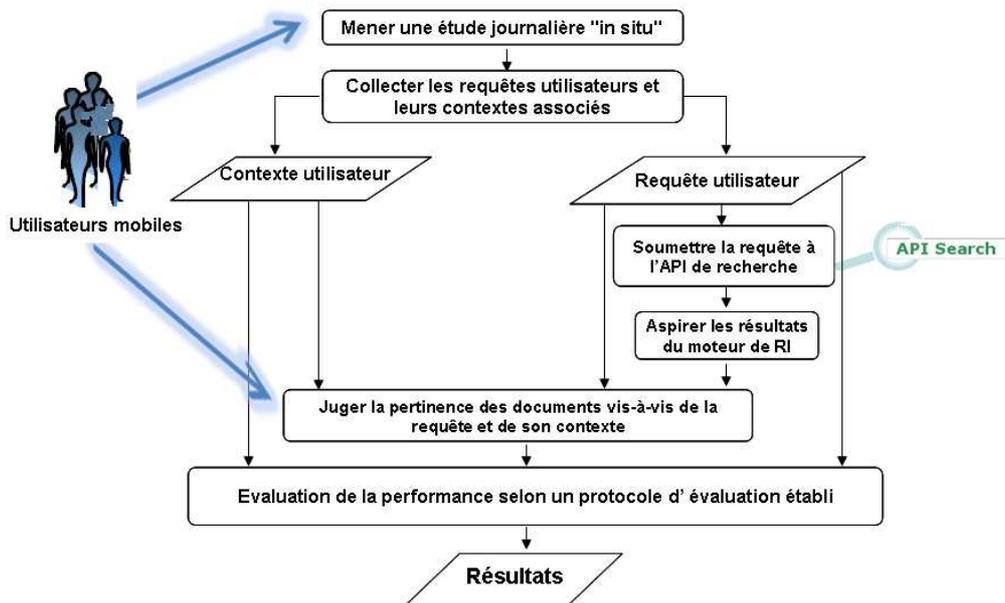


FIG. 3.9 – Cadre d'évaluation basé étude journalière.

générales sont à respecter, notamment :

- fixer le nombre de participants et la durée de l'étude journalière ;
 - s'assurer que tous les participants ont fait déjà l'expérience d'utiliser des moteurs de recherche sur le Web, en utilisant un PC ou un téléphone portable ;
 - établir une description des activités que vous aimeriez que le participant note, notamment : demander la notation des éléments contextuels (la date, l'heure, le lieu, etc) et la requête de l'utilisateur mobile ;
 - pour éviter des participants oubliant d'enregistrer leurs activités, envoyer des rappels périodiques pour rappeler les participants à leur tâche.
2. Ensuite, traiter les entrées journalières pour en extraire les contextes, et les requêtes et soumettre ces dernières à un moteur de recherche Web via une API,
 3. Après, aspirer les N premiers documents résultats renvoyés par le moteur de recherche, et demander aux participants de donner les jugements de pertinence pour ces documents selon leurs requêtes et leurs contextes associés,
 4. Finalement, intégrer les requêtes utilisateurs et leurs contextes dans le/ou les protocoles d'évaluation.

Au terme de ces processus, nous obtenons une collection de test en contexte, composée d'une liste de requêtes contextuelles, une collection de documents avec des jugements de pertinence en contexte. Nous présentons dans la suite une description plus détaillée de cette collection de test.

7.2.2 Collection de test

7.2.2.1 Requêtes contextuelles

L'étude journalière fournit un ensemble de requêtes contextuelles. Alors que beaucoup d'information contextuelles peuvent être notées ou enregistrées, dans notre travail nous nous sommes focalisés sur le contexte spatio-temporel et les centres d'intérêts de l'utilisateur mobile. Plus formellement, nous représentons une requête contextuelle par le quadruplet :

$$Q_i^u = \langle q_i^u, l_i^u, t_i^u, g_i^u \rangle$$

où q_i^u (resp. l_i^u , t_i^u , et g_i^u) représente la *ième* requête (resp. temps, localisation, et centre d'intérêt) des entrées journalières d'un utilisateur u .

Chaque requête contextuelle, sera de plus annotée par une description du besoin en information précis de l'utilisateur et un résumé narratif décrivant un document pertinent pour cette requête.

Selon les besoins de l'application en cours, les données de localisation (l_i^u) et de temps (t_i^u) peuvent être exprimées par des données de bas niveau (si relevées des capteurs mobiles) ou par des concepts sémantiques.

Les centres d'intérêts de l'utilisateur (g_i^u) peuvent être spécifiés manuellement par les participants eux mêmes ou appris d'une façon automatique en exploitant les jugements de pertinence des documents des requêtes passées de l'utilisateur (voir ci-dessous) comme *feedback* utilisateur.

7.2.2.2 Documents

La collection de documents sera construite par la collecte des N premiers résultats de recherche renvoyés d'une API de recherche web (en accès libre) pour chaque requête démunie de son contexte. On veillera à éliminer les documents doublons ou vides de la liste et aussi à garder trace de l'ordre initial des documents, en vue d'une utilisation comme *baseline*.

7.2.2.3 Jugements de pertinence en contexte

Les jugements de pertinence pour les documents de la collection sont collectés, comme dans les campagnes d'évaluation en RI, via un outil d'assessment (en ligne). Pour ce faire, chaque participant à l'étude journalière, sera invité à juger la pertinence des documents renvoyés (parmi les top N aspirés) comme réponse à chacune de ses requêtes en tenant compte de sa requête mais aussi de son contexte. Les jugements de pertinence sont effectués sur une échelle graduelle : pertinent, peu pertinent, et non pertinent.

7.2.3 Stratégie de validation

Dans nos expérimentations, on vise à évaluer l'efficacité de la recherche personnalisée qui intègre les centres d'intérêt de l'utilisateur le long d'un ensemble de situations de recherche similaires. L'évaluation de l'efficacité consiste à comparer, pour chaque requête de test, les résultats initiaux retournés par un système de RI non personnalisé, notamment le moteur web utilisé lors de la construction de la collection de test considéré comme *baseline*, par rapport aux résultats obtenus par l'approche de personnalisation intégrant les centres d'intérêt de l'utilisateur comme une source d'évidence additionnelle.

Dans ce cadre, l'ensemble des requêtes des utilisateurs, sera divisé en deux sous ensembles : un sous-ensemble d'apprentissage pour apprendre les paramètres de la technique de personnalisation, et un sous ensemble de test pour évaluer l'efficacité de la technique. Avec un ensemble de K requêtes contextuelles par utilisateur, annotées par des informations temporelles, nous proposons deux protocoles d'évaluation appropriés : l'apprentissage/test dans un ordre chronologique et la validation croisée. La seule recommandation à observer alors est de respecter un *minimum* de 25 requêtes de tests [254] pour que les résultats du processus d'évaluation soient significatifs. Ces deux stratégies d'évaluation sont décrites dans ce qui suit.

7.2.3.1 Apprentissage/test dans un ordre chronologique

Cette première stratégie prend les requêtes dans leur ordre chronologique d'émission, utilise les requêtes $Q_1^u \cdot \cdot \cdot Q_{i-1}^u$ comme l'ensemble d'apprentissage, et teste sur les requêtes suivantes $Q_i^u \cdot \cdot \cdot Q_K^u$. Cette stratégie est la plus simple et la plus naturelle, cependant les résultats peuvent être critiqués pour être dépendant de la séquence des données telle qu'elle se présente.

7.2.3.2 Validation croisée *k-fold cross validation*

Notre seconde stratégie de validation consiste en un scénario qui se base sur la méthode de validation croisée et ce pour ne pas biaiser les résultats avec un seul jeu de test. Le processus de validation croisée est décrit comme pour le premier cadre d'évaluation 7.1.3. Cela consiste à diviser la collection de test en k sous ensembles de tailles égales d'utiliser $k - 1$ sous ensembles pour l'apprentissage et le $k^{\text{ème}}$ sous ensemble pour le test. Puis, réitérer le processus k fois, chaque fois un des k sous-ensembles est utilisé comme l'ensemble de test et les autres $k-1$ sous-ensembles reconstituent l'ensemble d'apprentissage. A la fin, calculer une moyenne des performances sur l'ensemble des jeux de test.

L'avantage de cette méthode comparativement à la première stratégie, est que toutes les requêtes sont utilisées à la fois pour l'apprentissage et pour le test et par conséquent on évite le biais lié à un choix figé de l'ensemble d'apprentissage.

Nous estimons que les deux stratégies sont applicables, et en dépit de leur différence notamment dans le nombre de requêtes qu'elles permettent de tester, on s'attend à ce que les résultats de l'évaluation soient cohérents entre les deux stratégies.

7.2.4 Métriques d'évaluation

De même que pour le cadre d'évaluation par simulation contextuelle, nous utilisons les métriques classiques de précision pour les X premiers documents restitués (précision@ X) et la mesure orientée rang de gain cumulatif (DCG) pour l'évaluation de l'efficacité de recherche personnalisée. Plus précisément, nous calculons des valeurs de précisions moyennes à X documents pertinents, $X \in \{5, 10, 15, \text{et} 20\}$, sur l'ensemble des requêtes de test. De même nous calculons des valeurs de gain cumulatif moyennes aux même points de coupure. Pour cela, nous utilisons la version normalisée du gain cumulatif (nDCG) qui est donnée par :

$$nDCG@X = \frac{DCG@X}{IDCG@X}$$

où IDCG@ X est le classement idéal des documents au rang X obtenu par un tri par pertinence décroissante des documents de la liste de résultats. Cette version normalisée permet de comparer les requêtes entre elles et de calculer un gain moyen sur toutes les requêtes pour obtenir une mesure globale de performance du SRI.

7.2.5 Application du cadre d'évaluation

Nous avons mené une étude journalière où nous avons demandé à des personnes volontaires de noter leurs requêtes alors qu'ils sont mobiles et de les annoter par le jour, l'heure et la place de leur émission. 7 participants (3 femmes et 4 hommes) dont l'âge varie entre 21 et 36 ans, ont participé à notre étude journalière (principalement des personnes de notre laboratoire). Tous les participants ont déjà une expérience dans la recherche sur le Web en utilisant un ordinateur de bureau ou un appareil mobile.

L'étude journalière a duré un mois et a généré 79 requêtes, avec une moyenne de 11,28 requêtes par participant (min=3, max=35). La taille des requêtes varie entre 1 et 5, avec une moyenne de 2,99. Le tableau 3.5 illustre des exemples de requêtes obtenues, chacune d'elle annotée par l'identifiant de l'utilisateur, l'heure, la date et le lieu d'émission de la requête.

ID user	Date	Heure	Place	Requête
1	20-févr	14h30	place de la concorde	"histoire obélisque"
2	27-févr	11h10	périphérique	"parking relais bordeaux"
6	16-févr	16h30	musée	"exposition beaubourg artistes"
7	02-mars	19h40	station bus	"tisseo horaire bus 2"

TAB. 3.5 – Un exemple des requêtes récupérées de l'étude journalière.

Nous avons extrait les informations sur la localisation et le temps associées à chaque requête et nous les avons représentées manuellement sous forme de concepts de nos ontologies temporelle et spatiale. Nous avons finalement obtenu 36 situations, avec une moyenne de 5 situations par utilisateur ($min=2, max=12$) et une moyenne de 3 ($min = 1, max = 8$) requêtes par situation.

Pour construire notre collection de documents, nous avons soumis les 79 requêtes au service de recherche en ligne Yahoo boss, et aspiré les 50 premières pages Web retournées pour chaque requête, obtenons une collection de 3950 documents. Chaque participant à l'étude journalière a ensuite jugé les pages Web résultats de ses requêtes via un outil de jugement en ligne ⁵. Les jugements de pertinence ont été établis sur une échelle de trois valeurs : pertinent, peu pertinent et pas pertinent.

Les centres d'intérêts des utilisateurs ont été générés sur la base du *feedback* des participants comme décrit dans la section (5.2.1) et sont intégrés dans la stratégie de validation décrite ci-dessus, comme suit :

⁵<https://osirim.irit.fr>

1. Pour chaque requête dans le sous ensemble d'apprentissage, on génère son centre d'intérêt G en exploitant un sous ensemble de documents pertinents listés dans le fichier de jugements de pertinence comme décrit dans la section (5.2.1). Nous avons utilisé les 5 premiers documents selon les jugements de pertinence établis par l'utilisateur.
2. On met à jour le centre d'intérêt G avec les documents des requêtes d'apprentissage dans le même groupe, en utilisant la formule 3.4.
3. On utilise ce centre d'intérêt pour réordonner les 50 premiers résultats de recherche de chaque requête de test, comme décrit dans la section (6.2.2).

En conclusion, cette première étude journalière nous a permis de vérifier la faisabilité de notre cadre d'évaluation et nous a fourni un premier ensemble de requêtes en contextes réels pour évaluer notre approche. L'étude journalière nous a également permis d'avoir un *feedback* des participants sur leurs requêtes et le contexte de leurs émissions et surtout leur perception personnel de la pertinence des documents retournés à leurs requêtes.

7.2.6 Résultats et discussions

Dans nos expérimentations, on vise à estimer le paramètre de mise en correspondance des situations ainsi que l'évaluation de l'efficacité de la recherche personnalisée qui intègre les centres d'intérêts de l'utilisateur.

7.2.6.1 Estimation expérimentale du paramètre de la mise en correspondance des situations

Le processus de mise en correspondance des situations permet d'identifier les situations les plus similaires tout en choisissant un seuil de similarité entre les situations. Le but ici est de fixer ou d'identifier la meilleure valeur du seuil β de la comparaison des situations similaires. Nous proposons de classer manuellement les situations similaires, puis de comparer ces classements manuels avec les résultats obtenus automatiquement en variant la valeur du seuil de similarité β (définie dans la section 4.3) dans l'intervalle $[0..1]$.

Le classement manuel des situations est basé sur un *feedback* des participants qui nous a guidé d'un côté à identifier les situations qu'ils estiment similaires et d'un autre côté à fixer les valeurs des poids α_j dans la formule 3.5. Selon le retour des participants, la localisation a plus de poids dans la détermination du contexte de la recherche que les facteurs de temps. La valeur du poids de la dimension de localisation est fixée à 0.5, celle du moment de la journée à 0.3 et celle du jour de la semaine à 0.2.

Pour calculer la pertinence de l'algorithme de similarité, on a défini deux mesures de rappel et de précision comme suit : $P = \frac{CAG}{AG}$; $R = \frac{CAG}{MG}$, où AG est le nombre total de situations construites automatiquement par notre algorithme de similarité, MG est le nombre de situations similaires classées manuellement et CAG est le nombre de situations similaires identifiées correctes par l'algorithme de similarité. La figure 5.2(b) illustre l'effet de la variation du seuil de similarité β dans l'intervalle $[0..1]$ sur les mesures de rappel (R) et de précision (P). Les résultats montrent que la meilleure performance est obtenue à partir

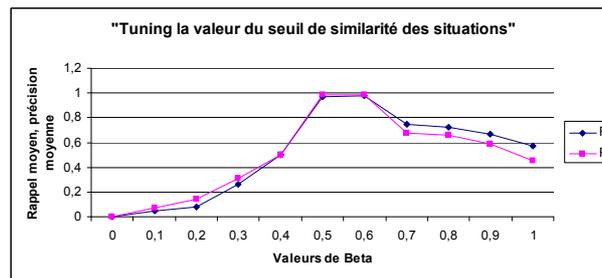


FIG. 3.10 – Effet du paramètre Beta sur la pertinence de la similarité entre situation.

d'une valeur $\beta = 0.6$ avec une valeur de rappel de 0.97 et de précision de 0.98, ce qui montre l'efficacité de notre algorithme de mise en correspondance des situations.

Dans la suite, nous utilisons la valeur de seuil optimal $\beta = 0.6$ pour la similarité des situations lors de l'évaluation de l'efficacité de notre approche de personnalisation de recherche présentée ci-dessous.

7.2.6.2 Évaluation de l'efficacité de la recherche personnalisée

L'objectif de nos expérimentations ici est d'évaluer l'efficacité de la recherche personnalisée qui intègre les centres d'intérêt de l'utilisateur selon les deux stratégies d'évaluation identifiées dans la section 7.2.3 et de vérifier la cohérence des résultats à travers les deux protocoles d'évaluation proposés.

Dans cette évaluation, nous étudions dans un premier temps, l'effet de la variation du paramètre de combinaison du score original du document, ici représenté par une fonction des rangs retournée par Yahoo BOSS, et le score personnalisé du document obtenu selon notre approche, sur la pertinence des résultats de recherche. Les figures 5.2.2, 5.2.2 et 3.12(a), 3.12(b), montrent les améliorations apportées par notre approche de personnalisation respectivement en termes de précision et de gain cumulatif, calculés à différents points (5, 10 et 20) en variant la valeur du paramètre de combinaison γ de l'équation 3.11 dans l'intervalle $[0..1]$, obtenus selon chacune des deux stratégies de validation.

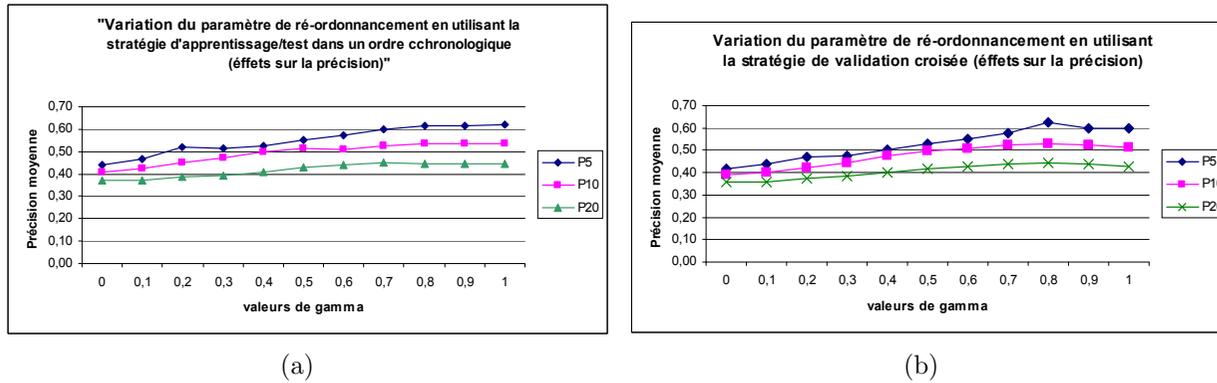


FIG. 3.11 – Effet du paramètre gamma dans la combinaison des scores sur la précision en utilisant. (a) la stratégie d'apprentissage/test en ordre chronologique, (b) la validation croisée.

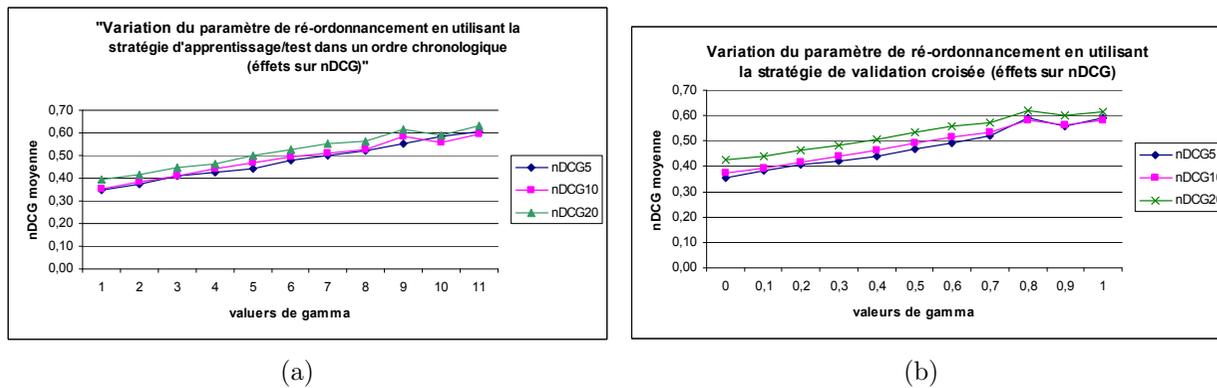


FIG. 3.12 – Effet du paramètre gamma dans la combinaison des scores sur la nDCG en utilisant. (a) la stratégie d'apprentissage/test en ordre chronologique, (b) la validation croisée.

Les résultats obtenus montrent que les meilleures performances sont obtenues avec une valeur de γ comprise entre 0.8 et 1 pour les deux stratégies de validation. Cela est probablement dû au fait que les 50 premiers documents retournés par Yahoo BOSS correspondent bien au sujet de la requête et que la seule caractéristique distinctive est la façon dont ils correspondent aux centres d'intérêt de l'utilisateur.

Dans un second temps, nous avons comparé l'efficacité de notre modèle de recherche personnalisée par rapport à la *baseline* en utilisant la meilleure valeur γ pour chaque stratégie de validation. Les tableaux 3.6 et 3.7 montrent les résultats obtenus par notre modèle de recherche personnalisé et par la *baseline* respectivement en termes des moyennes des précisions et nDCG obtenues à différents points (5, 10 et 20) pour l'ensemble des requêtes de test selon chacune des stratégies de validation. Afin d'évaluer si les différences observées sont significatives d'un point de vue statistique, nous avons également mené un *t*-test sur les moyennes des précisions et nDCG obtenues à différents points (5, 10 et 20) entre la *baseline* et notre approche de personnalisation. Nous supposons encore que la différence entre les moyennes est significative si $p < 0.05$ (noté * dans les tableaux 3.6 et 3.7). Les résultats obtenus montrent qu'en général

Stratégie de validation	system/ amélioration	Précision moyenne		
		P5	P10	P20
apprentissage/ test ordre chronologique	baseline	0.44	0.41	0.37
	notre approche	0.62	0.53	0.45
	amélioration	40.91%*	31.14%*	20.72%*
k-fold cross validation	baseline	0.42	0.39	0.36
	notre approche	0.60	0.52	0.43
	amélioration	43.03%*	32.14%*	19.58%*

TAB. 3.6 – Comparaison des moyennes des Précision@X entre notre approche personnalisée et la *baseline* en utilisant les deux stratégies d'évaluation.

Stratégie de validation	system/ amélioration	nDCG moyenne		
		nDCG5	nDCG10	nDCG20
apprentissage/ test ordre chronologique	baseline	0.35	0.35	0.40
	notre approche	0.60	0.59	0.63
	amélioration	74.06%*	67.65%*	58.80%*
k-fold cross validation	baseline	0.35	0.37	0.42
	notre approche	0.59	0.58	0.61
	amélioration	66.65%*	55.84%*	44.48%*

TAB. 3.7 – Comparaison des moyennes des nDCG@X entre notre approche personnalisée et la *baseline* en utilisant les deux stratégies d'évaluation.

des améliorations significatives sont apportées par notre approche de personnalisation en termes de précision moyenne à différents niveaux de coupures. De

plus, notre approche améliore la $nDCG$ obtenue par la recherche standard et améliore ainsi la qualité des documents en tête de la liste des résultats de la recherche.

Au niveau local de chaque requête, nous avons remarqué de légères dégradations de la pertinence sur quelques requêtes (eg. "entrée pitié-salpêtrière hôpital", "distributeur automatique de billets à ramonville") qui sont très précises. Cela rejoint les résultats mentionnés dans [?] concernant la limite de l'efficacité de la personnalisation sur des requêtes très précises.

La comparaison des résultats selon les deux stratégies d'évaluation laisse observer qu'il y a une certaine différence dans les améliorations de notre approche de personnalisation selon les deux stratégies de validation. Pour déterminer si une stratégie d'évaluation pourrait être meilleure qu'une autre, nous avons effectué un t -test. Plus précisément, nous avons supposé l'hypothèse nulle (notée H_0) stipulant que les deux protocoles d'évaluation atteignent des niveaux de performance similaires, ici évaluée entre les moyennes obtenues sur P_5 , P_{10} , P_{20} , $nDCG_5$, $nDCG_{10}$ et $nDCG_{20}$ sur les requêtes communes au deux protocoles. Cette hypothèse serait rejetée au seuil de signification fixé à 5%. Nous avons obtenu une valeur p de 0.5 pour P_5 , 0.434 pour P_{10} , 0.478 pour P_{20} , 0.463 pour $nDCG_5$, 0.387 pour la $nDCG_{10}$ et 0.365 pour la $nDCG_{20}$, qui sont toutes supérieures à 0.05. Nous pouvons alors accepter l'hypothèse nulle et conclure qu'il n'y a pas de différence significative entre les deux stratégies de validation.

8 Conclusion

Nous avons présenté au cours de ce chapitre notre contribution portant sur la conception d'une approche d'accès contextuel à l'information guidée par un profil situationnel pour un utilisateur mobile. Cette approche est basée sur la modélisation, l'évolution et l'exploitation des profils situationnels de l'utilisateur mobile dans le processus de RI. Ce modèle présente les avantages suivants :

1. L'avantage du modèle de représentation sémantique de la situation de recherche selon des concepts spatio-temporels est liée d'un côté à sa capacité à regrouper des données géographiques et temporelles de bas niveaux dans des représentations plus significatives pour l'utilisateur. Et d'un autre côté, le modèle de la situation est facilement extensible pour inclure d'autres dimensions de l'environnement de l'utilisateur.
2. L'exploitation de la technique de raisonnement à partir de cas a le double avantage de définir un mécanisme de liaison des situations et leur centres d'intérêts tout en permettant une flexibilité pour la modélisation de cha-

cun d'entre eux, puis d'offrir un mécanisme pour la sélection du profil le plus adéquat à une nouvelle tâche de recherche.

3. Finalement, la personnalisation du processus de recherche consiste à exploiter le profil utilisateur dans le ré-ordonnancement des résultats de recherche des requêtes liées à la même situation spatio-temporelle de l'utilisateur mobile.

Nous avons ensuite présenté deux cadres d'évaluation orientés-contexte pour l'évaluation de notre approche de personnalisation.

Le premier cadre est basé sur simulation du contexte et de l'activité de recherche. Ce cadre d'évaluation a permis de faire une première évaluation rapide de l'efficacité de la recherche personnalisée en comparaison à une recherche standard. Les résultats des expérimentations ont démontré l'efficacité de notre approche de personnalisation pour des requêtes courtes et ambiguës pouvant être exprimées dans divers contextes simulés. Le gain de performance de la recherche personnalisée étant variable entre les requêtes, cette variabilité dépend principalement du degré de spécificité des requêtes mais aussi du contenu des documents de la collection de test.

Le second cadre d'évaluation était construit sur la base des entrées d'une étude journalière qui nous a fourni de vrais contextes et des requêtes émises par de vrais utilisateurs mobiles. Ce nouveau cadre nous a permis de tester et de comparer deux protocoles d'évaluation : par ordre chronologique des requêtes et selon la méthode de la cross-validation.

Les résultats expérimentaux montrent bien l'efficacité des différentes composantes de notre approche, notamment :

1. une cohérence des résultats obtenus selon chacun des protocoles d'évaluation testés.
2. une précision du mécanisme de sélection de la situation de recherche la plus adéquate.
3. un gain de performance significatif de la recherche personnalisée comparativement à la recherche classique sur l'ensemble des requêtes de test selon chacun des protocoles d'évaluation.

En l'état, l'expérimentation que nous avons proposée présente trois principales limites. Premièrement, la courte durée de l'étude journalière nous nous a pas permis de tester la dimension saison et le nombre de situations de recherche obtenues par utilisateur reste faible (une moyenne de 5 situations). Deuxièmement, de par ces 3 950 pages web, la collection de test est peu volumineuse en comparaison des collections TREC. Cependant, il suffit de consacrer un peu plus d'efforts pour récolter plus de requêtes et de jugements manuels des documents permettant d'obtenir davantage de contextes à analyser, pour passer l'échelle et construire un cadre plus puissant. Troisièmement, nous n'avons pas comparé notre approche de personnalisation à d'autres approches de personnalisation proposées dans le domaine pour conclure de l'efficacité de notre

modèle situationnel de l'utilisateur comparativement aux autres approches proposés dans ces travaux.

Malgré ces limites, nous estimons qu'en général le cadre d'évaluation basé étude journalière est applicable aux diverses propositions issues du domaine de la RI contextuelle.

Chapitre 4

Identification de la sensibilité de la requête à la localisation

1 Introduction

Le chapitre 1 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. Les défis majeurs dans ce cadre, consistent à identifier les caractéristiques des requêtes induisant chacune des tâches de recherche à prédire, et ensuite à identifier 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 du besoin global ou local derrière la requête (selon que l'utilisateur souhaite des résultats de recherche généraux ou spécifiques/proches de sa localisation), en exploitant des caractéristiques de la requête apprises à partir des résultats de la recherche qu'elle retourne. Nous proposons ensuite d'adapter les résultats de la recherche selon le type prédit de la requête. Nous évaluons notre approche selon deux protocoles d'évaluation adapté d'un côté à la classification du type de besoin des requêtes et d'un autre côté à son intégration dans le processus d'adaptation de la recherche.

Ce chapitre est organisé comme suit. La section 2 présente les problématiques et les motivations de notre contribution. La section 3 présente l'architecture générale de notre approche d'accès contextuel guidé par le type de la requête. La section 4 décrit notre approche de détection de la sensibilité de la requête à la localisation. La section 5 présente notre proposition pour un accès contextuel à l'information en prenant en compte la sensibilité de la requête à la localisation. La section 6 présente l'évaluation de notre approche. La dernière section conclut le chapitre.

2 Problématique et motivations

Le but fondamental des techniques d'identification du type du besoin derrière la requête [36, 200, 68, 258, 247] 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. En particulier, un besoin en information d'un utilisateur mobile (ou son intention de recherche) peut être :

- **une intention globale** : si l'utilisateur cherche des pages globales de contenu général, exemple avec des requêtes telles que "*ringtones*", "*grey's anatomy*", "*mothers day music*",
- **une intention locale** : si l'utilisateur désire plutôt des pages Web centrées autour de sa localisation. Selon que l'utilisateur exprime sa localisation dans sa requête ou pas, on peut distinguer alors entre :
 - **une intention locale explicite** : exemple avec des requêtes telles que "*movie london*", "*indianapolis zoo*", "*nj lottery*",
 - **une intention locale implicite** : exemple avec des requêtes telles que "*airport shuttle*", "*train ticket*", "*jobs*",

Selon l'analyse faite dans [53], plus de 30% des besoins informationnels des utilisateurs mobiles sont liés à leur localisation. Cependant, d'après l'étude de [258] uniquement 50% des requêtes sensibles à la localisation émises par les utilisateurs sur le Web, contiennent des localisations explicites. On s'attend à ce que ce pourcentage soit plus important pour des utilisateurs mobiles. En effet, vu la difficulté de la saisie des requêtes sur ces petits appareils, l'utilisateur s'attend à ce que le système détecte lui-même la nature de sa requête et qu'elle soit automatiquement complétée par le contexte de sa localisation.

Détecter la sensibilité à la localisation d'une requête émise par un utilisateur mobile constitue un défi important à la communauté en RI. On peut distinguer trois problèmes principaux :

1. Comment distinguer les requêtes globales des requêtes sensibles à la localisation ? Cette question est critique pour déterminer s'il convient de retourner des résultats de recherche génériques ou plutôt locaux.
2. Comment différencier les requêtes locales explicites des requêtes locales implicites ? Cette spécification supplémentaire est importante car elle permet de décider s'il convient d'ajouter la localisation de l'utilisateur dans le cas d'une requête locale implicite ou pas dans le cas d'une requête locale explicite.
3. Comment adapter les résultats de la recherche selon le type de la requête ?

S'il est relativement facile de détecter les requêtes locales explicites, en comparant les mots de la requête aux entrées d'une base de données géographiques, il reste cependant plus difficile de détecter les requêtes locales implicites. En effet, les moteurs de recherche locale peuvent traiter certaines catégories prédéfinies de requêtes sensibles à la localisation (eg. restaurants, hôtels, parcs, etc), mais pas toutes les requêtes.

En guise d'exemple, la figure 4.1 et la figure 4.2 présentent un aperçu des résultats de la recherche retournés respectivement par Google et Google Local pour la requête locale implicite "looking to hire a dj".

Comme l'on peut constater sur ces deux figures aucune des fonctionnalités

- [Re: Looking to hire a DJ](#)
 28 Mar 2011 ... \$250 for a wedding DJ is atrociously low. The cheapest equipment money can buy is still thousands of dollars to properly do a wedding. ...
[washingtondc.craigslist.org](#) > [northern virginia](#) > [gigs](#) > [event gigs](#)
- [DJS-FOR-HIRE.com | Looking for a Dj? Worldwide Dj agency](#)
 We are a dedicated online directory of Dj profiles available to the public who can contact Djs directly to hire for any event with minimal Fuss!
[djs-for-hire.com/-](#) - [Cached](#) - [Similar](#)
- [Looking to hire a DJ | Calgary | Yelp](#)
 Does anyone recommend a good DJ in Calgary that plays good dance music for a company Christmas party? They should have music from 80's up to today's dance ...
[www.yelp.ca/topic/calgary-looking-to-hire-a-dj](#) - [Cached](#)
- [How to hire a DJ - How to hire a wedding DJ](#)
 Check at least 5-10 references on the DJ or service you are looking to hire (either names and numbers or written comment letters). Remember that DJ's can ...
[www.alltimefavorites.com/tips/dj.htm](#) - [Cached](#) - [Similar](#)
- [Looking to Hire a DJ in Essex? | upentertainments Mobile DJ Blog](#)
 Looking to Hire a DJ in Essex? By upentertainments | In: News | 8 Comments. 17 January 2011. You are here: Home / News / Looking to Hire a DJ in Essex? ...
[www.upentertainments.co.uk/blog/news/looking-to-hire-a-dj-in-essex](#)
- [Hire a DJ for your party in Durban and areas | Facebook](#)
 This group is for anybody looking to hire a DJ for their private functions. DJs may join this group and use us as your Agent. Honesty regarding commission ...
[www.facebook.com/group.php?gid=6928918534](#) - [Cached](#) - [Similar](#)
- [Sydney DJ Hire](#)
 The number one website you need when looking for Sydney DJ Hire online! ... If you are looking to Hire a DJ in Sydney, then you have come to the right place ...
[www.sydneydjhire.com/-](#) - [Cached](#)
- [Hire DJs, DJ Jobs | Hire Wedding and Club DJs | IWantA-DJ](#)
 Using our DJ FOR HIRE service, perhaps your looking for an experienced and professional WEDDING DJ, CLUB DJ or MOBILE DISCO DJ, we have them to suit all ...
[www.iwanta-dj.com/-](#) - [Cached](#) - [Similar](#)
- [Hire A DJ, Find A DJ](#)
 If you're looking to hire a DJ you've come to the right place! DJ Craig.Net is disc jockey music, sound and lighting for your special event. ...
[www.djcraig.net/events.html](#) - [Cached](#)
- [Looking to hire a DJ for my... - dj forums - The Worldwide DJ...](#)
 17 posts - 11 authors - Last post: 3 Apr 2008
 Location: [Philadelphia](#). Posts: 2. Track: anything by The Cure. LKK is an unknown quantity at this point.
 Looking to hire a Philly DJ for my. ...
[www.djforums.com](#) > ... > [General](#) > [Mobile DJ](#) - [Cached](#)

FIG. 4.1 – Un aperçu des résultats de Google retournés pour la requête locale implicite "looking to hire a dj".

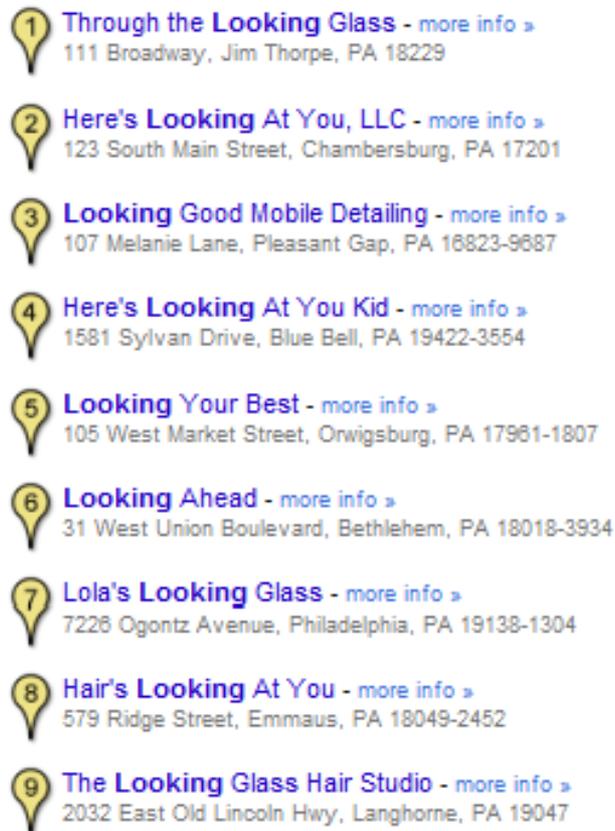


FIG. 4.2 – Un aperçu des résultats de Google Local retournés pour la requête locale implicite “*looking to hire a dj*”.

de Google ou Google Local n’ont permis de retourner des résultats appropriés à la localisation de l’utilisateur. Dans Google les résultats sont diversifiés sur plusieurs localisations et dans Google Local les résultats sont tout simplement erronés et ne correspondent même pas à la thématique de la requête.

Pour toutes ces raisons, nous nous sommes donc intéressés à la prédiction de l’intention de l’utilisateur (globale ou locale) derrière sa requête sur le *Web*. Notre approche de prédiction du type du besoin consiste à exploiter les top “K” documents retournés par un moteur de recherche standard sur le Web comme source d’évidence pour identifier le type de la requête [8]. Notre intuition, est qu’une requête qui retourne des documents contenant peu de références vers des lieux géographiques est susceptible d’être globale, alors qu’une requête qui renvoie des documents répartis uniformément sur de nombreux lieux géographiques sans inclure un pourcentage important de documents ne contenant aucune localisation, est susceptible d’être sensible à la localisation.

La plupart des approches dans le domaine [89, 258, 247] définissent uniquement un ensemble de caractéristiques pour classer les requêtes, aucune approche ne propose de modéliser la sensibilité de la requête à la localisation. De plus,

la classification des requêtes est faite uniquement en deux classes : globale ou locale implicite. Cela suppose qu'il y a un premier mécanisme pour détecter si une requête est locale explicite, dans le cas contraire un second mécanisme distinguera si la requête est globale ou locale.

Contrairement à ces approches, nous proposons une nouvelle approche pour identifier automatiquement la sensibilité de la requête à la localisation de l'utilisateur. Comparativement à ces approches, notre approche se distingue par :

1. nous proposons de formaliser la notion de sensibilité de la requête à la localisation de l'utilisateur en construisant un profil local de la requête en utilisant un modèle de langue,
2. nous proposons d'identifier en un seul mécanisme les trois classes possibles des requêtes : globale, locale-implicite et locale-explicite,
3. nous nous basons dans notre classification sur deux mesures calculées sur le profil de la requête, l'une est issue du domaine de la théorie de l'information, notamment la mesure "Kullback-Leibler Divergence", et la seconde de la théorie des probabilités en l'occurrence la mesure de Kurtosis, qui sont utilisées comme nos attributs de classification.

3 Architecture de notre approche d'accès contextuel guidé par le type de la requête

L'architecture générale de notre approche pour un accès guidé par le type de la requête est décrite dans la figure 4.3. Cette architecture est basée sur le modèle client serveur :

- du côté serveur, nous avons quatre composantes : la composante de recherche (moteur Web), la composante d'indexation géographique, la composante de classification conçue pour classer les requêtes en fonction de leur sensibilité à la localisation, la composante d'adaptation de la recherche selon le type de la requête,
- du côté du client, nous avons deux composantes : l'interface de recherche et le service de localisation de l'utilisateur.

Les deux principales composantes de cette architecture sont celle de classification qui supporte notre méthode de détection du type de la requête, et celle d'adaptation qui supporte notre méthode d'exploitation du type de la requête dans le processus de recherche. Ces deux processus sont illustrés dans la figure 4.4.

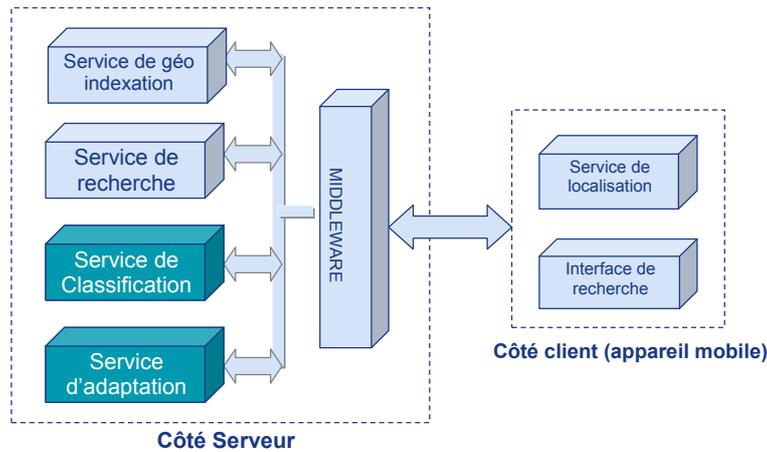


FIG. 4.3 – Architecture générale du système s'accès contextuel intégrant notre approche de classification.

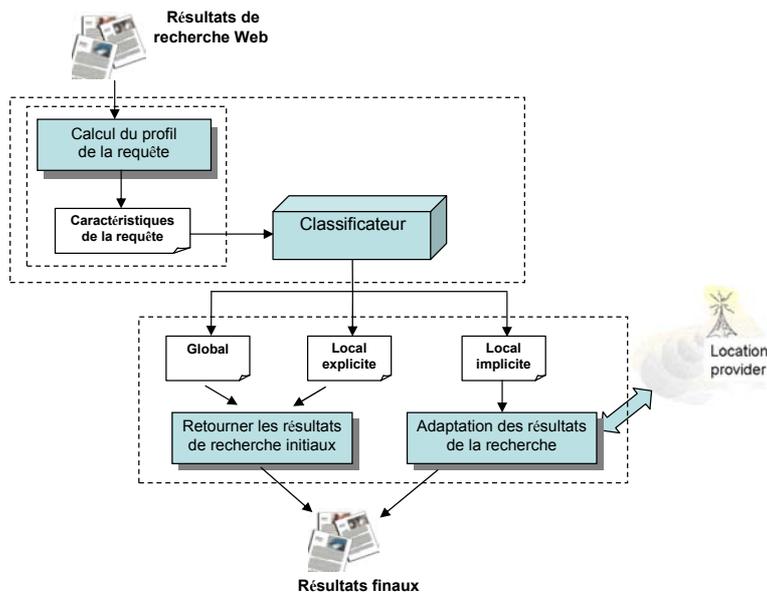


FIG. 4.4 – Schéma général de notre modèle de classification des requêtes.

1. **Le processus de classification** : le processus de classification se base sur un profil de la requête représenté par un modèle de langue de localisation de la requête conçu à partir des résultats de recherche qu'elle retourne. Plus exactement, le classificateur exploite les deux caractéristiques des requêtes calculées sur ce profil, notamment les deux mesures de Kurtosis et Kullback-Leibler Divergence, calculées sur le modèle de langue de la requête.
2. **Le processus d'adaptation** : selon les résultats de la classification, les résultats initiaux de la recherche peuvent être modifiés pour prendre en compte la localisation de l'utilisateur. L'adaptation des résultats de

recherche s'effectue essentiellement dans le cas de requêtes de type local implicite. Nous avons proposé d'étudier les deux méthodes par reformulation de requête et par ré-ordonnancement des résultats de la recherche.

4 Prédiction du type du besoin derrière la requête

Notre approche de prédiction du type de besoin derrière la requête est basée sur un modèle de langue de localisation de la requête, également appelé profil local de la requête. Il est appris sur la base des résultats de la recherche retournés par un moteur de recherche général sur le Web. Deux attributs caractérisant la requête sont calculés ensuite sur ce profil, notamment la divergence de Kullback-Leibler et Kurtosis, dans le but de classer les requêtes dans les trois catégories : globale, locale implicite et locale explicite.

Nous présentons dans cette section :

1. la modélisation du profil local de la requête,
2. les attributs de classification basés sur le profil local de la requête.

4.1 Modélisation du profil local de la requête

Les requêtes Web sont généralement courtes, ce qui rend difficile la tâche de les analyser en se basant sur les termes qui les composent. Un bon moyen d'analyser une requête serait alors d'avoir un aperçu des documents qu'elle retourne. Cela est possible par exemple en examinant les K premiers documents dans la liste initiale des résultats de recherche, puis en effectuant des mesures sur les propriétés statistiques des termes dans cet ensemble de documents [64], [142].

Dans un contexte de modèle de langue en RI, les documents sont généralement ordonnés dans la liste des résultats de la recherche, selon leur probabilité d'avoir généré la requête. Plus formellement, étant donné une requête Q et un document D , le score du document D face à la requête Q est déterminé par la probabilité que la requête soit générée à partir du modèle de langue du document :

$$P(Q|D) = \prod_{w \in Q} P(w|D)^{q_w} \quad (4.1)$$

Où q_w est la fréquence du terme w dans la requête Q . Le modèle de langue du document, $P(w|D)$, est généralement estimé par la méthode du maximum de vraisemblance, sur la base des fréquences des termes dans le document D [63].

En utilisant le modèle de langue pour ordonner les documents retournés à une requête, [64] construit un modèle de langue de la requête, $P(w|Q)$, à partir des K premiers documents dans la liste des résultats de recherche, comme suit :

$$P(w|Q) = \sum_{D \in R} P(w|D) \frac{P(Q|D)}{\sum_{D' \in R} P(Q|D')} \quad (4.2)$$

Où R est la liste des top K documents considérés. Nous notons ce modèle de langue de la requête, calculé sur la base des termes de la requête, le profil de contenu “*the content-profile*” de la requête Q .

Notre objectif étant de déterminer l’intention locale ou pas d’un utilisateur derrière sa requête, nous nous sommes intéressés à identifier la nature de la sensibilité de la requête à la localisation. Ainsi, nous voulons examiner un “profil local” de la requête, par analogie au “profil de contenu “ décrit ci-dessous. Pour cela, nous définissons notre modèle de langue de localisation de la requête comme suit :

$$\hat{P}(l|Q) = \sum_{D \in R} \hat{P}(l|D) \frac{P(Q|D)}{\sum_{D' \in R} P(Q|D')} \quad (4.3)$$

Où l est une localisation (à partir d’une base de données géographique) et

$$\hat{P}(l|D) = \begin{cases} 1 & \text{si } l \in L_D \\ 0 & \text{sinon} \end{cases} \quad (4.4)$$

Où L_D est l’ensemble des localisations contenues dans le document D .

Il est souvent nécessaire de lisser les modèles de maximum de vraisemblance, comme $\hat{P}(l|Q)$, pour éviter les probabilités nulles. Nous proposons de lisser $\hat{P}(l|Q)$ en utilisant la technique de lissage de Jelinek-Mercer [113] par interpolation avec un modèle de référence (“*background model*”). Nous avons donc proposé d’utiliser la distribution des localisations à travers les documents de la collection C comme notre modèle de référence. Nous définissons notre modèle de localisation de la collection comme suit :

$$\hat{P}(l|C) = \frac{1}{|C|} \sum_{D \in C} \hat{P}(l|D) \quad (4.5)$$

Notre estimation peut alors être interpolée linéairement avec ce modèle de référence telle que :

$$P(l|Q) = \lambda \hat{P}(l|Q) + (1 - \lambda) \hat{P}(l|C) \quad (4.6)$$

Où λ est un le paramètre de lissage, il est expérimentalement fixé à 0.9 le long de notre travail.

La figure 4.6 montre un exemple de profils de requêtes de différents types.

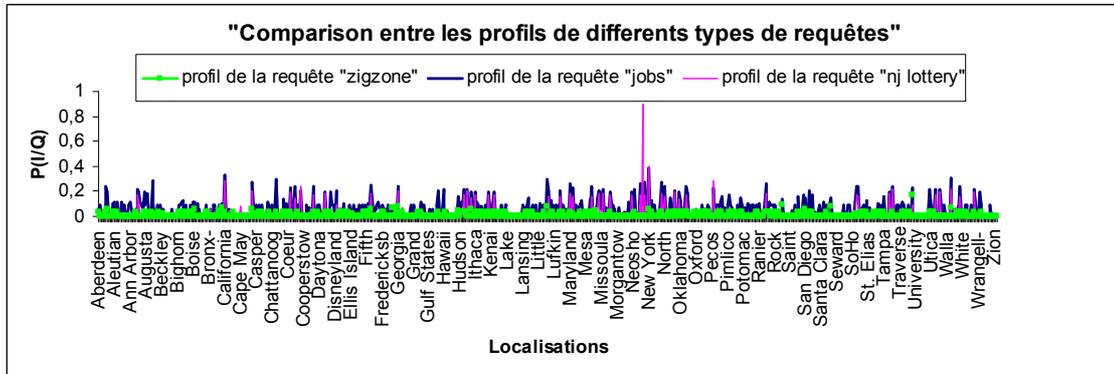


FIG. 4.6 – Comparaison entre le profil local d’une requête globale “*zigzone*”, une requête locale implicite “*jobs*” et une requête locale explicite “*nj lottery*”.

Comme on peut le constater sur la figure 4.6, la requête globale “*zigzone*” présente un profil quelque peu plat, la requête locale implicite “*jobs*” a un profil qui présente plusieurs pics distinctifs répartis sur plusieurs localisations, alors que la requête locale explicite “*nj lottery*” présente un profil avec principalement un pic très distinctif à côté de la localisation “*new jersey*”. Cela ne peut que soutenir notre hypothèse ici.

Pour mener à bien cette analyse et confirmer ces hypothèses, nous nous sommes proposés de calculer deux mesures, que nous estimons appropriées, pour cerner les différences que présenteraient les profils des requêtes. Ces mesures sont : la Divergence de Kullback-Leibler et Kurtosis.

Les données de base nécessaires pour produire ces deux mesures peuvent être efficacement calculées pendant la phase d’indexation, i.e *off-line*. Ainsi, le calcul final de ces caractéristiques pourrait être rapidement effectué au moment du traitement de la requête en utilisant ces données de base.

Ces deux mesures caractérisant les requêtes seront ensuite utilisées comme attributs de classification, comme présenté dans la section 6, pour prédire l’intention de l’utilisateur derrière sa requête.

4.2.1 La mesure Kullback-Leibler Divergence

Pour calculer une première caractéristique des requêtes à partir de leur modèles de langue, nous proposons de mesurer la différence entre la distribution des localisations dans les documents retournés pour une requête, et la distribution des localisations dans l’ensemble des documents de la collection.

Formellement, ceci peut être représenté en mesurant la divergence de Kullback-Leibler “*KL divergence*” [137] entre le modèle de localisation de la collection et le profil local de la requête, comme suit :

$$D_{KL}(P(l|Q), P(l|C)) = \sum_{l \in L} P(l|Q) \log \frac{P(l|Q)}{P(l|C)} \quad (4.7)$$

Nous nous référons à cette caractéristique par “KL divergence de localisation”, et l’on note : *locationKL*. Dans notre cadre, une grande valeur de *locationKL* indique une requête sensible à la localisation. En guise d’exemple, la table 4.1 présente des valeurs de *locationKL* obtenues par des requêtes de différents types.

Requête	Type de la requête	mesure de LocationKL
”zigzone”	global	1.23
”jobs”	local implicite	134.65
”nj lottery”	local explicite	31.01

TAB. 4.1 – Exemple de valeurs de *locationKL* pour différents types de requêtes

Comme on peut le constater, la forme quelque peu hérissée du profil de la requête “*jobs*” est clairement capturée par cette mesure avec une assez grande valeur de *locationKL*. De plus, la requête globale “*zigzone*” présente une valeur de *locationKL* beaucoup plus petite, signalant sa non sensibilité à la localisation. Nous notons aussi, que bien que la valeur de *locationKL* montre l’écart ou la différence de la distribution des localisations entre les documents retournés pour une requête par rapport aux documents de la collection, elle ne permet pas pour autant de bien distinguer entre les requêtes dépendantes d’une seule localisation, i.e. les requêtes locales explicites, des requêtes locales implicites.

4.2.2 La mesure de Kurtosis

Une autre façon de capturer les caractéristiques de la forme de la distribution des localisations dans le profil de la requête, est de regarder son coefficient *kurtosis* ou coefficient d’aplatissement.

La mesure de Kurtosis, correspond à une mesure de l’aplatissement ou a contrario de la pointicité, de la distribution d’une variable aléatoire réelle. En effet, un coefficient d’aplatissement élevé indique que la distribution est plutôt pointue, et à l’opposé, un coefficient d’aplatissement faible indique une distribution relativement aplatie.

Nous organisons les localisations dans un ordre décroissant de $P(l|Q)$, alors les caractéristiques sont les propriétés statistiques de la décroissance de $P(l|Q)$.

Plus spécifiquement, le coefficient kurtosis est défini, selon [13], comme le 4^{ème} moment standardisé d'une distribution de probabilité.

Formellement, étant donné une variable aléatoire réelle X d'espérance μ et d'écart-type σ , le kurtosis est calculé comme suit :

$$Kurtosis = \frac{\mu_4}{\sigma^4} \quad (4.8)$$

Où $\mu_i = E[(X - E[X])^i]$ est le moment central d'ordre i .

Par application à notre cas ici, nous nous attendons à ce que des valeurs supérieures de Kurtosis, signifiant que la majorité de la variance est le résultat de rares écarts extrêmes, correspondraient à des requêtes locales explicites. Et que à l'opposé, de faibles valeurs de Kurtosis, signifiant des écarts fréquents de taille modeste, correspondraient à des requêtes locales implicites. La table 4.2 présente à titre d'exemple quelques valeurs de kurtosis calculées pour les profils de différents types de requêtes.

Requête	Type de la requête	mesure de Kurtosis
"zigzone"	global	41.39
"jobs"	local implicite	2.05
"nj lottery"	local explicite	119.65

TAB. 4.2 – Exemple de valeurs de *kurtosis* pour différents types de requêtes

On peut observer que la valeur de kurtosis pour la requête "*nj lottery*" exprime clairement la nature pointue de sa courbe et permet de bien la différencier de la requête locale implicite "*jobs*", qui présente en revanche une très faible valeur de kurtosis.

5 Accès contextuel à l'information guidé par la sensibilité de la requête à la localisation

Notre méthode proposée pour identifier la sensibilité de la requête à la localisation est exploitée comme une source d'évidence pour adapter les résultats de la recherche selon le type de la requête. En particulier, l'adaptation des résultats de la recherche s'effectuera pour les requêtes locales implicites en considérant la localisation de l'utilisateur comme facteur supplémentaire pour déterminer la pertinence des résultats à retourner à l'utilisateur.

L'adaptation du processus de RI nécessiterait alors d'un côté, une méthode pour la détermination de la localisation de l'utilisateur, et d'un autre côté, une méthode pour la prise en compte de la pertinence des documents vis-à-vis de cette localisation. La détermination de la localisation de l'utilisateur mobile peut être effectuée en utilisant une des techniques de localisation GSM basée

sur le réseaux cellulaire, sur la technologie mobile elle même (incluant des modules GPS), ou sur une combinaison de ces deux dernières. Pour l'exploitation de la localisation dans le processus de sélection des documents, nous avons proposé de tester les deux approches d'adaptation par re-formulation de la requête et par ré-ordonnancement des résultats initiaux du SRI. Elles sont décrites dans la suite de cette section.

5.1 Adaptation par re-formulation de la requête

Pour adapter les résultats de la recherche mobile étant donnée une requête de type local implicite, une façon intuitive serait de ré-écrire la requête initiale Q en ajoutant la localisation de l'utilisateur l_u .

Plus spécifiquement, soit $Q = \{t_1, t_2, \dots, t_n\}$ les termes de la requête initiale, une nouvelle requête de recherche peut être alors formulée comme $Q_l = \{t_1, t_2, \dots, t_n, l_u\}$. Par exemple, si la requête initiale est “*jobs*” et la localisation de l'utilisateur a été déterminée comme “*Jefferson City*”, la nouvelle requête sera reformulée par “*jobs Jefferson City*” par simple ajout de la localisation à la requête initiale.

La nouvelle requête sera alors soumise automatiquement au moteur de recherche Web pour obtenir de nouveaux résultats de recherche qui a priori correspondraient mieux à la localisation de l'utilisateur. En effet, étant donné que la localisation “*Jefferson City*” est maintenant explicitement incluse ou exprimée dans la requête, le moteur de recherche est plus susceptible de retourner des offres de *jobs* centrées autour de “*Jefferson City*”.

Cette approche d'adaptation est la plus simple. Cependant elle peut induire quelques erreurs dues à la limite de la précision du classificateur. En effet, l'intention de l'utilisateur pour chercher des résultats locaux pourrait ne pas être sa seule intention, ou même peut ne pas exister dans le cas de requêtes mal classées.

5.2 Adaptation par ré-ordonnancement des résultats de la recherche

Une autre approche d'adaptation consiste à utiliser la technique de ré-ordonnancement pour ajuster les effets de la localisation sur le classement final des documents dans la liste des résultats de recherche.

Cela consiste à reprendre les “ K ” premiers documents retournés par la requête initiale Q et d'exercer comme un effet de levier sur ces résultats de recherche. D'un côté, en remontant en tête de la liste les documents qui correspondent

bien à la localisation de l'utilisateur l_u , et de l'autre côté, en poussant plus bas les documents qui n'y correspondent pas.

Plus formellement, les résultats initiaux de la recherche sont ré-ordonnés en combinant pour chaque document retourné d_k , son score original $scoreo(d_k, Q)$, exprimé par une fonction de son rang dans la liste retournée par le moteur Web, et un score de localisation $scorel(d_k, l_u)$ pour obtenir un score final $scoref(d_k)$ défini comme suit :

$$scoref(d_k) = (1 - \alpha) * scoreo(d_k, Q) + \alpha * scorel(d_k, l_u) \quad (4.9)$$

Où le paramètre de combinaison α varie dans l'intervalle $[0, 1]$, $k \in [1, K]$.

Le score original $scoreo(d_k, Q)$ est exprimé par une fonction de son rang dans la liste retournée par le moteur Web. Le score de localisation $scorel(d_k, l_u)$ est donné par une fonction de correspondance textuelle qui permet de vérifier si la localisation de l'utilisateur l_u et/ou ses variantes existent dans le document d_k . Il est calculé comme suit :

$$scorel(d_k, l_u) = \frac{tf(l_u)}{\sum_{l \in L_{d_k}} tf(l)} \quad (4.10)$$

Où L_{d_k} est l'ensemble de toutes les localisations contenues dans le document d_k , $tf(l)$ est la fréquence de la localisation l et/ou ses variantes dans le document d_k .

6 Évaluation expérimentale

L'objectif de notre démarche expérimentale est double : d'abord évaluer l'efficacité de notre approche basée modèle de langue pour l'identification du type de la requête (locale explicite, locale implicite ou globale), puis évaluer l'efficacité de l'accès contextuel à l'information intégrant le mécanisme de détection du type de la requête.

6.1 Description des ressources expérimentales

Nous décrivons dans cette section, notre collection de requêtes, les outils d'extraction des caractéristiques des requêtes ainsi que l'outil de classification *Weka* que nous avons utilisé pour tester différents classificateurs de l'état de l'art.

6.1.1 Collection de requêtes

Dans le but d'évaluer la classification des requêtes selon le type global, local explicite ou local implicite, nous avons sélectionné, de façon aléatoire un ensemble initial de 1000 requêtes, issues du *log* de requête du moteur américain de recherche web AOL [177].

Après une étape de filtrage pour éliminer les requêtes adultes, en double ou de type navigationnel, nous avons retenu un ensemble de 200 requêtes (elles sont données dans l'appendice A). Nous avons restreint le nombre de requêtes de test du au double effort nécessaire pour l'annotation manuelle du type des requêtes et la construction des jugements manuels de pertinence des documents.

Nous avons soumis cet ensemble de requêtes à une annotation manuelle par deux juges, pour attribuer à chaque requête une étiquette indiquant si elle est locale implicite, locale explicite ou globale. Le critère pour assigner le type de la requête est basé sur l'attente de l'utilisateur de voir des résultats de recherche locaux en haut de la liste retournée par le moteur de recherche ou plutôt des pages web générales.

En fin de cette opération d'annotation manuelle, la composition du type des requêtes dans notre échantillon de test est de 20,5% de locales explicites, 25,5% locales implicites et 54% globales.

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

Nous spécifions ici les outils utilisés pour pouvoir construire les caractéristiques de nos requêtes.

1. **Google AJAX Search API**² : pour obtenir les 50 premières pages Web qui correspondent à chaque requête, nous avons utilisé le moteur de recherche Web standard Google via son AJAX API de recherche. Cette version de Google est une interface de recherche générale (par opposition à l'API de recherche locale de Google), elle n'exploite pas les adresses IP des requêtes, ni les catégories de recherche Google lors de la réponse à une requête. Pour illustrer notre propos, la figure 4.7 montre un *snapshot* des premiers résultats de la recherche retournés par cette API pour la requête "airport shuttle", qui est a priori une requête sensible à la localisation d'un utilisateur mobile. Nous pouvons constater que les résultats retournés ne sont pas biaisés par les fonctionnalités de Google local.
2. **Collection de documents** : pour pouvoir calculer le profil de la collection, nous avons augmenté notre petite collection de documents formée par l'ensemble des 50 premiers documents résultats de nos requêtes de

²<http://code.google.com/apis/ajaxsearch/web.html>



FIG. 4.7 – Un snapshot des résultats de la recherche retournés par l’API Google pour la requête “airport shuttle” : il illustre la présence et la diversité des localisations (surlignées en rouge).

test par une partie (à peu près 1/5) de la collection Web de trec WT10g³, pour avoir au total une collection de près de 300000 documents.

3. **Base de données géographiques** : pour pouvoir identifier les noms des localisations dans les pages Web de notre collection, nous avons eu recours à une base de données géographiques. Étant donné que les requêtes dans le *log* AOL sont en anglais, nous avons focalisé notre identification de localisations sur les états, les villes et les comtés des États Unis d’Amérique, en utilisant la liste des places géographiques téléchargée via le web à partir du bureau américain de recensement⁴.

Le processus d’extraction des noms géographiques consiste à comparer les mots du texte de chaque document aux localisations dans cette base de données géographiques. Les mots correspondants aux entrées de la base, sont retenus avec leur nombre d’occurrence par document. Nous notons qu’à ce stade nous nous sommes confrontés au problème de l’ambiguïté des noms géographiques, qui est un problème bien connu dans le domaine

³http://ir.dcs.gla.ac.uk/test_collections/wt10g.html

⁴U.S. Census Bureau : <http://www.census.gov/>

de la RI géographique [117, 18].

En effet, un même nom de localisation dans un document, peut être associé à plusieurs entrées de notre base de données géographiques, cela est communément appelé “*geo-geo ambiguity*” (eg. “Virginia” est une ville de l’état de l’Illinois mais aussi de Minnesota). Pour pallier ce problème, nous avons décidé d’associer la même occurrence de localisation dans un document à une seule entrée dans la base de données.

Un autre type d’ambigüité est connu sous l’appellation “*geo-non-geo ambiguity*”, cela correspond à des noms qui peuvent référer à la fois des places géographiques mais au même temps des noms de personnes (eg. “Denzel Washington”) ou d’objets (eg. “Orange”). Pour remédier à ce problème, nous avons eu recours à une liste de mots vides pour éliminer les mots tels que : orange, mobile, etc, et à un outil de reconnaissance d’entités nommées [190] pour différencier les noms de personnes des noms de localisations.

6.1.3 Outil de classification des requêtes par apprentissage supervisé basé sur le logiciel Weka

Weka [92] est une suite de logiciels d’apprentissage automatique, utilisée dans beaucoup de domaines d’application en particulier pour l’éducation et la recherche. Il supporte plusieurs outils d’exploration de données standards, et en particulier, plusieurs classificateurs statistiques y sont implémentés.

Nous avons testé avec plusieurs classificateurs individuels standards : les arbres de décision, Naive Bayes, SVM, classificateurs à base de règles, mais aussi des méta classificateurs utilisant des algorithmes de *boosting* et des classificateurs combinés basés sur des méthodes de *voting*.

Toutes les techniques de Weka reposent sur la supposition que les données d’apprentissage et/ou de test sont disponibles dans un unique fichier qui définit une relation binaire, où chaque type de donnée est décrit par un nombre fixe d’attributs (les attributs ordinaires, numériques ou symboliques, mais d’autres types d’attributs sont aussi supportés).

6.2 Protocole d’évaluation

Cette section présente les stratégies de validation que nous avons mises en place pour l’évaluation de l’efficacité de nos attributs de classification pour l’identification du type des requêtes et celle de l’efficacité de la recherche intégrant notre approche de classification des requêtes.

6.2.1 Protocole d'évaluation de l'efficacité de la technique de classification

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. Cela signifie que le résultat de la classification d'une requête, est dit être correcte seulement s'il correspond à l'annotation manuelle.

Le principe de l'évaluation est basé sur la méthode de la 10-fold cross-validation qui consiste à :

- subdiviser l'ensemble des requêtes Q en 10 sous ensembles de taille $|Q|/10$,
- apprendre sur 9 sous-ensembles et tester avec le sous-ensemble restant,
- répéter ce processus 10 fois en alternant les sous ensembles d'apprentissage et de test,
- prendre la moyenne des performances sur les 10 jeux de test.

Le protocole d'évaluation est basé donc sur les deux phases d'apprentissage et de test. Pour chacun des classificateurs testés, nous spécifions ces deux phases comme suit :

1. La phase d'apprentissage selon cette méthode consiste à sauvegarder des exemples de requêtes dans trois classes représentant les types de la requête : global, local explicite et locale implicite. Chaque exemple d'une requête q correspond à son descripteur contenant les caractéristiques de la requête ($F(q)=(\text{valeur de l'acationKL}, \text{valeur de Kurtosis})$).
2. La phase de test consiste à créer le descripteur 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 requête.

Pour mesurer la performance des classificateurs à prédire le type des requêtes, nous avons utilisé les métriques standards de précision, rappel, F-mesure et d'*accuracy*.

6.2.2 Protocole d'évaluation de l'efficacité de l'accès contextuel guidé par le type de la requête

Pour pouvoir évaluer l'efficacité de l'accès contextuel dans le cas des requêtes locales implicites, et en l'absence d'information sur la localisation des utilisateurs ayant émis les requêtes dans notre *log*, nous avons associé pour chaque requête une localisation simulée extraite à partir du profil de la requête. Nous avons ensuite, appliqué les deux approches par ré-ordonnancement et par reformulation de la requête pour les requêtes classées comme locales implicites par notre classificateur.

Nous avons utilisé les 50 premiers résultats de recherche retournés par Google AJAX API (déjà utilisés pour la classification des requêtes) comme notre *baseline*. Nous avons ensuite utilisé ces mêmes documents pour l'approche de ré-ordonnancement. Puis, pour l'approche par re-formulation, nous avons soumis une seconde fois les requêtes initiales (classées comme locales implicites) augmentées des localisations simulées à Google AJAX API et récupéré les 50 premiers documents résultats.

Nous avons soumis les requêtes, avec leurs localisations et les 10 premiers documents retournés par le moteur Web pour chaque approche à deux juges (via un outil d'assesment en ligne) pour assigner les pertinences de recherche, sans aucune information sur la source des documents. Le jugement de pertinence est fait en tenant compte à la fois de la localisation et de la thématique de la requête pour les requêtes locales implicites. Pour les requêtes mal classées, la localisation est ignorée dans le jugement de pertinence. Les jugements de pertinence sont effectués en utilisant une échelle à trois niveaux de pertinence : très pertinent, pertinent, et non pertinent .

Pour mesurer la performance de la recherche nous avons utilisé les deux métriques du gain cumulatif normalisé - Discounted Cumulative Gain (nDCG) et la précision@X calculées à différents rangs (Top1, Top3, Top5, Top7 et Top10) parmi les premiers rangs des résultats. En effet, dans une recherche localisée on s'attend à avoir des résultats pertinents dès les premiers rangs étant donné que la pertinence de la recherche localisée est plus précise qu'une recherche thématique.

6.3 Résultats et discussion

Nous avons évalué l'efficacité du modèle de combinaison en utilisant les protocoles d'évaluation présentés dans la section 6.2. Plus précisément, nos objectifs d'évaluation à ce niveau consiste à :

- Évaluer en un premier temps l'efficacité de notre méthode d'identification du type de la requête. Ceci est réalisé de deux manières : d'abord une évaluation de la performance en utilisant les classificateurs de l'état de l'art, puis par comparaison à une autre approche d'identification de type de requêtes.
- Évaluer en un second temps l'efficacité du processus de recherche intégrant notre méthode d'identification du type de la requête par rapport à une *baseline*.

6.3.1 Évaluation de l'efficacité de nos attributs de classification

L'évaluation de la prédiction du type de la requête 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 est d'évaluer l'efficacité de nos attributs de classification locationKL et Kurtosis à identifier le type de la requête parmi les classes : globale, locale implicite et locale explicite.

Dans la table 4.3 nous présentons les 5 plus faibles et/ou plus fortes valeurs de kurtosis et de locationKL pour chaque classe de requêtes obtenues sur notre ensemble de requêtes de test.

Nous pouvons constater un contraste claire entre les valeurs de locationKL et de Kurtosis entre les différentes classes de requêtes. Ce qui nous amène à confirmer une corrélation possible entre la classe de la requête et ses valeurs caractéristiques de locationKL et Kurtosis.

Nous avons alors testé nos attributs de classification avec différents types ou familles de classificateurs : arbres de décision - *Decision trees* (J48 une implémentation du classificateur C4.5 [188]), la classification naïve bayésienne - *Naive Bayes classifier* [82], les machines à vecteurs de support - *Support Vector Machines* (SVM) [181], classificateur à base de règles - *rule-based classifier* (Ripper) [57] et un meta classificateur (Classification Via Regression) [80].

La table 4.4 présente les valeurs des métriques d'évaluation précision, rappel, F-mesure et *accuracy* obtenues par chaque classificateur.

Nous pouvons constater que tous les types de classificateurs étaient capables de distinguer entre les différents types de requêtes, atteignant des valeurs de F-mesure comparables allant de 85% à 88%. Le meta classificateur "Classification Via Regression" présente la meilleure *accuracy* avec une valeur de F-mesure de 88%.

Cette première évaluation démontre bien l'efficacité de notre approche de classification basée sur le profil de la requête à identifier les requêtes sensibles à la localisation. En effet, une classification basée sur deux caractéristiques de ce profil, notamment locationKL et Kurtosis, permet de distinguer avec précision les trois classes de requêtes globale, locale explicite et locale implicite avec une *accuracy* atteignant 88%.

Requête	LocationKL	Kurtosis	Classe
“area codes”	268.07	0.62	locale implicite
“yellow pages”	134.65	2.05	
“national parks”	95.25	9.17	
“restaurants”	62.21	7.61	
“looking to hire DJ”	62.05	1.78	
“las vegas special events contacts”	8.81	418.73	locale explicite
“chicago magazine”	16.28	281.51	
“campgrounds in the mountains of n.c”	27.53	158.07	
“road conditions brevard county florida”	31.05	135.54	
“park tudor indianapolis”	29.16	133.80	
“outer space decor”	0.35	80.48	globale
“funny pictures”	0.40	69.33	
“dictionary”	0.59	77.81	
“ringtones”	0.75	84.23	
“southpark cartman”	1.74	55.07	

TAB. 4.3 – Les 5 plus faibles et/ou plus fortes valeurs de LocationKL et Kurtosis pour chaque classe de requêtes dans notre ensemble de requêtes de test.

Classificateur	Classe	Précision	Rappel	F-mesure	Accuracy
Meta	locale-implicite	0.81	0.74	0.78	88%
	locale-explicite	0.91	0.95	0.93	
	globale	0.90	0.92	0.91	
	moyenne	0.88	0.88	0.88	
SVM	locale-implicite	0.78	0.74	0.76	87%
	locale-explicite	0.95	0.85	0.90	
	globale	0.89	0.93	0.91	
	moyenne	0.87	0.87	0.87	
J48	locale-implicite	0.78	0.82	0.8	87.5%
	locale-explicite	0.92	0.88	0.90	
	globale	0.91	0.90	0.90	
	moyenne	0.88	0.87	0.88	
Ripper	locale-implicite	0.73	0.80	0.77	86.5%
	locale-explicite	0.92	0.90	0.91	
	globale	0.91	0.88	0.90	
	moyenne	0.87	0.86	0.87	
Bayes	locale-implicite	0.75	0.71	0.73	85%
	locale-explicite	0.94	0.80	0.87	
	global	0.86	0.93	0.90	
	moyenne	0.85	0.85	0.85	

TAB. 4.4 – Résultats de classification des requêtes de test.

6.3.2 Une évaluation comparative avec l’approche de Gravano et al. [89]

L’objectif de cette évaluation est de comparer notre approche à une approche de classification proposée dans le domaine décrite dans Gravano et al. [89].

La raison principale pour laquelle nous avons choisi de comparer notre approche à celle de Gravano et al. est que les deux approches se basent sur la même source d’évidence pour définir les attributs de classification, notamment les “ n ” premiers résultats de recherche retournés par un moteur Web.

6.3.2.1 Aperçu de l’approche de Gravano et al. [89]

Dans l’approche de Gravano et al. les requêtes sont classées en deux classes globales ou locales. Plusieurs attributs de classification sont proposés, tous définis sur la base de fréquences et ou de distributions des localisations dans les “ n ” premiers résultats retournés par un moteur de recherche Web :

- le nombre moyen (par page Web retournée) de localisations dans les résultats récupérés.
- le nombre total des localisations à travers tous les documents retournés, divisé sur le nombre de documents retournés.
- la fraction des pages qui comportent au moins une localisation,
- le pourcentage de mots qui correspondent à des localisations à travers toutes les pages,
- l’écart type du nombre total de localisations par page Web retournée.
- ...

Certaines de ces mesures sont réalisées avec distinction du type de la localisation dans la hiérarchie des places (pays, état, ville), ou sont faites sur tous les types confondus. Ces différents attributs de classification ont été testés seuls ou combinés avec plusieurs classificateurs de l’état de l’art.

Le classificateur Ripper permet d’atteindre l’une des meilleures performances de classification en utilisant une seule règle basée uniquement sur le nombre moyen de localisations géographiques par page Web retournée, si ce nombre dépasse un certain seuil, la requête est considérée comme locale sinon elle est considérée comme globale.

Nous avons implémenté l’approche de Gravano et al. en utilisant ce classificateur et cette règle et nous l’avons comparé à notre approche avec le même classificateur appris avec nos deux attributs de classification.

6.3.2.2 Résultats de l'évaluation comparative

Comme mentionné plus haut, l'approche de Gravano et al. est désignée seulement pour classifier les requêtes en deux classes locales ou globales. De ce fait, nous comparons les performances de classification de nos deux approches uniquement sur la base de ces deux classes. La table 4.5 reporte les valeurs de

Approche	[89] approche			notre approche		
	locale	globale	moyenne	locale	globale	moyenne
Précision	0.76	0.88	0.82	0.87 (14%)	0.92 (5%)	0.90 (10%)
Rappel	0.62	0.93	0.77	0.91 (47%)	0.89 (-4%)	0.89 (16%)
F-measure	0.68	0.90	0.79	0.89 (31%)	0.90 (0%)	0.89 (13%)
Accuracy	86%			90% (5%)		

TAB. 4.5 – Performance de la classification sur les classes locale et globale : comparaison entre notre approche et l'approche de [89] sur l'ensemble de nos requêtes de test.

précision, rappel, F-mesure et *accuracy* obtenues par le classificateur Ripper selon les deux approches.

La gain de performance est calculé selon la formule :

$$\text{Gain de performance} = (P_{\text{notre-approche}} - P_{[89]\text{approche}}) / P_{[89]\text{approche}} * 100$$

où P représente la valeur de la métrique d'évaluation de la performance (précision, rappel, F-mesure ou *accuracy*) pour chaque approche. Il est noté entre parenthèse) dans le table.

Nous constatons, qu'en général, notre approche atteint de meilleurs résultats de classification que l'approche [89], avec une amélioration de 13% de la métrique F-measure et 5% d'*accuracy*. Le gain de performance de notre approche est principalement sur la classification des requêtes locales avec un gain de 31% en F-measure. Les performances de classification des requêtes globales restent à peu près similaires entre les deux approches.

6.3.3 Évaluation de l'efficacité de l'accès contextuel guidé par le type de la requête

Nous nous intéressons dans cette seconde série d'expérimentations à évaluer l'efficacité de l'accès contextuel guidé par le type de la requête selon les deux approches par re-formulation de la requête et par ré-ordonnancement des résultats de recherche. Cette évaluation est basée sur l'ensemble des requêtes classées comme locales implicites par notre classificateur.

6.3.3.1 Estimation expérimentale du paramètre de ré-ordonnement

Dans un premier temps nous étudions l'impact du paramètre α de combinaison du score original du document avec le score géographique (cf. formule 4.9) sur les performances de recherche. Nous avons calculé la précision de la recherche et la nDCG aux premiers rangs (1, 3, 5, 7 et 10) pour chacune des valeurs de α dans l'intervalle $[0, 1]$. Pour $\alpha = 1$, il s'agit de la recherche basée seulement sur le score géographique du document. Pour $\alpha \in [0, 9]$, il s'agit de la recherche basée sur la combinaison du score original du document et de son score géographique. Pour $\alpha = 0$, il s'agit de la recherche classique. La figure 4.8 (respect. figure 4.9) montre l'effet de la variation du paramètre de combinaison α dans l'intervalle $[0, 1]$ sur l'efficacité de la recherche en termes de la moyenne des précisions P1, P2, P3, P5, P7 et P10 (respect. en termes de la moyenne des gains cumulatifs nDCG1, nDCG3, nDCG5, nDCG7 et nDCG10) sur l'ensemble des requêtes de test.

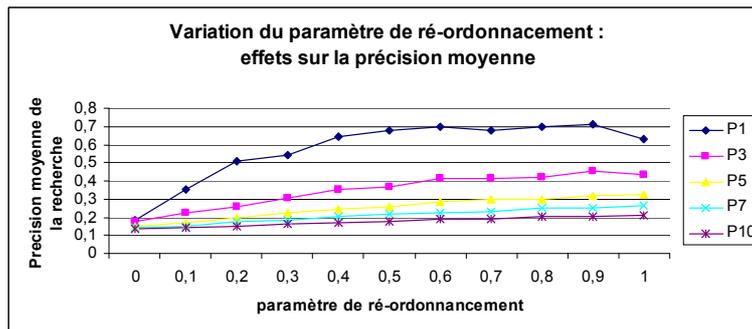


FIG. 4.8 – Variation de l'efficacité de la recherche en termes de précision moyenne en fonction du paramètre de ré-ordonnement α .

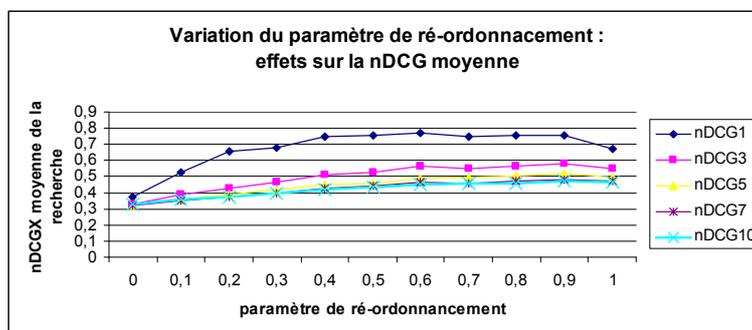


FIG. 4.9 – Variation de l'efficacité de la recherche en termes de nDCG moyenne en fonction du paramètre de ré-ordonnement α .

Nous constatons que la meilleure valeur de α est identifiée à 0.9, c'est-à-dire que la meilleure performance est obtenue pour une contribution faible du score original du document. Ceci peut être expliqué par le fait que les 10 premiers

documents répondent bien à la thématique de la requête et que la seule caractéristique distinctive est la façon dont ils correspondent à la localisation de l'utilisateur.

6.3.3.2 Comparaison de l'efficacité de l'accès contextuel sur toutes les requêtes

Dans un second temps, nous présentons les résultats de comparaison des performances de recherche entre la *baseline*, la méthode par ré-ordonnement utilisant la valeur optimale du paramètre $\alpha = 0.9$ et la méthode par reformulation de la requête. Afin d'évaluer si les différences observées sont significatives d'un point de vue statistique, nous avons mené un t-test sur les moyennes des précisions et nDCG obtenues à différents points (1, 3, 5, 7 et 10) entre les différentes approches. Nous supposons que la différence entre les moyennes est significative si $p < 0.05$.

La figure 4.10 et la figure 4.11 illustrent respectivement la précision et nDCG moyennes obtenues à différents points de coupure (1, 3, 5, 7 et 10) par chaque méthode sur l'ensemble des requêtes de test.

Les résultats montrent, en général, une amélioration des performances de nos deux méthodes d'accès contextuel par rapport à la *baseline*.

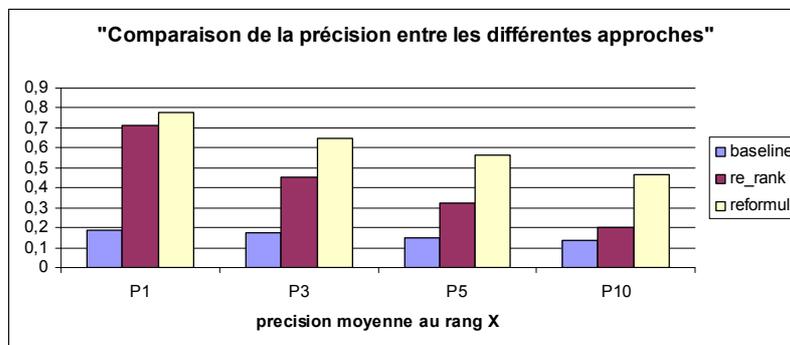


FIG. 4.10 – Comparaison des moyennes de Précision@X entre nos approches contextuelles et la *baseline* sur toutes les requêtes de test.

En particulier, le calcul du gain de performance entre notre méthode de ré-ordonnement et la *baseline* nous donne des améliorations de 51.51% à P1, 36.14% à P3, 25.59% à P5, 18.18% à P7, 12.96% à P10, 102.27% à nDCG1, 77.21% à nDCG3, 62.59% à nDCG5, 50.75% à nDCG7 et 47.40%. Nous constatons que le meilleur gain en terme de précision et de nDCG est obtenu au rang 1, et qu'il diminue à mesure que les rangs augmentent. Le calcul d'un t-test sur les moyennes des valeurs de précision et nDCG obtenues par ces deux

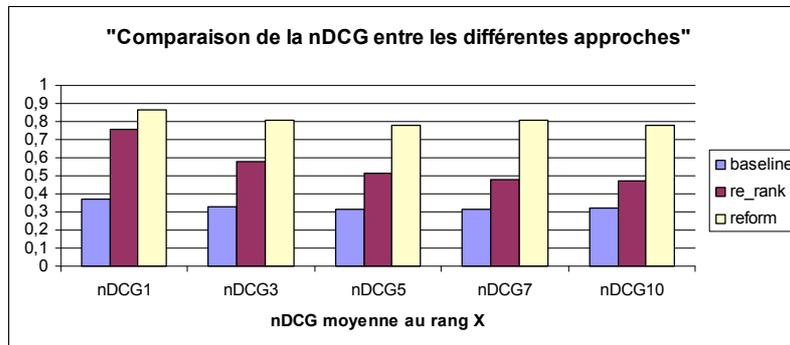


FIG. 4.11 – Comparaison des moyennes de $nDCG@X$ entre nos approches contextuelles et la *baseline* sur toutes les requêtes de test.

méthodes, permet de considérer qu'il y a des différences significatives entre les deux méthodes, au seuil de confiance de 0.05.

De plus, la méthode par re-formulation améliore les performances de recherche comparativement à la *baseline* mais aussi à la méthode par réordonnement notamment à partir du rang 3. En effet, la méthode par re-formulation permet de récupérer des documents pertinents que la requête initiale n'a pas retournés ou qui étaient plus bas dans la liste des résultats initiaux du SRI standard ($>$ aux 50 premières pages aspirées). Le calcul d'un t-test sur les moyennes des valeurs de précision et $nDCG$ obtenues par ces deux méthodes, permet de considérer que les gains de performance de 40.70% à P3, 58.60% à P5, 67.17% à P7, 79.91% à P10, 39.31% à $nDCG3$, 51.22% à $nDCG5$, 69.07% à $nDCG7$ et 62.19% à $nDCG10$ obtenus par la méthode par re-formulation comparativement à la méthode par re-ordonnement étaient aussi significatifs au seuil de confiance de 0.05.

6.3.3.3 Comparaison de l'efficacité de l'accès contextuel sur les requêtes mal classées

Par ailleurs, nous avons voulu étudier l'effet de nos approches guidées par le type de la requête sur les performances des résultats de recherche dans le cas de requêtes mal classées (notamment des requêtes globales).

La figure 4.12 et la figure 4.13 montrent respectivement les valeurs de la précision et de la $nDCG$ à différents rangs (Top1, Top2, Top3, Top5, Top7 et Top10) obtenues par chaque approche moyennées uniquement sur les requêtes mal classées. Comme l'on peut constater nos deux approches dégradent légèrement la précision et la $nDCG$ aux différents rangs (Top1, Top2, Top3, Top5, Top7 et Top10). Après calcul du t-test sur les moyennes des précisions et $nDCG$ entre nos approches et la *baseline*, nous constatons :

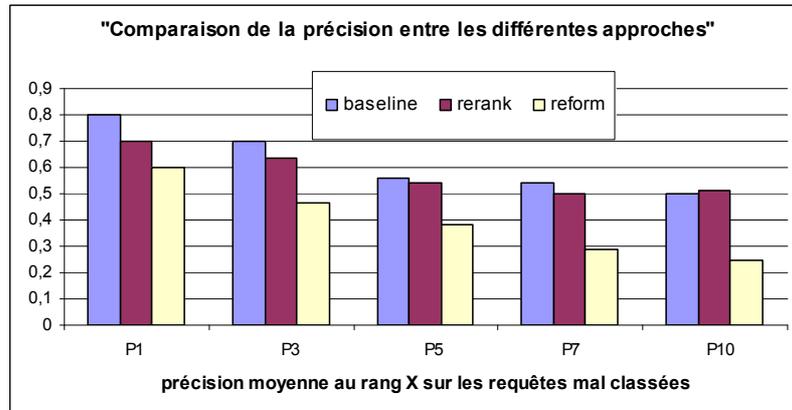


FIG. 4.12 – Comparaison des moyennes de Précision@X entre nos approches contextuelles et la *baseline* sur les requêtes mal classées.

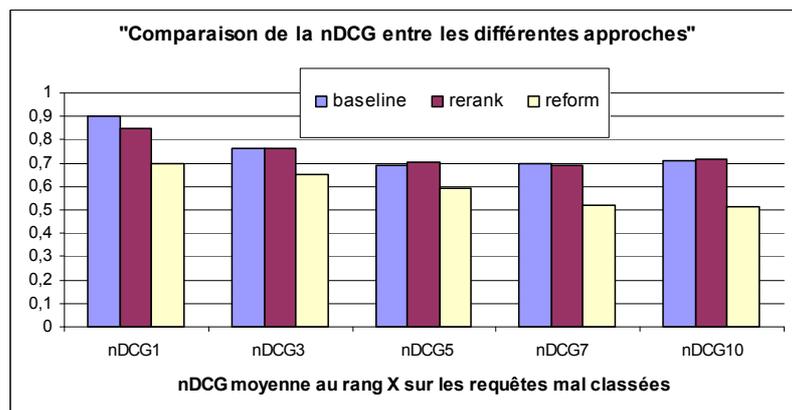


FIG. 4.13 – Comparaison des moyennes de nDCG@X entre nos approches contextuelles et la *baseline* sur les requêtes mal classées.

- que cette dégradation de performance n'est pas significative pour l'approche de re-ordonnancement au seuil 0.05,
- et que cette dégradation de performance est significative pour l'approche de re-formulation pour les rangs 7 et 10 avec une p -valeur > 0.05 .

7 Conclusion

Nous avons présenté au cours de ce chapitre notre contribution portant sur la proposition d'une méthode pour la détermination du type du besoin derrière la requête (global ou local) et son exploitation pour adapter l'accès à l'information en tenant compte du type de la requête. Notre approche pour l'identification du type de la requête est basée sur la définition d'un modèle de langue de localisation de la requête à partir des " n " premiers résultats de

recherche retournés par un moteur de recherche web général.

Ce modèle présente l'avantage de formaliser la notion de sensibilité de la requête à la localisation. De plus, il nous offre une source d'évidence pour calculer deux caractéristiques de la requête, notamment les mesures de divergence de Kullback-Leibler et de Kurtosis. Ces deux mesures, utilisées comme attributs de classification, nous ont permis de bien classer les trois types de requêtes : global, local explicite et local implicite, avec une F-mesure globale de 88%.

Nous avons testé deux techniques d'adaptation de la recherche pour intégrer les résultats de la classification des requêtes selon leur sensibilité à la localisation, pour adapter les résultats de la recherche d'un moteur web général. Nous avons montré leur efficacité à améliorer les performances de recherche.

Chapitre 5

Agrégation prioritaire de plusieurs éléments contextuels : proposition et évaluation

1 Introduction

L'objectif fondamental de l'accès contextuel à l'information est d'optimiser la pertinence des résultats de recherche, en impliquant deux étapes complémentaires : définition du contexte du besoin en information de l'utilisateur, puis adaptation de la recherche en prenant la représentation du contexte, en considération dans le processus de sélection de l'information. 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 plusieurs éléments contextuels. Le défi majeur dans ce cadre, consiste à combiner les pertinences individuelles d'un document issues de plusieurs dimensions contextuelles pour définir un seul score de pertinence du document. Ce problème a été généralement abordé par l'adoption des techniques de combinaison linéaire appliquées indépendamment des préférences des utilisateurs sur les dimensions de la pertinence. Une piste de recherche intéressante serait de rendre l'utilisateur un acteur dans la détermination d'un tel modèle d'agrégation.

Notre contribution [5]¹ porte sur la proposition d'un modèle de pertinence multidimensionnelle basé sur la combinaison de la pertinence thématique (traditionnellement à la base de la pertinence calculée par les SRI standards) avec des facteurs de pertinence contextuelle notamment : les centres d'intérêts et la localisation de l'utilisateur mobile. Ce modèle de pertinence multidimensionnelle

¹Ce travail a été réalisé en collaboration avec le professeur Gabriella Pasi et Célia da Costa Pereira de l'université degli Studi di Milano, Italie.

présente la particularité d’exploiter deux opérateurs d’agrégation prioritaire permettant d’adapter les résultats de la recherche selon les préférences de l’utilisateur exprimées sur les critères de pertinence. Ainsi, pour une même requête et un même document, plusieurs scores de pertinence peuvent être calculés en fonction de l’ordre de priorité exprimé par un utilisateur sur les différents critères de pertinence. Nous avons également proposé un cadre d’évaluation centré-utilisateur pour tester l’efficacité des opérateurs d’agrégation prioritaire comparativement à l’approche standard par combinaison linéaire.

Ce chapitre est organisé comme suit. La section 2 présente les motivations pour la proposition de ce cadre d’agrégation prioritaire. La section 3 présente le cadre d’agrégation prioritaire, notamment les opérateurs d’agrégation “*scoring*” et “*and*”. La section 4 présente l’application de ce cadre à notre problème de combinaison multi-dimensionnelle pour la RI mobile et les dimensions de pertinence que nous avons considérées dans notre travail. La section 5 présente notre cadre d’évaluation et les résultats obtenus. La dernière section conclut le chapitre.

2 Problématique et motivations

Un des défis majeurs de la RI contextuelle mobile après la modélisation du contexte, concerne l’intégration des dimensions contextuelles modélisées dans le processus de sélection de l’information. Lorsque plusieurs dimensions contextuelles sont considérées et lorsque on peut définir un score contextuel du document relativement à chacune de ces dimensions, l’intégration des dimensions contextuelles revient alors à la combinaison des scores contextuels individuels du document. Le principal problème qui se pose alors est : comment définir la fonction de combinaison des pertinences individuelles d’un document issues des dimensions contextuelles pour définir un seul score de pertinence globale du document. Juste comme exemple introductif, supposons un modèle contextuel intégrant à la fois plusieurs facteurs contextuels (e.g. les centres d’intérêts et la localisation de l’utilisateur). Et supposons la requête “*événement culturel*”, la pertinence d’un document par rapport à cette requête peut être exprimée selon la dimension des centres d’intérêts de l’utilisateur “*événement jazz*” et de la dimension de localisation de l’utilisateur “*Londres*”.

Ce problème a été principalement abordé en adoptant des techniques de combinaison appliquées de manière indépendante des préférences des utilisateurs sur les divers facteurs de pertinence considérés [192, 87, 229]. Cependant, les utilisateurs peuvent être différemment intéressés à chaque critère contextuel en fonction de leur préférences et de leurs besoins informationnels actuels [223]. Pour illustrer cela, revenons à notre exemple, selon les préférences

de l'utilisateur nous pouvons imaginer plusieurs scénarios possibles. Dans un premier scénario, l'utilisateur pourrait favoriser des documents sur son centre d'intérêt "*événement jazz*" bien que la localisation de l'événement est différente de sa ville "*Londres*". Dans un autre scénario, l'utilisateur qui ne désire pas se déplacer, pourrait préférer des documents sur n'importe quel "*événement culturel*" dans "*Londres*" (indépendamment s'il est un "*événement de jazz*" ou non) sur des documents parlants "*d'événements de jazz*" dans d'autres places en dehors de "*Londres*".

Nos objectifs dans cette contribution portent sur l'expérimentation d'autres opérateurs de combinaison de critères qui permettraient de prendre en considération la relation d'ordre de priorité que pourrait exprimer un utilisateur sur l'ensemble des dimensions de pertinence intégrées dans le modèle de combinaison de pertinence multi-dimensionnelle proposé.

Plus précisément, pour pouvoir prendre en compte la relation d'ordre de priorité entre les dimensions de pertinence, nous avons adopté les deux opérateurs d'agrégation prioritaire multi-critères définis dans [265]. Ces opérateurs ont la particularité de rendre le calcul du poids d'un critère dans le score final d'un document, dépendant du score de satisfaction du critère de plus haute priorité. Nous avons proposé d'appliquer ces deux opérateurs dans le cadre d'un modèle de RI mobile intégrant les trois dimensions de pertinence : thématique, centre d'intérêt et localisation de l'utilisateur. En vue de tester l'efficacité de ces opérateurs, nous avons proposé un cadre d'évaluation centré-utilisateur. Cette évaluation est basée sur la comparaison des listes des résultats de recherche de ces opérateurs par rapport à la technique standard de combinaison linéaire, à travers divers scénarios d'évaluation construits en considérant différents types d'utilisateurs selon leurs préférences d'ordre sur les dimensions testées.

3 Présentation de l'approche d'agrégation prioritaire multi-critères

Le problème de l'agrégation prioritaire est typique dans le cas où nous désirons modéliser une relation d'ordre (\succ) entre plusieurs critères. Dans de tels cas, le manque de satisfaction d'un critère de priorité élevée, ne peut être compensé par la satisfaction d'un critère de priorité inférieure.

Dans cette section nous donnons une synthèse des opérateurs d'agrégation prioritaire de multi-critères proposés par Yagger [265] et exploitée dans [66] et [67] pour la modélisation du problème d'agrégation prioritaire de plusieurs dimensions de pertinence. Dans la section 3.1 la modélisation du problème de l'agrégation prioritaire multi-critères est introduite comme un problème de prise de décision multi-critères où les alternatives possibles à la sélection sont

les documents de la collection considérée. Dans la section 3.2 nous décrivons les opérateurs d'agrégation basés-priorité “*scoring*” et “*and*”.

3.1 La modélisation du problème : agrégation prioritaire multi-critères

Le problème d'agrégation prioritaire multi-critères est présenté comme un problème décisionnel. Plus formellement, soit un cadre de processus décisionnel, dans lequel nous avons les éléments suivants :

- L'ensemble C des critères considérés : $C = \{C_1, \dots, C_n\}$; afin de simplifier la notation, nous désignons aussi par C_i la fonction qui évalue le i^{e} critère, nous supposons de plus qu'après que l'utilisateur ait exprimé son choix de préférence sur l'ordre des n critères considérés, nous dénotons par C_1 le critère le plus préféré (i.e., au premier rang de la liste des préférences de l'utilisateur), et par C_n le critère le moins préféré (i.e., au dernier rang de la liste des préférences de l'utilisateur), et nous supposons que C_i est préféré à C_j si et seulement si $i < j$.
- La collection de documents D ;
- Le score de satisfaction du document d par rapport au critère de pertinence j , noté : $C_j(d)$;
- Une fonction d'agrégation F pour calculer pour chaque document $d \in D$ un score global sur la base des scores d'évaluation des critères considérés : $RSV(d) = F(C_1(d), \dots, C_n(d))$.

Pour chaque critère $C_i \in C$, un poids d'importance est calculé de sorte qu'il soit à la fois dépendant du document et de l'utilisateur. Plus spécifiquement, le calcul du poids d'un critère dépend à la fois de :

- l'ordre de préférence exprimé par l'utilisateur sur l'ensemble des critères,
- le poids calculé pour le critère C_{i-1} (d'une plus grande priorité par rapport à C_i),
- le score de satisfaction du document par rapport au critère de pertinence C_{i-1} .

En d'autres termes, pour un document d , pour chaque critère C_i un poids d'importance $\lambda_i \in [0, 1]$ est calculé, qui varie suivant les documents considérés. Les poids associés aux critères ordonnés par l'utilisateur selon ses préférences, sont calculés comme suit :

- pour chaque document d , le poids du critère de plus haute priorité C_1 est fixé à 1, i.e., par définition on a : $\forall d \lambda_1 = 1$;

- le poids des autres critères C_i , $i \in [2, n]$, sont calculés comme suit :

$$\lambda_i = \lambda_{i-1} \cdot C_{i-1}(d), \quad (5.1)$$

Où $C_{i-1}(d)$ est le score de satisfaction du critère C_{i-1} par le document d , et λ_{i-1} est le poids d'importance du critère C_{i-1} .

L'instanciation de ce modèle d'agrégation est effectué par définition de la fonction F , i.e. l'opérateur d'agrégation des scores individuels des critères de pertinence. Dans la suite, nous décrivons deux fonctions d'agrégation permettant d'instancier ce modèle.

3.2 Les opérateurs d'agrégation prioritaire

Dans cette section, nous présentons deux formalisations alternatives de l'opérateur d'agrégation des priorités proposé F , notamment : l'opérateur prioritaire “*scoring*” puis l'opérateur prioritaire “*and*”.

3.2.1 L'opérateur prioritaire “*scoring*”

Cet opérateur permet de calculer la valeur du score global à partir des scores individuels de chaque critère, où le poids de chaque critère dans le score global dépend à la fois des poids et des scores de satisfaction des critères les plus prioritaires : plus est le score de satisfaction des critères de plus hautes priorités, plus le score de satisfaction d'un critère de moindre priorité influence peu le score global d'un document.

L'opérateur d'agrégation “*scoring*”, dénoté F_s , est formellement défini comme suit : $F_s : [0, 1]^n \rightarrow [0, n]$ et pour un document donné d ,

$$F_s(C_1(d), \dots, C_n(d)) = \sum_{i=1}^n \lambda_i \cdot C_i(d). \quad (5.2)$$

Où chaque C_i représente l'une des dimensions de pertinence considérées.

Pour illustrer le fonctionnement de cet opérateur, prenons un exemple simple avec quatre critères de pertinence (i.e., $n = 4$) : C_1, C_2, C_3 et C_4 . Par convention, C_1 est le critère de plus important et C_4 le critère le moins important. Soit un document d_i pour lequel on a calculé le score de satisfaction par rapport à chaque critère comme suit : $C_1(d_i) = 0.6$, $C_2(d_i) = 0.8$, $C_3(d_i) = 0.9$, et $C_4(d_i) = 1$, le calcul des poids des critères est alors donné par :

- $\lambda_1 = 1$;

- $\lambda_2 = \lambda_1 \cdot C_1(d_i) = 0.6$;
- $\lambda_3 = \lambda_2 \cdot C_2(d_i) = 0.48$;
- $\lambda_4 = \lambda_3 \cdot C_3(d_i) = 0.432$;

et le score global du document RSV_s sera

$$RSV_s(d_i) = 1 \cdot 0.6 + 0.6 \cdot 0.8 + 0.48 \cdot 0.9 + 0.432 \cdot 1 = 1.944.$$

Maintenant, si on échange le score de satisfaction de C_2 et C_3 de sorte que $C_2(d_i) = 0.9$ et $C_3(d_i) = 0.8$, le calcul des poids des critères devient :

- $\lambda_1 = 1$;
- $\lambda_2 = \lambda_1 \cdot C_1(d_i) = 0.6$;
- $\lambda_3 = \lambda_2 \cdot C_2(d_i) = 0.54$;
- $\lambda_4 = \lambda_3 \cdot C_3(d_i) = 0.432$;

et le nouveau score global du document RSV_s sera par conséquent :

$$RSV_s(d_i) = 1 \cdot 0.6 + 0.6 \cdot 0.9 + 0.54 \cdot 0.8 + 0.432 \cdot 1 = 2.004.$$

Ce résultat est justifié par le fait que le critère de seconde priorité C_2 est mieux satisfait par le document d_i dans ce cas qu'il ne l'était dans le premier cas.

3.2.2 L'opérateur prioritaire “and”

La particularité de cet opérateur, qui le distingue aussi de l'opérateur traditionnel “min”, est que la mesure dans laquelle le critère le moins satisfait est considéré dans le score global, dépend de son importance pour l'utilisateur. En d'autres termes, si le critère le moins satisfait n'est pas du tout important pour l'utilisateur, alors son score de satisfaction ne doit pas être considéré, tandis que s'il est le critère le plus important pour l'utilisateur, alors seulement son score de satisfaction est considéré.

De cette façon, si on considère un document d , pour lequel le critère C_k ayant le score de satisfaction le plus bas, est au même temps le critère le moins important, alors le score global de satisfaction du document sera plus grand que $C_k(d)$; ça ne sera pas C_k comme cela aurait été le cas avec l'opérateur traditionnel “min”, moins un critère est important, moins il a de chance pour représenter le score global de satisfaction du document.

L'opérateur d'agrégation “and”, dénoté F_a , est défini comme suit : $F_a : [0, 1]^n \rightarrow [0, 1]$, et pour un document d ,

$$F_a(C_1(d), \dots, C_n(d)) = \min_{i=1, n} (\{C_i(d)\}^{\lambda_i}). \quad (5.3)$$

Comme simple exemple illustratif, supposons que l'évaluation de la pertinence d'un document d_i est effectuée selon 4 dimensions de pertinence, et que les scores de pertinence individuels sont : $C_1(d_i) = 0.9$, $C_2(d_i) = 0.7$, $C_3(d_i) = 0.9$, et $C_4(d_i) = 0.6$, avec $C_1 \succ C_2 \succ C_3 \succ C_4$, on a :

- $\lambda_1 = 1$;
- $\lambda_2 = \lambda_1 \cdot C_1(d_i) = 0.9$;
- $\lambda_3 = \lambda_2 \cdot C_2(d_i) = 0.63$;
- $\lambda_4 = \lambda_3 \cdot C_3(d_i) = 0.567$;

et le score global du document sera alors :

$$\begin{aligned} F_a(C_1(d_i), \dots, C_4(d_i)) &= \min(0.9^1, 0.7^{0.9}, 0.9^{0.63}, 0.6^{0.567}) \\ &= \min(0.9, 0.72, 0.93, 0.74) = 0.72. \end{aligned}$$

Si nous échangeons les scores de satisfaction des critères C_2 et C_3 de sorte que $C_2(d_i) = 0.9$ $C_3(d_i) = 0.7$, le calcul des poids des critères nous donne :

- $\lambda_1 = 1$;
- $\lambda_2 = \lambda_1 \cdot C_1(d_i) = 0.9$;
- $\lambda_3 = \lambda_2 \cdot C_2(d_i) = 0.81$;
- $\lambda_4 = \lambda_3 \cdot C_3(d_i) = 0.567$;

et le score global du document devient :

$$\begin{aligned} F_a(C_1(d_i), \dots, C_4(d_i)) &= \min(0.9^1, 0.9^{0.9}, 0.7^{0.81}, 0.6^{0.567}) \\ &= \min(0.9, 0.909, 0.749, 0.748) = 0.748. \end{aligned}$$

Ce résultat est justifié par le fait que le critère de seconde priorité C_2 est mieux satisfait par le document d_i dans ce cas qu'il ne l'était dans le premier cas.

4 Combinaison multi-dimensionnelle pour la RI mobile

Nous proposons d'appliquer le modèle d'agrégation prioritaire pour la combinaison de pertinence multi-dimensionnelle pour la RI mobile. Nous avons voulu intégrer deux dimensions contextuelles en l'occurrence les centres d'intérêts et la localisation de l'utilisateur pour le calcul de la pertinence finale d'un document retourné par un SRI. Le calcul de la pertinence finale d'un document sera alors dépendant du score de satisfaction des trois critères suivants :

$$C = \{\text{thème}, \text{centres d'intérêts}, \text{localisation}\}.$$

La pertinence finale d'un document d , par rapport à une requête Q , les centres d'intérêts I et la localisation de l'utilisateur L , sera présentée par un score global $RSV(d)$ définie par :

$$RSV(d) = F(\text{thème}(d, Q), \text{Intérêts}(d, I), \text{localisation}(d, L)).$$

Où F est l'opérateur d'agrégation prioritaire, $\text{thème}(d, Q)$, $\text{Intérêts}(d, I)$, et $\text{localisation}(d, L)$ sont respectivement la fonction évaluant le score individuel *thème*, *centres d'intérêts*, et *localisation*.

Dans la suite de cette section, nous présentons une description et une définition formelle de nos trois critères ainsi que des fonctions de score associées.

4.1 La dimension thème

La dimension ou critère “Thème” fait référence à la pertinence thématique traditionnellement calculée par les systèmes de RI standards. La pertinence thématique est généralement mesurée selon l’un des nombreux modèles de RI.

L’un des modèles de référence en RI est le modèle probabiliste [119] avec le schéma de pondération BM25 comme fonction d’appariement document requête. Pour cette raison, nous avons adopté ce modèle pour le calcul de la pertinence thématique. Bien évidemment, il est tout à fait possible de la calculer en utilisant n’importe quel autre modèle de RI.

BM25 est une fonction de recherche basée “sac de mots”, elle permet d’ordonner l’ensemble des documents sur la base des occurrences des termes de la requête donnée dans les documents.

Plus précisément, soit une requête Q contenant les termes t_1, \dots, t_n , le score de la pertinence thématique *thème* d’un document d est donné par :

$$\text{thème}(d, Q) = \sum_{i=1}^n IDF(t_i) * \frac{f(t_i, d) * (k_1 + 1)}{f(t_i, d) + k_1 * (1 - b + b * \frac{|d|}{avgdl})} \quad (5.4)$$

Où $f(t_i, d)$ est la fréquence du terme t_i dans le document d , $|d|$ est le nombre de mots appartenant au document d , et $avgdl$ la taille moyenne des documents dans la collection à partir de laquelle les documents sont retournés. k_1 et b sont des paramètres, fixés par défaut tels que : $k_1 = 2.0$ et $b = 0.75$. $IDF(t_i)$ est la fréquence inverse de document (*inverse document frequency*) qui mesure l’importance du terme t_i dans l’ensemble du corpus, généralement calculé comme :

$$IDF(t_i) = \log \frac{N - n(t_i) + 0.5}{n(t_i) + 0.5},$$

Où N est le nombre total de documents dans la collection, et $n(t_i)$ est le nombre de documents contenant le terme t_i .

4.2 La dimension centres d’intérêts

Comme première dimension contextuelle intégrée à notre modèle de pertinence multi-dimensionnelle, nous avons voulu considérer les centres d’intérêts

de l'utilisateur. Les centres d'intérêts sont en effet considérés comme le facteur contextuel le plus important qui réduit la difficulté d'une requête ambiguë dans une tâche de recherche *ad hoc* [175]. La dimension "Intérêts" permet ainsi de mesurer le degré de similitude d'un document avec le profil utilisateur représenté par ces centres d'intérêts.

Pour intégrer la dimension "Intérêts" dans notre travail ici, nous avons exploité le modèle sémantique du profil utilisateur proposé dans [69], où les centres d'intérêts de l'utilisateur sont présentés comme un graphe de concepts pondérés de l'ontologie ODP. Chaque concept c_j de l'ODP est représenté par un vecteur de termes \vec{c}_j extraits des pages web classées sous ce concept, ainsi que ces concepts descendants. Les documents de la collection sont également représentés par des vecteurs de termes pondérés. Nous avons déjà décrit cette approche dans notre première contribution dans cette thèse, nous nous référons donc à la section ?? du chapitre 3, pour une description plus détaillée de la méthode de modélisation des centres d'intérêts.

Le calcul du score de pertinence "Intérêts" d'un document d selon le profil utilisateur I , est effectué selon une mesure de similarité basée-termes, notamment la mesure de similarité cosinus entre le document d et les top k concepts représentant le profil utilisateur I , comme suit :

$$Intérêts(d, I) = \sum_{c_j \in I \wedge j \in [1, k]} sw(c_j) * \cos(\vec{d}, \vec{c}_j) \quad (5.5)$$

Où $sw(c_j)$ est le poids du concept c_j dans le profil utilisateur I .

4.3 La dimension localisation

Comme seconde dimension contextuelle, nous proposons d'intégrer la dimension de localisation comme notre troisième critère de pertinence, en plus des deux critères de thème et de centre d'intérêt.

En effet, nous sommes également conscients de l'importance de la localisation géographique pour la recherche mobile, notamment pour des requêtes sensibles à la localisation représentant plus de 30% des requêtes des mobinautes selon l'étude récente de [53].

La prise en compte des besoins en information localisés ou géographiques des utilisateurs et la localisation de résultats de la recherche est un problème connu dans le domaine de la RI géographique [117]. Parmi les diverses fonctions d'ordonnement géographique définies dans la littérature, nous avons adopté la fonction de pondération géographique présentée dans [255], comme notre score de pertinence géographique ou de localisation d'un document.

Soit une hiérarchie de places géographiques GH , une place géographique L contenue dans la requête de l'utilisateur (ou inférée automatiquement par un service de localisation), un document d , le score de la pertinence de *localisation* est donnée par la formule suivante :

$$localisation(d, L) = f(L) + \sum_{L_i \in descendants(L)} f(L_i) \quad (5.6)$$

Où $f(L)$ fait référence au nombre d'occurrences de la localisation L dans le document d , les localisations descendantes L_i de la localisation donnée L sont identifiées à partir de GH .

5 Évaluation expérimentale

Nous nous confrontons dans notre travail d'évaluation à l'absence d'une collection d'évaluation standard adaptée pour évaluer un SRI qui intègre des critères de pertinences contextuelles et des préférences utilisateur. Nous avons donc proposé un cadre d'évaluation qui intègre l'utilisateur et son contexte dans le processus d'évaluation via la simulation contextuelle.

Notre cadre expérimental suit donc la démarche d'évaluation basée-simulation [207]. L'objectif de notre évaluation expérimentale est double : 1) étudier l'efficacité de la combinaison de la pertinence multi-critères par rapport à la pertinence thématique standard, et 2) comparer l'effet des opérateurs d'agrégation prioritaire sur la performance de la recherche par rapport à la combinaison linéaire standard.

Dans la suite de cette section, nous présentons en premier lieu notre cadre expérimental puis nous présentons et discutons les résultats obtenus.

5.1 Cadre expérimental

Pour des fins expérimentales, nous utilisons une partie de la branche de l'ontologie ODP composé de l'ensemble des pages Web classées sous la région des États-Unis. La raison derrière notre choix est double. D'un côté l'ODP est la base de données de contenu web la plus largement distribuée en accès libre et manuellement classée. De plus, ces pages Web sont à la fois classées sous des concepts thématiques mais aussi sous des places géographiques, ce qui nous permet d'étudier l'ensemble de nos critères : thématique, centre d'intérêt et localisation. La figure 5.1 illustre cette double structuration de l'ODP.

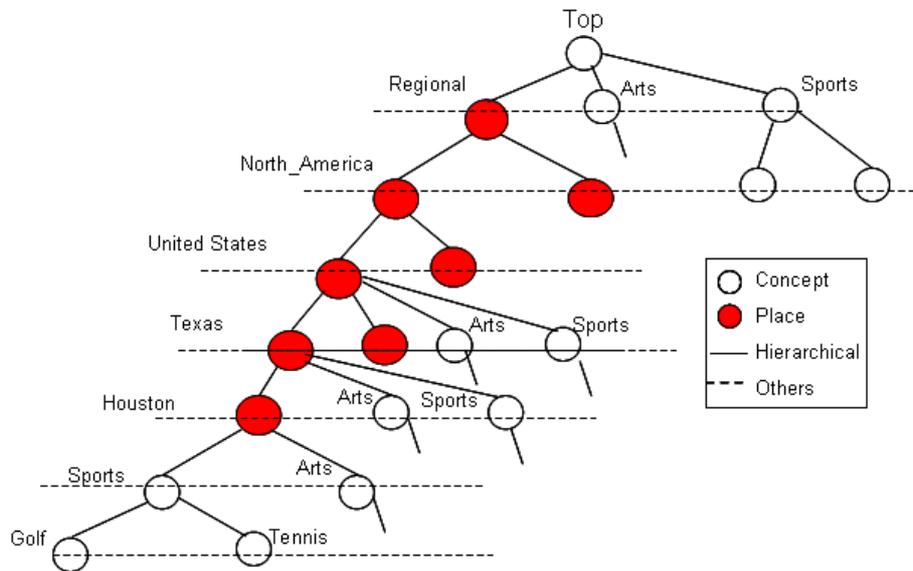


FIG. 5.1 – Un aperçu de la structure hiérarchique de l'ODP.

Nous avons ainsi aspiré un ensemble de 171541 pages web, que nous avons divisé en deux sous-ensembles : un ensemble de test (T) constitué de 2/3 des pages web aspirées, et un sous-ensemble d'apprentissage pour apprendre les profils utilisateurs (P), composé des pages web restantes. Les documents ont été aléatoirement affectés par concept et localisation à chacun des deux sous-ensembles.

5.1.1 Collection de documents

Le sous-ensemble (T) est utilisé comme notre collection de documents pour la recherche. Nous l'avons indexé en utilisant la plate forme de recherche libre service Terrier² [169]. Pour chaque document nous gardons aussi un index vers le concept ainsi que la localisation sous lesquels il est classé dans l'ontologie ODP. Cette information est exploitée comme une source d'évidence pour assigner les jugements de pertinence, comme il sera décrit ci-dessous.

5.1.2 Contextes utilisateurs

- **Les centres d'intérêts** : nous avons simulé 30 profils utilisateur. Par souci de simplification, nous avons affecté un centre d'intérêt pour chaque utilisateur, i.e. $k = 1$ et $sw(c_1) = 1$ dans la formule 5.5. Les centres d'intérêts sont simulés en choisissant aléatoirement 30 concepts de l'onto-

²<http://terrier.org/>

logie ODP, ils sont présentés dans la table 5.1. Plus précisément, chaque centre d'intérêt est représenté comme un concept de l'ontologie ODP en utilisant un ensemble de documents de (P) classés sous ce concept, comme décrit dans la section 4.2.

Concept_id	Concept
1	Health/Mental Health
2	Travel and Tourism/Lodging/Hotels and Motels
3	Arts and Entertainment/Artists and Galleries
4	Business and Economy/Automotive/Sales
5	Society and Culture/Religion
6	Business and Economy/Animals
7	Business and Economy/Restaurants and Bars
8	Business and Economy/Beauty and Cosmetic Services
9	Business and Economy/Shopping/Books
10	Business and Economy/Shopping/Jewelry
11	Business and Economy/Shopping/Clothing
12	Arts and Entertainment/Libraries
13	Arts and Entertainment/Museums
14	Arts and Entertainment/Theater
15	Arts and Entertainment/Events
16	Travel and Tourism/Parks
17	Travel and Tourism/Camping and RVs
19	Recreation and Sports/Summer Camps
20	Recreation and Sports/Golf
21	Recreation and Sports/Fishing
22	Society and Culture/Clubs and Lodges
23	Society and Culture/Organizations
24	Arts and Entertainment/Dance
25	Education/School Districts
26	Education/Colleges and Universities
27	News and Media/Newspapers
28	News and Media/Radio
29	Health/Dentistry
30	Health/Hospitals

TAB. 5.1 – La liste des 30 concepts testés représentant les centres d'intérêts des utilisateurs.

- **Les localisations** : les localisations assignées aux utilisateurs sont choisies de façon aléatoire parmi les villes des USA. Nous supposons une localisation différente pour chaque utilisateur. Elles sont simplement ajoutées aux requêtes des utilisateurs.

5.1.3 Collection de requêtes

Pour simuler l'activité de recherche des utilisateurs, nous avons conçu un ensemble de 6 requêtes pour chaque utilisateur. Elles sont formulées selon plusieurs stratégies.

Plus spécifiquement, chaque requête sera formulée dans l'une des stratégies suivantes :

1. un ensemble de termes décrivant un besoin en information particulier de l'utilisateur autour du concept représentant son centre d'intérêt ;
2. le terme le plus fréquent du concept représentant son centre d'intérêt ;
3. les deux termes les plus fréquents du concept représentant son centre d'intérêt ;
4. les trois termes les plus fréquents du concept représentant son centre d'intérêt ;
5. deux termes (ou plus) communs, parmi les 10 termes les plus pondérés, entre le concept représentant son centre d'intérêt et d'autres concepts ;
6. un besoin en information exprimé avec les termes d'un autre concept différent du centre d'intérêt actuel de l'utilisateur.

Notre but derrière ces diverses stratégies est double : (1) couvrir le cas des requêtes courtes et ambiguës, et (2) couvrir aussi bien des situations où l'utilisateur exprime des requêtes en ligne avec ces centres d'intérêts que des requêtes formulées sur de nouveaux besoins (pas encore présents dans son profil).

Au final, nous avons obtenu un ensemble de 180 requêtes ($30 * 6$) (présentées en totalité dans l'appendice B). La table 5.2 donne un exemple de requêtes construites selon chaque stratégie. Les cinq premières requêtes sont exprimées en utilisant des termes représentatifs du concept "*Travel and Tourism/Lodging/Hotels and Motels*", la dernière requête avec les termes du concept "*Travel and Tourism/Parks*".

Stratégie	Les termes de la requête
1	"book room spa wifi"
2	"hotel"
3	"hotel service"
4	"hotel service chain"
5	"location reserve resort"
6	"park activity"

TAB. 5.2 – Un exemple de requêtes formulées selon chaque stratégie.

5.1.4 Scénario d'évaluation et jugements de pertinence

L'ordre de priorité (\succ) que peut choisir un utilisateur sur les trois critères de pertinence (thème, intérêts et localisation), nous permet de définir six scénarios d'évaluation possible :

- **TIL** : Thème \succ Intérêts \succ Localisation,
- **TLI** : Thème \succ Localisation \succ Intérêts,
- **ITL** : Intérêts \succ Thème \succ Localisation,
- **ILT** : Intérêts \succ Localisation \succ Thème,
- **LTI** : Localisation \succ Thème \succ Intérêts,
- **LIT** : Localisation \succ Intérêts \succ Thème.

Chaque paire (requête, document) contenue dans les listes des résultats de la recherche sera jugée en fonction de chaque scénario d'évaluation. Les jugements de pertinence sont assignés en deux phases : d'abord une pertinence est assignée selon chaque critère à part, puis un jugement final est effectué par combinaison des pertinences individuelles.

1. **L'attribution individuelle des jugements de pertinence** : est effectuée automatiquement pour chaque critère, en exploitant comme source d'évidence les localisations et les concepts sous lesquels les documents sont initialement classés dans l'ontologie.

L'attribution individuelle des jugements de pertinence selon chaque critère est effectuée comme suit :

- si un document est classé sous un concept qui correspond au thème de la requête, il est jugé pertinent du point de vue thématique, sinon il est jugé non pertinent pour ce critère.
- si un document est classé sous un concept qui correspond au centre d'intérêt actuel de l'utilisateur, il est jugé pertinent pour le profil utilisateur, sinon il est jugé non pertinent pour ce critère.
- si un document est classé sous la localisation de l'utilisateur, il est jugé pertinent pour la localisation de l'utilisateur, sinon il est jugé non pertinent pour ce critère.

2. **Le jugement de pertinence finale** : est effectué en combinant les pertinences individuelles selon une échelle de pertinence à 4 niveaux : très pertinent, pertinent, peu pertinent et pas pertinent. L'algorithme 2 suivant illustre la démarche d'attribution du jugement de pertinence selon le degré de satisfaction des critères prioritaires.

Algorithme 2 Attribution automatique de jugements de pertinence selon le degré de satisfaction des critères prioritaires

si le document ne satisfait pas le critère de premier degré **alors**
- le document n'est pas pertinent.
sinon
si le document ne satisfait pas le critère du second degré **alors**
- le document est peu pertinent.
sinon
si le document ne satisfait pas le critère de troisième degré **alors**
- le document est pertinent.
sinon
- le document est très pertinent.
finsi
finsi
finsi

5.1.5 Métriques d'évaluation

Pour estimer la qualité des listes de résultats produites selon les différentes approches, nous avons utilisé les mesures standards de la MAP, nDCG@X et Précision@X. Pour les mesurer nous avons utilisé le programme standard *trec_eval*³. Les valeurs calculées sont moyennées sur l'ensemble des résultats de toutes les requêtes de test et/ou selon chacun des scénarios d'évaluation.

5.2 Résultats et discussion

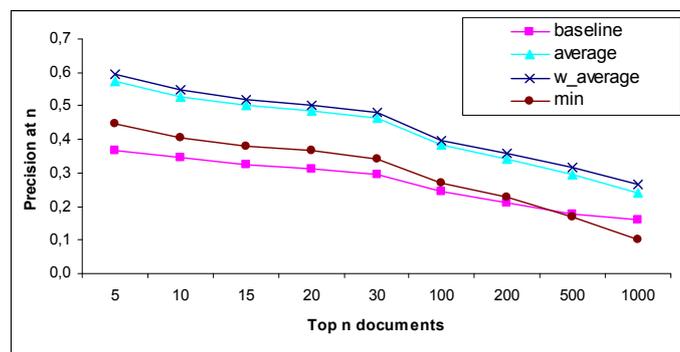
Nous avons évalué l'efficacité du modèle de combinaison en utilisant les scénarios d'évaluation présentés dans la section 5.1.4. Plus précisément, nos objectifs d'évaluation à ce niveau consiste à :

- Évaluer en un premier temps l'efficacité de la combinaison de la pertinence multi-dimensionnelle comparativement à un système classique à pertinence thématique,
- Évaluer en un second temps l'efficacité des opérateurs d'agrégation prioritaire par rapport à des opérateurs de combinaison linéaire standard.

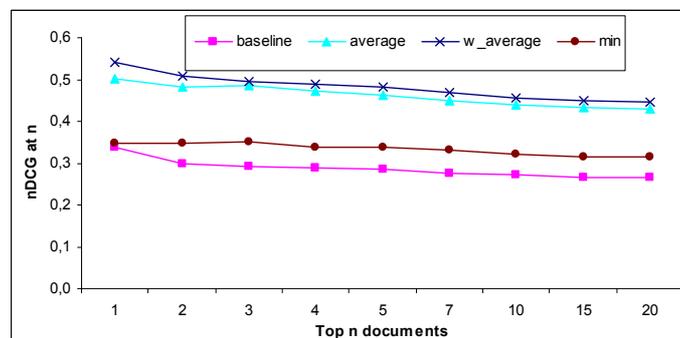
³http://trec.nist.gov/trec_eval/

5.2.1 Évaluation de l'efficacité de la combinaison de la pertinence multi-dimensionnelle

Dans cette première série d'expérimentations, nous effectuons pour chaque requête, une recherche standard *baseline* en utilisant le système de recherche Terrier [169] ; nous calculons ensuite un score personnalisé (selon les centres d'intérêts) et un score basé localisation, pour tous les documents dans notre collection, comme décrit dans la section 4.2 respectivement dans la section 4.3. Pour pouvoir combiner ces scores individuels, nous avons d'abord normalisé chacun d'eux en divisant sur le score maximum. Nous avons testé les principaux opérateurs de combinaison standard, notamment : le *min*, la moyenne "*average*" et la moyenne pondérée "*weighted average*" (noté *w-average*) où nous avons affecté des poids pour les critères de pertinence, en adéquation avec leur degré de priorité dans le scénario d'évaluation. Nous avons conduit des tests avec les différents scénarios d'évaluation présentés ci-dessus. Les figures 5.2(a) et 5.2(b) illustrent les résultats de performance mesurée respectivement selon la précision et nDCG à différents points de coupure, moyennée sur l'ensemble des scénarios d'évaluation.



(a)



(b)

FIG. 5.2 – Performance en termes de (a) précision, (b) nDCG à n moyennée sur l'ensemble des scénarios d'évaluation.

Nous pouvons constater, qu'en général, la combinaison des critères de per-

tinence améliore la performance par rapport à la *baseline*. En particulier, la méthode de combinaison linéaire pondérée atteint les meilleures performances en termes de nDCG et de précision à différents points de coupure. Nous pouvons conclure alors en faveur de l'efficacité du modèle de combinaison de la pertinence multi-critères et valider une hypothèse à la base de notre travail : combiner les trois dimensions de pertinence contextuelle améliore la pertinence des résultats par rapport à la dimension thématique seule. Des conclusions similaires sont obtenues dans des travaux de la RI géographique notamment dans Palacio et al. [170] qui combinent les dimensions spatiale, temporelle et thématique.

5.2.2 Évaluation de l'efficacité des opérateurs d'agrégation prioritaire

Dans cette seconde série d'expérimentations, nous avons combiné nos trois critères de pertinence en utilisant les opérateurs d'agrégation prioritaire “*scoring*” et “*and*”. Puis, nous avons comparé les listes des résultats de recherche obtenues avec la liste obtenue par l'opérateur standard “*weighted average*” comme notre nouvelle *baseline*. Les figures 5.2.2 et 5.2.2 montrent les performances de recherche respectivement en termes de précision et nDCG à différents points de coupure, moyennées sur l'ensemble des scénarios d'évaluation.

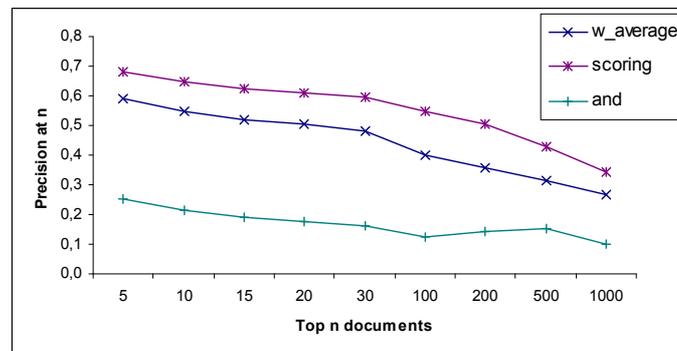


FIG. 5.3 – Performance en termes de précision à n moyennée sur l'ensemble des scénarios d'évaluation.

Nous pouvons constater que l'opérateur d'agrégation prioritaire “*scoring*” améliore les performances de recherche par rapport à son homologue standard “*weighted average*”.

Cependant, l'opérateur d'agrégation prioritaire “*and*” dégrade les performances de recherche. Ceci peut être expliqué par le fait que la logique d'agrégation de cet opérateur ne convient pas à nos scénarios d'évaluation. Par ailleurs, ces résultats rejoignent les conclusions dans [146] selon lesquelles la combinaison par l'opérateur logique “*min*” donne de faible performance comparativement à un opérateur faisant la somme des pertinences individuelles.

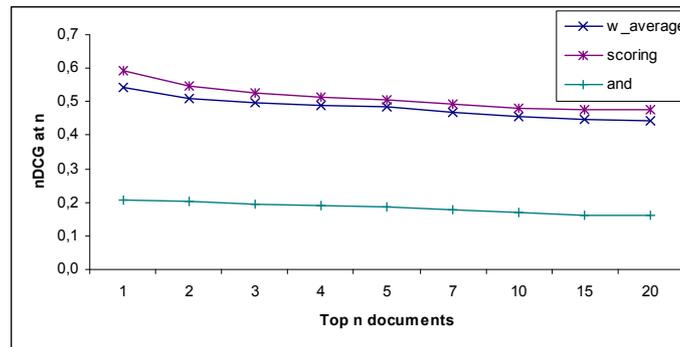


FIG. 5.4 – Performance en termes de nDCG à n moyennée sur l’ensemble des scénarios d’évaluation.

Afin d’évaluer si les différences observées, entre l’opérateur d’agrégation prioritaire “*scoring*” et son homologue “*weighted average*”, sont significatives d’un point de vue statistique, nous avons mené un t-test sur les moyennes de la MAP obtenues entre les différentes approches. La table 5.3, illustre les performances de recherche en termes de la MAP calculées pour chaque scénario d’évaluation.

Scénario	W-average	Scoring	Gain
TIL	0.0419	0.0478	14.08% (*)
TLI	0.0316	0.0413	30.70% (*)
ITL	0.0378	0.1150	204.23% (*)
ILT	0.0265	0.1086	309.81% (*)
LTI	0.8255	0.8312	0.69% (*)
LIT	0.8315	0.8315	0.00% (-)

TAB. 5.3 – Comparaison de la MAP calculée sur les résultats de recherche des opérateurs “*scoring*” et “*weighted average*” pour chaque scénario d’évaluation.

Les résultats confirment que le gain de performance enregistré par l’opérateur “*scoring*” par comparaison à la *baseline* est statistiquement significatif (noté * dans la table) avec des p -valeurs < 0.01 pour la majorité des scénarios d’évaluation. Cependant, le gain de performance est différent d’un scénario à l’autre. En particulier, nous notons que la meilleure amélioration de performance est observée sur le scénario où le profil de l’utilisateur est le critère d’ordre premier. En opposé, peu et pas d’améliorations sont observées sur les MAP des deux scénarios LTI respectivement LIT où la localisation est le critère le plus important. Par ailleurs, nous notons que les MAP des deux scénarios LTI et LIT sont quelque peu meilleures que les autres scénarios. Cela est probablement dû au fait que le score du critère de localisation est calculé avec une meilleure précision, que le score basé thématique et centre d’intérêt.

Nous avons de plus analysé les performances de recherche pour chaque type de requête issu de nos différentes stratégies de construction. La table 5.4

compare la moyenne de la MAP à travers les six scénarios d'évaluation pour chaque stratégie, calculée sur la liste des résultats de recherche retournée par l'opérateur "*scoring*" et l'opérateur "*weighted average*".

Stratégie	W-average	Scoring	Gain
1	0.2915	0.3207	10.02%(*)
2	0.3033	0.3305	8.96%(*)
3	0.3113	0.3387	8.83%(*)
4	0.3152	0.3437	9.04%(*)
5	0.2822	0.3141	11.31%(*)
6	0.2912	0.3277	12.52%(*)

TAB. 5.4 – Comparaison de la MAP calculée sur les résultats de recherche des opérateurs "*scoring*" et "*weighted average*" pour chaque stratégie de formulation des requêtes.

Comme on peut le constater sur cette table, l'opérateur "*scoring*" présente un gain de performance selon toutes les stratégies de formulation des requêtes. Ce gain est significatif (noté * dans la table) avec une p -valeur < 0.01 par comparaison à l'opérateur de combinaison linéaire standard "*weighted average*" à travers les six stratégies de formulation de requêtes. Particulièrement, dans les stratégies 5 et 6, où les requêtes sont exprimées par des termes communs à plusieurs concepts ou des termes non présents dans le profil actuel de l'utilisateur, le potentiel d'amélioration de la performance de recherche est important.

En fin, nous pouvons conclure en faveur de l'efficacité de l'opérateur d'agrégation "*scoring*" pour la combinaison de nos critères de pertinence thématique, centres d'intérêts et localisation.

6 Conclusion

Nous avons présenté au cours de ce chapitre un modèle de combinaison de pertinences multi-critères pour un accès contextuel à l'information. Nous avons instancié ce modèle dans un environnement mobile, en considérant deux facteurs contextuels importants pour l'utilisateur mobile, en particulier ses centres d'intérêts et sa localisation.

L'avantage majeur de cette méthode d'agrégation est que l'utilisateur est libre de définir un ordre de priorité sur les dimensions de pertinence pour chaque requête. De cette façon, le score de pertinence finale d'un document dans la liste des résultats de la recherche est dépendant des préférences de l'utilisateur exprimées sur les critères d'agrégation.

Nous avons ensuite présenté un cadre d'évaluation orienté-utilisateur pour l'évaluation de notre modèle de combinaison de pertinence multi-dimensionnel. Les résultats expérimentaux montrent en général l'efficacité de notre modèle de combinaison, et en particulier :

1. la combinaison des dimensions de pertinences améliore la performance des résultats de la recherche comparativement à un système de recherche standard basé uniquement sur la pertinence thématique ;
2. le modèle d'agrégation prioritaire basé sur l'opérateur "*scoring*" permet d'avoir un gain de performance significatif de la recherche comparativement à la technique classique de combinaison linéaire sur l'ensemble des scénarios d'évaluation et les stratégies de formulation de requêtes testés ;
3. l'opérateur d'agrégation prioritaire "*and*" n'est pas adapté pour l'agrégation de nos critères de pertinence dans notre cadre de recherche mobile.

Conclusion générale

Synthèse

Les travaux présentés dans ce mémoire se situent dans le contexte général de l'accès contextuel à l'information et plus particulièrement dans le cadre d'un environnement mobile. Nos contributions présentées dans ce mémoire ont porté sur quatre volets : la modélisation du profil situationnel de l'utilisateur, l'identification du type de besoin derrière la requête, l'agrégation prioritaire de pertinences multidimensionnelles et l'évaluation du processus d'accès contextuel à l'information.

1. Notre méthode de personnalisation se base sur la conception d'un profil flexible de l'utilisateur qui varie selon le changement de sa situation spatio-temporelle. La modélisation du profil utilisateur est basée sur l'exploitation de son contexte spatio-temporel et de son historique de recherche et des ontologies de domaines Web, temporelle et spatiale permettant de dégager une représentation sémantique du contexte spatio-temporel de recherche et des centres d'intérêts selon un graphe de concepts. Chaque situation de recherche est modélisée sur 4 dimensions sémantiques de l'environnement de l'utilisateur. La première dimension correspond au type de la localisation, la seconde dimension correspond à la saison, la troisième dimension correspond au jour de la semaine et la quatrième au moment de la journée. Les valeurs de chaque dimension sont issues de l'abstraction des données de bas niveau en terme de coordonnées GPS et de l'heure de et de la date. A chaque situation sont associés les centres d'intérêt de l'utilisateur modélisés comme des concepts extraits d'une ontologie thématique. Cela consiste à collecter les documents jugés pertinents implicitement pour un ensemble de requêtes liées à une même situation de recherche, et ensuite les exploiter pour extraire les concepts les plus pertinents issus de l'ontologie. Une approche de raisonnement à partir de cas (RàpC) est utilisée pour adapter le processus de personnalisation à la flexibilité du profil utilisateur. La phase de sélection du profil est basée sur une mesure de similarité des situations. La ré-utilisation du profil utilisateur est basée sur une technique de ré-ordonnement des résultats de recherche. La phase de révision est basée sur le *feedback*

utilisateur qui permet de mettre à jour la base des cas. La mise à jour consiste en l'évolution du profil situationnel qui est basée sur l'augmentation de l'importance des concepts récurrents dans une même situation de recherche.

2. Le type de besoin derrière la requête pour un utilisateur mobile peut être global, local explicite ou local implicite. Notre approche pour l'identification du type de la requête est basée sur la définition d'un modèle de langue de localisation de la requête à partir des " n " premiers résultats de recherche retournés par un moteur de recherche web général. Ce modèle présente l'avantage de formaliser la notion de sensibilité de la requête à la localisation. De plus, il nous offre une source d'évidence pour calculer deux caractéristiques de la requête, notamment les mesures de divergence de *Kullback-Leibler* et de *Kurtosis*. Ces deux mesures sont ensuite utilisées comme attributs de classification pour la prédiction du type de la requête : global, local explicite et local implicite. Nous avons proposé deux techniques d'adaptation de la recherche par reformulation puis par ré-ordonnancement, en vue d'intégrer les résultats de la classification des requêtes selon leur sensibilité à la localisation, pour adapter les résultats de recherche d'un moteur web général.
3. Notre contribution pour la proposition d'un modèle de pertinence multidimensionnelle est basée sur la combinaison de la pertinence thématique (traditionnellement à la base de la pertinence calculée par les SRI standards) avec des facteurs de pertinence contextuelle notamment : les centres d'intérêts et la localisation de l'utilisateur mobile. Ce modèle de pertinence multidimensionnelle présente la particularité d'exploiter deux opérateurs d'agrégation prioritaire permettant d'adapter les résultats de la recherche selon les préférences de l'utilisateur exprimées sur les critères de pertinence. Dans ce modèle, le poids d'un critère dans la combinaison est dépendant du degré de satisfaction du critère de priorité supérieure. Ainsi, pour une même requête et un même document, plusieurs scores de pertinence peuvent être calculés en fonction de l'ordre de priorité exprimé par un utilisateur sur les différents critères de pertinence.
4. Le dernier volet de notre contribution consiste en la définition de cadre d'évaluation appropriés pour l'accès contextuel à l'information.

Nous avons proposé deux cadres d'évaluation orientés-contexte pour l'évaluation de notre approche de personnalisation. Le premier cadre est basé sur simulation du contexte et de l'activité de recherche. Ce cadre d'évaluation a permis de faire une première évaluation rapide de l'efficacité de la recherche personnalisée en comparaison à une recherche standard. Les résultats des expérimentations ont démontré l'efficacité de notre approche de personnalisation pour des requêtes courtes et ambiguës pouvant être exprimées dans divers contextes simulés. Le gain de performance de la recherche personnalisée étant variable entre les requêtes, cette variabilité dépend principalement du degré de spécificité des requêtes mais aussi du

contenu des documents de la collection de test.

Le second cadre d'évaluation est basé sur une "étude journalière" (*diary study*) pour évaluer l'efficacité de notre approche de personnalisation en utilisant des requêtes et des contextes de recherche réels. Ce nouveau cadre nous a permis de tester et de comparer deux protocoles d'évaluation : par ordre chronologique des requêtes et selon la méthode de la cross-validation. Les résultats expérimentaux montrent bien l'efficacité des différentes composantes de notre approche, notamment la cohérence des résultats obtenus selon chacun des protocoles d'évaluation testés, la précision du mécanisme de sélection de la situation de recherche la plus adéquate et un gain de performance significatif de la recherche personnalisée comparativement à la recherche classique sur l'ensemble des requêtes de test selon chacun des protocoles d'évaluation.

Nous avons validé notre approche de détection de la sensibilité de la requête à la localisation de l'utilisateur sur une collection de requêtes annotées manuellement issue du *log* de recherche d'AOL, en testant plusieurs classificateurs du domaine et par comparaison à une approche de l'état de l'art. Nous avons intégré les résultats de la classification dans une approche d'adaptation des résultats de recherche selon le type de la requête. Les jugements de pertinence sont construits en considérant le type des requêtes. Nous avons montré son efficacité à améliorer la performance de la recherche par comparaison à un SRI standard.

Nous avons également proposé un cadre d'évaluation centré-utilisateur pour tester l'efficacité des opérateurs d'agrégation prioritaire comparativement à l'approche standard par combinaison linéaire. Il est basé sur une évaluation par simulation basée sur des évidences issues de la structuration de l'ontologie du Web ODP. Les résultats des expérimentations montrent que la combinaison des dimensions de pertinences améliore la performance des résultats de la recherche comparativement à un système de recherche standard basé uniquement sur la pertinence thématique et comparativement à la technique classique de combinaison linéaire.

Limites et perspectives

Les évaluations expérimentales de nos contributions ont montré leurs efficacités selon plusieurs aspects, et ouvrent des perspectives portant sur l'amélioration de nos modèles de représentation et exploitation du contexte d'une part et l'évaluation du modèle d'accès contextuel à l'information dans un scénario de recherche réaliste d'autre part.

Plus particulièrement, nos perspectives portent essentiellement sur les points suivants :

1. Modélisation du profil situationnel de l'utilisateur :
 - (a) Lors de la mise à jour des centres d'intérêts des utilisateurs, nous avons favorisé la spécificité sur la généralité. En d'autres termes nous avons choisi d'ajouter un nouveau profil pour chaque nouvelle situation de recherche. Certains centres d'intérêts de l'utilisateur peuvent être communs dans des situations généralisées sur une ou plusieurs dimensions du modèle de la situation. Dans ce sens, notre perspective consiste à mettre en place un mécanisme de généralisation des profils situationnels. Cela permettra de comparer les centres d'intérêts à travers des situations de recherche différentes pour essayer de dégager des centres d'intérêts communs à une ou plusieurs situations ou des centres d'intérêts indépendants de la situation de l'utilisateur.
 - (b) Enrichir le modèle des situations par d'autres dimensions contextuelles. Une piste de recherche dans ce sens est d'inclure le contexte social défini par les personnes à côté de l'utilisateur ou appartenant au même groupe. Cela permet d'un côté d'affiner la pertinence des résultats selon le contexte social actuel de l'utilisateur, et d'un autre côté d'inférer la pertinence des résultats selon le profil social dans le cas où un contexte spatio-temporel actuel est nouveau pour l'utilisateur.
 - (c) Lors de l'évaluation de l'efficacité de notre approche de personnalisation, nous avons montré son efficacité vis-à-vis d'une approche de RI standard n'exploitant pas de profils utilisateurs. Il faut étendre l'évaluation pour comparer l'efficacité de notre approche de personnalisation comparativement à d'autres approches de personnalisation présentées dans le domaine.
 - (d) Lors de l'évaluation de l'efficacité de notre approche de personnalisation nous avons remarqué une légère dégradation des performances sur des requêtes très spécifiques. Une solution à ce problème est de proposer de contrôler le processus de personnalisation par un mécanisme de détection ou de calcul d'un degré de spécificité/ambiguïté des requêtes.
 - (e) Lors de notre expérimentation, nous étions confrontés à deux cas présentant un problème de multi-langues. Dans le premier cas un utilisateur a jugé pertinent des documents retournés pour sa requête en deux langues différentes (anglais et français). Dans le second cas un autre utilisateur a soumis des requêtes en français puis en anglais relatives à un même centre d'intérêt. En effet, les processus de construction du profil et de sa mise à jours ne permettent pas la comparaison et l'intégration de termes ou de concepts de langues différentes. Une perspective dans ce sens est d'exploiter les liens de

langue dans la structure de l'ODP pour essayer de relier les concepts multi-langues similaires.

2. Prédiction du type de besoin :
 - (a) Notre modèle de détection de la sensibilité de la requête à la localisation est basé sur des caractéristiques issues uniquement des "n" premiers résultats de recherche retournés par un moteur web. Une perspective de recherche est d'enrichir ce modèle par des caractéristiques issues du *log* de recherche pour améliorer la pertinence de classification.
 - (b) Le modèle de détection de la sensibilité de la requête à la localisation a été testé sur un échantillon relativement petit (200 requêtes). Notre perspective à court terme dans ce sens consiste à élargir la collection de teste en continuant l'effort d'annotation manuelle des requêtes.
 - (c) Nous avons proposé un modèle de détection de la sensibilité des requêtes mobiles à la localisation de l'utilisateur. Une perspective de recherche à long terme dans le même sens consiste à proposer un modèle pour la détection de la sensibilité de la requête au temps.
3. Agrégation de pertinences multidimensionnelles :
 - (a) Nous avons testé le modèle d'agrégation prioritaire pour la combinaison d'uniquement deux dimensions contextuelles en plus de la dimension thématique. Notre perspective dans ce sens porte sur l'amélioration de notre modèle multidimensionnel pour inclure d'autres critères contextuels, notamment les dimensions temporelle et sociale.
 - (b) L'évaluation expérimentale a montré l'efficacité du modèle d'agrégation prioritaire comparativement à un système classique basé uniquement thématique ainsi qu'à un système de combinaison linéaire standard. Une particularité de ce modèle est que l'utilisateur est sollicité pour spécifier pour chaque requête son ordre de préférence sur les critères de pertinence. Une perspective de recherche dans ce sens est de proposer d'alléger cette contrainte en apprenant un ordre de préférence sur les interaction passées de l'utilisateur portant sur des requêtes similaires.
4. Évaluation du modèle d'accès contextuel à l'information :
 - (a) à court terme, notre perspective porte sur l'évaluation de l'efficacité de recherche contextuelle dans un contexte de recherche réel en utilisant des données de vrais utilisateurs issus des fichiers *logs* d'un moteur de recherche. Ceci permet d'évaluer nos approches avec un nombre plus important d'utilisateurs, de leurs vrais interactions et contextes.
 - (b) à long terme, nous envisageons à mettre en place un cadre d'évaluation standard pour la RI contextuelle où l'évaluation sera vigoureusement contrôlée en utilisant une collection de test réelle dont les requêtes

sont émises par des vrais utilisateurs, les requêtes seront en plus annotées par des données contextuelles (temps, localisation, etc), les interactions des utilisateurs sont exploitables pour apprendre des profils réels.

Annexe A

Collection de requêtes utilisées pour la validation de notre approche de classification

Le tableau [A.1](#) présente l'ensemble des requêtes utilisées pour le test de la performance de notre approche d'identification de la sensibilité de la requête à la localisation présentée dans le chapitre [4](#).

TAB. A.1: Collection de requêtes extraites du *log* d'AOL.

Numéro requête	Termes de la requête
1	train ticket
2	eye chart
3	inuyasha
4	national parks
5	popular careers
6	Bronx Zoo
7	weather
8	jobs
9	white pages
10	the wispy look
11	panera bread
12	lime for garden
13	pizza
14	movies times
15	The new buss federation
16	where to get flying lessons
17	macromedia flash player

Continued on next page

Tab. A.1 – continued from previous page

Numéro requête	Termes de la requête
18	printable party invitations
19	hunting and fishing regulations
20	the current newspaper
21	ringworm
22	starbucks store
23	ipod questions
24	funny pictures
25	state park jobs
26	tattous
27	vehicle registration
28	department of education
29	art museums
30	music
31	zigzone
32	state trooper
33	amusement parks
34	animal shelter
35	city bus schedules
36	post office
37	alaskan cruises
38	airport shuttle
39	las vegas special events contacts
40	hippie clothing
41	air travel information
42	area codes
43	home depot
44	bank of america
45	circuit city
46	movies times
47	space needle
48	st. christopher school west covina
49	apartments on michigan avenue chicago
50	find atm machine
51	indianapolis zoo
52	free readings
53	t2 judgement day
54	fermilab tours
55	Lincoln Memorial
56	humorous jokes
57	lottery
58	Giant Sequoia National Monument
59	dog heat

Continued on next page

Tab. A.1 – continued from previous page

Numéro requête	Termes de la requête
60	flame designs
61	indexed annuity
62	solar panels
63	the secret garden
64	elliptical trainer
65	website design hosting
66	rent car
67	lower heart rate
68	dictionary
69	ringtones
70	jokes
71	engineering college
72	ps 2 games
73	dinosaur
74	sentence composing kilgallon
75	nancy drew walkthrough
76	Virtue of the verbena tea
77	color book piglet
78	stories on the web
79	free online games nintendo
80	ms medications
81	yellow pages
82	outer space decor
83	game and fish
84	magnetism experiments
85	black holes
86	board of corrections
87	southpark cartman
88	birthday ecards
89	art council
90	winzip
91	public library
92	medwatch
93	recipes for kids
94	scrabble blast
95	tabular latex
96	what is an essay generalization
97	free avatars
98	looking to hire a DJ
99	health, nicotine
100	bing bang
101	campgrounds in the mountains of n.c.

Continued on next page

Tab. A.1 – continued from previous page

Numéro requête	Termes de la requête
102	mercury in aquarius
103	annual parades
104	work ethics attendance
105	cna training
106	ny department stores
107	virginia beach government
108	johnny pacar pics
109	is there a way to waterproof ceramics
110	motorcycle rides
111	radio message board
112	final fantasy
113	bicornuate uterus
114	cake decorating images
115	hotels in center texas
116	large silk flowers
117	soccer clubs
118	rental property in springfield ohio
119	road conditions brevard county florida
120	androgen hormone
121	martinsburg correctional center
122	restaurants
123	cinderella
124	verizon internet service
125	epson printer driver
126	window security virus patches
127	mrsa
128	server error
129	newspapers in burns tn
130	car insurance
131	anne arundel county yellow pages
132	what foods can my 12 month old eat
133	school district
134	rave movie theater
135	flower delivery new rochelle
136	cursive alphabet
137	13th letter hebrew alphabet
138	office skills
139	mirgraines and hormones
140	claremont college music
141	nj lottery
142	space wall mural
143	solar 20system

Continued on next page

Tab. A.1 – continued from previous page

Numéro requête	Termes de la requête
144	poetry by yusef komunyakaa
145	need speed most wanted
146	modem cannot find dsl in
147	jefferson city jobs
148	canon digital rebel flash information
149	honda dealer
150	examples of newton's three laws of motion
151	ballads of rock
152	ohio family getaways
153	christening gowns
154	infant clothing stores albany ny
155	military grade ranks
156	medical dictionary
157	unclaimed money indiana
158	neuropathy pain
159	pork roast cooking time
160	beauty and the beast lyrics
161	dogs for adoption
162	new jersey trailers
163	nano technology
164	mothers day music
165	tennessee newspapers
166	community college
167	oregon city school district
168	free grant to start a car wash
169	chip and dale disney characters
170	zoning holly hill florida
171	online food shopping
172	portable dvd player
173	riverton country club nj
174	what is spurs of the bone
175	how to clean your computer tower
176	dark and high elves fan art
177	graphic novel making tutorials
178	los angeles Taxi
179	medications for intermittent explosive disorder
180	white pages waterford michigan
181	florida lottery results
182	grey's anatomy
183	new job jitters
184	craigslist portland
185	arizona used mercedes parts

Continued on next page

Tab. A.1 – continued from previous page

Numéro requête	Termes de la requête
186	size of bow windows ohio
187	Statue of Liberty
188	desirable communities in milford pennsylvania
189	what is a quadrilateral
190	park tudor indianapolis
191	farmhouse decorating
192	chicago magazine
193	oriental area rugs
194	movable pen for fowl
195	auto trader
196	normal color stool
197	bowel movement consistency
198	currency exchange rate
199	claw tub faucets
200	marriage records cochise county arizona

Annexe B

Collection de requêtes utilisées pour la validation de notre approche d'agrégation multidimensionnelle

Le tableau B.1 présente l'ensemble des requêtes construites pour le test de la performance de notre approche d'agrégation prioritaire de la pertinence multidimensionnelle présentée dans le chapitre 5. La colonne 1 représente les stratégies de formulation des requêtes, telles que décrites dans la section 5.1.3 du chapitre 5. La colonne 2 représente l'id du concept représentant le centre d'intérêt de l'utilisateur et la colonne 3 l'id du concept lorsque la requête porte sur un centre d'intérêt autre que celui de l'utilisateur. Les concepts sont présentés dans la section 5.1.2 du chapitre 5. La dernière colonne représente les termes des requêtes.

TAB. B.1: Collection de requêtes simulées.

Stratégie	Concept1	Concept2	Place	Termes de la requête
1	1		Traverse City	headache drugs price
	2		San Francisco	book room spa wifi
	3		Seattle	paintings postmodern
	4		Houston	ford mustang motor
	5		Anchorage	mormon fellows
	6		Austin	animal testing alternatives
	7		Boston	pub irish
	8		Los Angeles	piercing tatoo
	9		Cambridge	economics bookstore
	10		Denver	engagement ring

Continued on next page

Tab. B.1 –continued from previous page

Stratégie	Concept1	Concept2	Place	Termes de la requête
	11		Cincinnati	baggy trousers men
	12		Bridgeport	position librarian
	13		Baltimore	current exhibitions photography
	14		Atlanta	action movie american
	15		Indianapolis	concert punk music
	16		Evergreen	nature ride picnic
	17		Saint Louis	tourism culture entertainment
	18		Wells	water supply hose
	19		Raymond	overnight camp mountains
	20		Ocean City	maryland courses seaside
	21		Homer	fish bowl crackers
	22		Columbia	rituals occult
	23		Greenville	non-profit social
	24		Cleveland	rock salsa lessons
	25		Little Rock	uniforms manufacturer
	26		San Diego	undergraduate lecturers
	27		Philadelphia	open positions journalist
	28		Salt Lake City	mormon worship service
	29		Virginia Beach	bad smell mouth
	30		Lexington	lung cancer ward
2	1		Traverse City	health
	2		San Francisco	hotel
	3		Seattle	art
	4		Houston	axl
	5		Anchorage	church
	6		Austin	fenc
	7		Boston	restaur
	8		Los Angeles	tan
	9		Cambridge	book
	10		Denver	jewelry
	11		Cincinnati	store
	12		Bridgeport	library
	13		Baltimore	museum
	14		Atlanta	perform
	15		Indianapolis	event
	16		Evergreen	park
	17		Saint Louis	histor
	18		Wells	campground
	19		Raymond	camp
	20		Ocean City	golf
	21		Homer	fish

Continued on next page

Tab. B.1 –continued from previous page

Stratégie	Concept1	Concept2	Place	Termes de la requête
	22		Columbia	lodge
	23		Greenville	organiz
	24		Cleveland	danc
	25		Little Rock	school
	26		San Diego	univers
	27		Philadelphia	newspap
	28		Salt Lake City	radio
	29		Virginia Beach	dental
	30		Lexington	medic
3	1		Traverse City	health mental
	2		San Francisco	hotel servic
	3		Seattle	art gallery
	4		Houston	axl lincoln
	5		Anchorage	chuch dioces
	6		Austin	fenc dog
	7		Boston	restaur guide
	8		Los Angeles	tan list
	9		Cambridge	book stock
	10		Denver	jewelry jewel
	11		Cincinnati	store locat
	12		Bridgeport	library collect
	13		Baltimore	museum history
	14		Atlanta	perform theatr
	15		Indianapolis	event festiv
	16		Evergreen	park activity
	17		Saint Louis	histor event
	18		Wells	campground camp
	19		Raymond	camp summer
	20		Ocean City	golf cours
	21		Homer	fish informat
	22		Columbia	lodge organiz
	23		Greenville	organiz membership
	24		Cleveland	danc folk
	25		Little Rock	school district
	26		San Diego	univers colleg
	27		Philadelphia	newspap daily
	28		Salt Lake City	radio guid
	29		Virginia Beach	dental list
	30		Lexington	medic informat
4	1		Traverse City	health mental servic

Continued on next page

Tab. B.1 –continued from previous page

Stratégie	Concept1	Concept2	Place	Termes de la requête
	2		San Francisco	hotel servic chain
	3		Seattle	art gallery paint
	4		Houston	axl lincoln exchang
	5		Anchorage	chuch dioces eckankar
	6		Austin	fenc dog livestock
	7		Boston	restaur guide informat
	8		Los Angeles	tan list salon
	9		Cambridge	book stock store
	10		Denver	jewelry jewel diamond
	11		Cincinnati	store locat women
	12		Bridgeport	library collect univers
	13		Baltimore	museum history locat
	14		Atlanta	perform theatr play
	15		Indianapolis	event festiv list
	16		Evergreen	park activity area
	17		Saint Louis	histor event tour
	18		Wells	campground camp rv
	19		Raymond	camp summer year
	20		Ocean City	golf cours list
	21		Homer	fish informat guide
	22		Columbia	lodg organiz grand
	23		Greenville	organiz membership event
	24		Cleveland	danc folk present
	25		Little Rock	school district calendar
	26		San Diego	univers colleg program
	27		Philadelphia	newspap daily articl
	28		Salt Lake City	radio guid station
	29		Virginia Beach	dental list society
	30		Lexington	medic informat health
5	1		Traverse City	hospit staff research
	2		San Francisco	locat reserv resort
	3		Seattle	photograph exhibit
	4		Houston	radio servic guid
	5		Anchorage	servic meet weekly
	6		Austin	store sale dog
	7		Boston	meet young girl
	8		Los Angeles	institut care woman
	9		Cambridge	staff store offer
	10		Denver	fashion woman retail
	11		Cincinnati	kid school price uniform
	12		Bridgeport	articl weekly psycholog

Continued on next page

Tab. B.1 –continued from previous page

Stratégie	Concept1	Concept2	Place	Termes de la requête
	13		Baltimore	art photograph exhibit
	14		Atlanta	children schedul
	15		Indianapolis	exhibit kid entertain
	16		Evergreen	children animal camp
	17		Saint Louis	park activity recr
	18		Wells	fee site mount
	19		Raymond	kid educat danc
	20		Ocean City	tourna price
	21		Homer	area campground fee
	22		Columbia	peopl list membership
	23		Greenville	lodg masonry admission
	24		Cleveland	event perform children
	25		Little Rock	admission schedul children
	26		San Diego	psycholog library directory
	27		Philadelphia	mail staff board
	28		Salt Lake City	daily program health
	29		Virginia Beach	research artiel medic clinic
	30		Lexington	medic program educat fee
6	1	13	Traverse City	museum history locat
	2	16	San Francisco	park activity
	3	7	Seattle	restaur guide informat
	4	17	Houston	histor event
	5	11	Anchorage	store locat women
	6	30	Austin	medic informat
	7	29	Boston	dental list
	8	7	Los Angeles	restaur guide
	9	15	Cambridge	event festiv list
	10	9	Denver	book stock
	11	23	Cincinnati	organiz membership event
	12	2	Bridgeport	hotel servic
	13	21	Baltimore	fish informat guide
	14	1	Atlanta	health mental
	15	13	Indianapolis	museum history
	16	5	Evergreen	chuch dioces
	17	22	Saint Louis	lodg organiz grand
	18	15	Wells	event festiv
	19	6	Raymond	fenc dog livestock
	20	26	Ocean City	univers colleg
	21	21	Homer	fish informat
	22	18	Columbia	campground camp rv
	23	2	Greenville	hotel servic chain

Continued on next page

Tab. B.1 –continued from previous page

Stratégie	Concept1	Concept2	Place	Termes de la requête
	24	25	Cleveland	school district
	25	8	Little Rock	tan list salon
	26	14	San Diego	perform theatr play
	27	24	Philadelphia	danc folk present
	28	10	Salt Lake City	jewelry jewel
	29	14	Virginia Beach	perform theatr
	30	11	Lexington	store locat

Bibliographie personnelle

- [1] O. Boudighaghen and L. Tamine. Personnalisation flexible d'un processus de recherche d'information mobile. In *huitième édition de la Conférence francophone en Recherche d'Information et Applications (CORIA 2011)*, pages 65–72, Avignon, France, 2011. support électronique.
- [2] O. Boudighaghen, L. Tamine, and M. Boughanem. Dynamically personalizing search results for mobile users. In *Proceedings of the 8th International Conference on Flexible Query Answering Systems*, pages 99–110, Roskilde, Danemark, 2009. Springer-Verlag.
- [3] O. Boudighaghen, L. Tamine, and M. Boughanem. A diary study-based evaluation framework for mobile information retrieval. In P.-J. e. a. CHENG, editor, *proceedings of the 6th Asia Information Retrieval Society Conference (AIRS 2010)*, pages 389–398, Taipei, Taiwan, 2010. Springer-Verlag.
- [4] O. Boudighaghen and L. Tamine. Spatio-Temporal Based Personalization for Mobile Search. In C. Jouis, I. Biskri, J.-G. Ganascia, M. Roux, editor, *Next Generation Search Engines : Advanced Models for Information Retrieval*, IGI-Global, Hershey, PA, USA, 2011 (to appear).
- [5] O. Boudighaghen, L. Tamine, G. Pasi, G. Cabanac, M. Boughanem and C. da Costa Pereira. Prioritized Aggregation of Multiple Context Dimensions in Mobile IR. In M.V.M. Salem et al. (Eds.) *proceedings of the 7th Asia Information Retrieval Society Conference (AIRS 2011)*, LNCS Vol. 7097, pages 169–180, Springer, Heidelberg, 2011.
- [6] O. Boudighaghen, L. Tamine, and M. Boughanem. Context-aware user's interests for personalizing mobile search. In *proceedings of the 12th International Conference on Mobile Data Management*, Luleå, Sweden, 2011. IEEE Computer Society.
- [7] O. Boudighaghen, L. Tamine, and M. Boughanem. Inferring situation-based interests for mobile users : A case based reasoning approach. In *proceedings of the 26th Symposium On Applied Computing, SAC2011*, Taichung, Taiwan, 2011. ACM.
- [8] O. Boudighaghen, L. Tamine, and M. Boughanem. Personalizing mobile web search for location sensitive queries. In *proceedings of the 12th Inter-*

-
- national Conference on Mobile Data Management*, Luleå, Sweden, 2011. IEEE Computer Society.
- [9] O. Boudghaghen, L. Tamine, M. Daoud, and C. Laffaire. Contextual evaluation of mobile search. In M. M. L. T. Bich-lien Doan, Joemon Jose, editor, *proceedings of the 2nd International Workshop on Contextual Information Access, Seeking and Retrieval Evaluation (CIRSE) in conjunction with ECIR 2010*, Milton Keynes, 2010. support électronique.
- [10] O. Boudghaghen, L.-L. Tamine, and M. Boughanem. Vers la définition du contexte d'un utilisateur mobile de système de recherche d'information. In *UbiMOB'09 : 5èmes journées Francophones Mobilité et Ubiquité*, Université des Sciences et Technologies de Lille, 2009.

Bibliographie

- [11] A. Aamodt and E. Plaza. Case-based reasoning : Foundational issues, methodological variations, and system approaches. *AI Communications*, 7(1) :39–59, 1994.
- [12] G. Abowd, C. Atkeson, J. Hong, S. Long, R. Kooper, and M. Pinkerton. Cyberguide : A mobile context-aware tour guide. *Wireless Networks*, 3(5) :421–433, 1997.
- [13] M. Abramowitz and I. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. NBS Applied Mathematics Series 55, National Bureau of Standards, Washington, DC, 1964.
- [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] E. Al-Masri and Q. H. Mahmoud. A context-aware mobile service discovery and selection mechanism using artificial neural networks. In *ICEC'06 : Proceedings of the 8th international conference on Electronic commerce*, pages 594–598. ACM, 2006.
- [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] B. Allen. Information needs : a person-in-situation approach. In *Proceedings of an International Conference on Research in needs, seeking and use in different context*, pages 111–122, London, UK, 1997. Taylor Graham Publishing.
- [18] E. Amitay, N. Harel, R. Sivan, and A. Soffer. Web-a-where : Geotagging web content. In *Proceedings of the 27st annual International ACM SIGIR conference on Research and development in information retrieval*, pages 273–280. ACM, 2004.
- [19] T. Andreasen. An approach to knowledge-based query evaluation : Fuzzy databases. *Fuzzy Sets and Systems*, 140(1) :75–91, 2003.
- [20] M. Arias, J. M. Cantera, and J. Vegas. Context-based personalization for mobile web search. In *the 2nd Int'l Workshop on Personalized Access*,

- Profile Management, and Context Awareness : Databases*, pages 36–42, 2008.
- [21] K. August, M. H. Hansen, and E. Shriver. Mobile web searching. *Bell Labs Technical Journal*, 6(2) :84–98, 2002.
- [22] R. Baeza-Yates and R. A. Ribeiro-Neto. *Modern Information Retrieval*. New York : ACM Press ; Harlow England : Addison-Wesley, cop., 1999.
- [23] M. Bazire and P. Brézillon. Understanding context before using it. In *Modeling and Using Context*, pages 29–40, 2005.
- [24] 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.
- [25] N. Belkin. Knowledge elicitation using discourse analysis. *International Journal of Man-Machine studies*, 27 :127–144, 1987.
- [26] 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.
- [27] V. Bellotti, B. Begole, E. Chi, N. Ducheneaut, J. Fang, E. Isaacs, T. King, M. Newman, K. Partridge, B. Price, P. Rasmussen, M. Roberts, D. Schiano, and A. Walendowski. Activity-based serendipitous recommendations with the magitti mobile leisure guide. In *CHI 2008 Proceedings On the Move*, pages 98–107, 2008.
- [28] R. BIERIG. *Event and map content personalisation in a mobile and context-aware environment*. Phd thesis in computer science, School of Computing, Robert Gordon University, Available from : <http://openair.rgu.ac.uk>, April 2008.
- [29] N. Bila, J. Cao, R. Dinoff, T. Ho, R. Hull, B. Kumar, and P. Santos. Mobile user profile acquisition through network observables and explicit user queries. In *9th Int'l conference on Mobile Data Management*, pages 98–107, 2008.
- [30] D. Bilal. Children's use of the yahoooligans! web search engine : cognitive, physical, and affective behaviors on fact-based search tasks. *Journal of the American Society for Information Science*, 51(7) :646–665, 2000.
- [31] P. Borlund and P. Ingwersen. The development of a method for the evaluation of interactive information retrieval systems. *Journal of Documentation*, 53(3) :223–250, 1997.
- [32] 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.) : Proceedings of the 21st ACM SIGIR Int'l. Conference on Research and Development*, pages 324–331. ACM, 1998.
- [33] 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.

- [34] M. Boughanem and J. Savoy, editors. *Recherche d'information états des lieux et perspectives*. Hermès Science Publications, <http://www.editions-hermes.fr/>, 2008.
- [35] N. Bouvin, C. B.G., K. Gronbaek, and F. Hansen. Hycon : a framework for context-aware mobile hypermedia. *New review of hypermedia and multimedia*, 9 :59–88, 2003.
- [36] A. Broder. A taxonomy of web search. In *ACM SIGIR Forum*, pages 3–10. ACM, 2002.
- [37] T. Brondsted, H. L. Larsen, L. B. Larsen, B. Lindberg, D. Ortiz-Arroyo, Z.-H. Tan, and H. Xu. Mobile information access with spoken query answering. In *ASIDE2005, Workshop on Applied Spoken Language Interaction in Distributed Environments*, 2005.
- [38] B. Brookes. The developping cognitive viewpoint in information science. In *De Mey, M. and al Editors*, pages 195–203, septembre 1977.
- [39] P. Brown, J. Bovey, and C. Xian. Context-aware applications : from the laboratory to the marketplace. *IEEE Personal Communications*, 4(5) :58–64, 1997.
- [40] P. Brown and G. Jones. Context-aware retrieval : Exploring a new environment for information retrieval and information filtering. *Personal and Ubiquitous Computing*, 5(4) :253–263, 2001.
- [41] 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.
- [42] J. Budzik and K. Hammond. User interactions with everyday applications as context for just-in-time information access. In *Proceedings of the 5th international conference on intelligent user interfaces*, pages 44–51, 2000.
- [43] M. Buffiere and F. Pichon. Knowledge based flexible query answering. International master in information technology, Aalborg University, Esbjerg, 2005.
- [44] J. M. Cantera, M. Arias, J. Cabrero, G. Garcia, A. Zubizarreta, J. Vegas, and P. de la Fuente. Mymose : Next generation search engine for mobile users. In *the 3rd edition of the Future of Web Search Workshop*, 2008.
- [45] J. Cao, K. M. Chan, G. Y.-K. Shea, and M. Guo. Location-aware information retrieval for mobile computing. In *L.T. Yang et al. (Eds.) : EUC 2004, LNCS 3207*, pages 450–459. Springer-Verlag Berlin Heidelberg, 2004.
- [46] G. Castelli, M. Mamei, and A. Rosi. The whereabouts diary. In *J. Hightower, B. Schiele, and T. Strang (Eds.) : LoCA 2007, LNCS 4718*, pages 175–192. Springer-Verlag, 2007.
- [47] G. Castelli, A. Rosi, M. Mamei, and F. Zambonelli. A simple model and infrastructure for context-aware browsing of the world. In *Proceedings*

- of the 5th Annual IEEE Int'l Conference on Pervasive Computing and Communications (PerCom'07), pages 229–238. IEEE Computer Society, 2007.
- [48] E. Chang, H. Meng, Y. chi Li, and T.-Y. Fung. Efficient web search on mobile devices with multi-modal input and intelligent text summarization. In *Proceedings of the 11th World Wide Web Conference*, 2002.
- [49] G. Chen and D. Kotz. A survey of context-aware mobile computing research. Technical report tr2000-381, Dept. of Computer Science, Dartmouth College, Nov 2000.
- [50] K. Cheverst, N. Davies, K. Mitchell, A. Friday, and C. Efstratiou. Developing a context-aware electronic tourist guide : Some issues and experiences. In *Conference on Human Factors in Computing Systems (CHI)*, pages 17–24. ACM Press, 2000.
- [51] K. Church, J. Neumann, M. Cherubini, and N. Oliver. Socialsearchbrowser : A novel mobile search and information discover tool. In *Proceedings of the 14th Int'l Conference on Intelligent User Interfaces*, pages 101–110. ACM, 2010.
- [52] K. Church and B. Smyth. Who, what, where & when : A new approach to mobile search. In *12th Int'l Conference on Intelligent User Interfaces*, pages 309–312. ACM, 2008.
- [53] K. Church and B. Smyth. Understanding the intent behind mobile information needs. In *13th int'l conference on Intelligent user interfaces*, pages 247–256. ACM, 2009.
- [54] K. Church, B. Smyth, K. Bradley, and P. Cotter. A large scale study of european mobile search behaviour. In *the 10th int'l conference on human computer interaction with mobile devices and services*, pages 13–22. ACM, 2008.
- [55] C. Cleverdon. The cranfield test on index language devices. *ASLIB Proceedings*, pages 173–194, 1967.
- [56] C. Cleverdon. Progress in documentation, evaluation tests of information retrieval systems. *Journal of Documentation*, 26(1) :55–67, 1970.
- [57] W. Cohen. Learning trees and rules with set-valued functions. In *Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence*, pages 709–716. AAAI Press, 1996.
- [58] W. Cohen, R. Schapire, and Y. Singer. Learning to order things. In *NIPS'97 : Proceedings of the 1997 conference on Advances in neural information processing systems 10*. MIT Press Cambridge, 1997.
- [59] G. Cong, C. S. Jensen, and D. Wu. Efficient retrieval of the top k most relevant spatial web objects. In *Proceedings of the VLDB Endowment, PVLDB 2(1)*, pages 337–348, 2009.
- [60] C. Cool. The concept of situation in information science. *Annual review of information science and technology*, 35 :5–42, 2001.

- [61] C. Cool and A. Spink. Issues of context in information retrieval : an introduction to the special issue. *Information Processing and Management*, 38(55) :605–611, 2002.
- [62] P. Coppola, V. Mea, D. Gaspero, and S. Mizzaro. The concept of relevance in mobile and ubiquitous information access. In *F. Crestani et al. (Eds.) : Mobile and Ubiquitous Info. Access Ws 2003*, pages 1–10. LNCS 2954, 2004.
- [63] W. Croft and J. Lafferty, editors. *Language Modeling for Information Retrieval*. Springer, 2003.
- [64] S. Cronen-Townsend, Y. Zhou, and W. Croft. Predicting query performance. In *Proceedings of 25th annual International ACM SIGIR conference on Research and development in information retrieval*, pages 299–306. ACM Press, 2002.
- [65] C. J. Crouch and B. Yang. Experiments in automatic statistical thesaurus construction. In *Proceedings of 15th annual international ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 77–88. ACM, 1992.
- [66] C. da Costa Pereira, M. Dragoni, and G. Pasi. Multidimensional relevance : A new aggregation criterion. In *Proceedings of 31st European Conference on Information Retrieval*, pages 264–275. Springer-Verlag Berlin, Heidelberg, 2009.
- [67] C. da Costa Pereira, M. Dragoni, and G. Pasi. A prioritized ”and” aggregation operator for multidimensional relevance assessment. In *Proceedings of the 11th International Conference of the Italian Association for Artificial Intelligence Reggio Emilia on Emergent Perspectives in Artificial Intelligence*, pages 72–81. Springer-Verlag Berlin, Heidelberg, 2009.
- [68] H. Dai, Z. Nie, L. Wang, L. Zhao, J. R. Wen, and Y. Li. Detecting online commercial intention (oci). In *15th international world wide web conference.*, pages 829–837. ACM, 2006.
- [69] M. Daoud, L. Tamine, and M. Boughanem. Learning user interests for session-based personalized search. In *ACM Information Interaction in context (IiX), London*, pages 57–64. ACM, octobre 2008.
- [70] M. Daoud, L. Tamine, and M. Boughanem. Towards a graph based user profile modeling for a session-based personalized search. *Knowledge and Information Systems*, 21(3) :365–398, 2009.
- [71] N. Davies, K. Mitchell, K. Cheverest, and G. Blair. Developing a context sensitive tourist guide. In *Proceedings of the First Workshop on Human Computer Interaction for Mobile Devices Glasgow*, pages 64–68, 1998.
- [72] S. Dobson. Leveraging the subtleties of location. In *PsOc-EUSAI’05 Proceedings of the 2005 joint conference on Smart objects and ambient intelligence : innovative context-aware services : usages and technologies*, pages 189–193. ACM New York, NY, USA, 2005.

- [73] Z. Dou, R. Song, and J. Wen. A large-scale evaluation and analysis of personalized search strategies. In *Proceedings of the 16th international conference on World Wide Web*, pages 581–590. ACM, 2007.
- [74] 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 information retrieval*, pages 72–79. ACM Press, 2003.
- [75] G. Dupret, V. Murdock, and B. Piwowarski. Web search engine evaluation using clickthrough data and a user model. In *Einat Amitay and Craig G. Murray and Jaime Teevan (Eds.) : WWW2007 workshop Query Log Analysis : Social and Technological Challenges*, 2007.
- [76] M. Eirinaki and M. Vazirgiannis. Web mining for web personalization. *ACM Trans. Internet Technol.*, 3(1) :1–27, 2003.
- [77] D. K. et X. Fu. Eliciting better information need descriptions from users of information search systems. *Information Processing and Management*, 43(1) :30–46, 2007.
- [78] W. FAN, M. GORDON, and P. PATHAK. Discovery of context specific ranking functions for effective information retrieval using genetic programming. *IEEE Transactions on knowledge and data engineering*, 16(4) :523–527, 2004.
- [79] D. A. Ford, J. Ruvolo, S. Edlund, J. Myllymaki, J. Kaufman, J. Jackson, and M. Gerlach. Tempus fugit : a system for making semantic connections. In *Proceedings of the 10th Int'l Conference on Information and knowledge management*, pages 520–522. ACM, 2001.
- [80] E. Frank, Y. Wang, S. Inglis, G. Holmes, and I. Witten. Using model trees for classification. *Machine Learning*, 32(1) :63–76, 1998.
- [81] 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.
- [82] N. Friedman, D. Geiger, and M. Goldszmidt. Bayesian network classifiers. *Machine Learning*, 29 :131–163, 1997.
- [83] N. Fuhr. Information retrieval : introduction and survey. post-graduate course on information retrieval, university of Duisburg-Essen, Germany, 2000.
- [84] S. Gauch, J. Chaffee, and A. Pretschner. Ontology-based personalized search and browsing. *Web Intelligence and Agent Systems*, 1(3-4) :219–234, 2003.
- [85] A. Göker and H. I. Myrhaug. User context and personalisation. In *Workshop on Case Based Reasoning and Personalization, in conjunction of the 6th European Conference on Case Based Reasoning ECCBR*, pages 1–8. Mehmet H. Göker, Barry Smyth (ed.), 2002.

- [86] A. Göker and H. I. Myrhaug. Evaluation of a mobile information system in context. *Information Processing and Management*, 44(1) :39–65, 2008.
- [87] K. Goenka, I. B. Arpinar, and M. Nural. Mobile web search personalization using ontological user profile. In *Proceedings of the 48th ACM Southeast Conference (ACMSE'10)*. ACM, 2010.
- [88] L. Goeuriot. *Découverte et caractérisation des corpus comparables*. Thèse en informatique, Université de Nantes, Nantes, France, Janvier 2009.
- [89] 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.
- [90] F. Gui, M. Adjouadi, and N. Rishe. A contextualized and personalized approach for mobile search. In *the Int'l Conference on Advanced Information Networking and Applications Workshops*, pages 966–971. IEEE Computer Society, 2009.
- [91] B. Habert. Des corpus représentatifs : de quoi, pour quoi, comment ? In M. Bilger, editor, *Linguistique sur corpus. Études et réflexions*, volume 31, pages 11–58. Cahiers de l'université de Perpignan, Presses Universitaires de Perpignan, 2000.
- [92] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. Witten. The weka data mining software : An update. *SIGKDD Explorations*, 11 :10–18, 2009.
- [93] S. HARTER. Variations in relevance assessments and the measurement of retrieval effectiveness. *Journal of the American Society for Information Science*, 47(1) :37–49, janvier 1996.
- [94] S. Hattori, T. Tezuka, and K. Tanaka. Activity-based query refinement for context-aware information retrieval. In *the 9th Int'l conference on Asian digital libraries*, pages 474–477. LNCS, 2006.
- [95] T. Haveliwala. Topic-sensitive page rank. In *International ACM World Wide Web conference*, pages 727–736, 2002.
- [96] T. Heimonen. Mobile findex : Facilitating information access in mobile web search with automatic result clustering. *Advances in Human-Computer Interaction*, 2008 :14, 2009.
- [97] D. Hiemstra and S. Robertson. Relevance feedback for best match term weighting algorithms in information retrieval. In A. Smeaton and J. Callan, editors, *Proceedings of the Joint DELOS-NSF Workshop on Personalisation and Recommender Systems in Digital Libraries, ERCIM Workshop*, pages 37–42, 2001.
- [98] 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.

- [99] A. Hwang, S. Ahern, S. King, M. Naaman, R. Nair, and J. Yang. Zurfer : Mobile multimedia access in spatial, social and topical context. In *Rainer Lienhart, Anand R. Prasad, Alan Hanjalic, Sunghyun Choi, Brian P. Bailey, Nicu Sebe (Ed.) : Proceedings of the 15th Int'l. conference on Multimedia*, pages 557–560. ACM, 2007.
- [100] P. Ingwersen. *Information retrieval interaction*. Taylor Graham, London, 1992.
- [101] 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.
- [102] 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.
- [103] P. INGWERSEN and K. JARVELIN. Information retrieval in context. In *Proceedings of the 27th ACM SIGIR Workshop on information retrieval in context*, pages 6–8, 2004.
- [104] P. Ingwersen and K. Järvelin. *The TURN : Integration of Information Seeking and Retrieval in Context*. SPRINGER, August 2005.
- [105] Y. Ioannidis and G. Koutrika. Personalized systems : Models and methods from an ir and db perspective. In K. Böhm, C. S. Jensen, L. M. Haas, M. L. Kersten, P.-A. Larson, and B. C. Ooi, editors, *Proceedings of the 31st International Conference on Very Large Data Bases*, 2005.
- [106] S. J. and L. M. An exploration in personalized and context-sensitive search. In *Proceedings of the 7th annual UK special interest group for computational linguists research colloquium*, 2003.
- [107] J. H. Jahnke, Y. Bychkov, D. Dahlem, and L. Kawasme. Implicit, context-aware computing for health care. In *Proceedings of the 1st Int'l Workshop on Modeling and Retrieval of Context (MRC 2004)*. Schulz, S. and Roth-Berghofer, Th. (Eds.), CEUR Workshop Proceedings Volume 114, 2004.
- [108] J. H. Jahnke, Y. Bychkov, D. Dahlem, and L. Kawasme. Context-aware views for mobile users (extended abstract). In *the 10th DELOS Thematic Workshop on Personalized Access, Profile Management, and Context Awareness in Digital Libraries, PersDL held in conjunction with the UM 2007*, pages 50–54, 2007.
- [109] 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.
- [110] 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.

- [111] K. Jarvelin and J. Kekalainen. Cumulated gain-based evaluation of ir techniques. *ACM Transactions on Information Systems*, 20(4) :422–446, 2002.
- [112] G. Jeh and J. Widom. Scaling personalized web search. In *WWW '03 : Proceedings of the 12th international conference on World Wide Web*, pages 271–279, New York, NY, USA, 2003. ACM.
- [113] F. Jelinek and R. L. Mercer. Interpolated estimation of markov source parameters from sparse data. In *Proceedings of the Workshop on Pattern Recognition in Practice*, 1980.
- [114] T. Joachims. Optimizing search engines using clickthrough data. In *In KDD'02 : Proceedings of the 8th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 133–142. ACM, 2002.
- [115] T. Joachims, L. Granka, H. Hembrooke, F. Radlinski, and G. Gay. Evaluating the accuracy of implicit feedback from clicks and query reformulations in web search. *ACM Transactions on Information systems*, 25(2) :7, 2007.
- [116] T. Joachims, L. Granka, B. Pan, H. Hembrooke, and G. Gay. Accurately interpreting clickthrough data as implicit feedback. In *Proceedings of the 28th annual int'l. ACM SIGIR conference on Research and development in information retrieval*, pages 154–161. ACM, 2005.
- [117] C. Jones and R. Purves. Geographical information retrieval. *International Journal of Geographical Information Science*, 22(3) :219–228, 2008.
- [118] G. Jones and P. Brown. Context-aware retrieval for ubiquitous computing environments. In *F. Crestani, et al. (Eds) : Mobile and Ubiquitous Information Access*, pages 371–374. LNCS, Vol. 2954, Springer-Verlag Berlin Heidelberg, 2004.
- [119] S. Jones, K. Walker, and S. Robertson. A probabilistic model of information retrieval : Development and comparative experiments. *Information Processing & Management*, 36(6) :779–808, 2000.
- [120] 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.
- [121] M. Kaenamponpan and E. O'Neill. An integrated context model : Bringing activity to context. In *In 1st International Workshop on Advanced Context Modelling, Reasoning And Management- UbiComp'04*, 2004.
- [122] M. Kamvar and S. Baluja. A large scale study of wireless search behavior : Google mobile search. In *the 24th annual SIGCHI conference on human factors in computing systems*, pages 701–709. ACM, 2006.
- [123] M. Kamvar and S. Baluja. Deciphering trends in mobile search. *Computer*, 40(8) :58–62, 2007.
- [124] M. Kamvar and S. Baluja. The role of context in query input : Using contextual signals to complete queries on mobile devices. In *Proceedings of*

- the 9th Conference on Human-Computer Interaction with Mobile Devices and Services*, pages 405–412. Adrian David Cheok, Luca Chittaro (Eds.), 2007.
- [125] A. K. Karlson, G. Robertson, D. C. Robbins, M. Czerwinski, and G. Smith. Fathumb : A facet-based interface for mobile search. In *Proceedings of the SIGCHI conference on Human Factors in computing systems (CHI'06)*, pages 711–720. ACM, 2006.
- [126] C. Keßler, M. Raubal, and C. Wosniok. Semantic rules for context-aware geographical information retrieval. In *P. Barnaghi, K. Moessner, M. Presser, and S. Meissner (Eds.), Smart Sensing and Context - 4th European Conference on Smart Sensing and Context (EuroSSC 2009), LNCS 5741*, pages 77–92. Springer, 2009.
- [127] J. Kekalainen and K. Jarvelin. Evaluating information retrieval systems under the challenges of interaction and multidimensional dynamic relevance. In H. Bruce and R. Fidel, editors, *Proceedings of the 4th CoLIS conference*, pages 253–270. Libraries Unlimited, Greenwood Village CO, ETATS-UNIS, 2002.
- [128] H. H. Kelley. The processs of causal attribution. *American Psychologist*, 28(2) :107–128, 2008.
- [129] L. R. Khan. *Ontology-based Information Selection*. Phd thesis, Faculty of the Graduate School, University of Southern California, August 2000.
- [130] 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.
- [131] K. Kim. Effects of emotion control and task on web searching behavior. *Information Processing and Management*, 44(1) :373–385, 2008.
- [132] L. Kolmonen. Mobile search engine survey. In *The Research Seminar on Telecommunications Business, Helsinki University of Technology*, <http://www.tml.tkk.fi/Opinnot/T-109.7510/2008/Eng1.pdf>, avril 2008.
- [133] A. Komninos and M. D. Dunlop. Calendar based contextual information as an internet content pre-caching tool. In *Proceedings of the ACM SIGIR 2005 Workshop on Information Retrieval in Context (IRiX)*, pages 37–40. ACM, 2005.
- [134] D. Kostadinov, M. Bouzeghoub, and S. Lopes. Accès personnalisé à des sources de données multiples : évaluation de deux approches de reformulation de requêtes. *Revue Ingénierie des Systèmes d'Information (ISI)*, pages 59–82, 2008.
- [135] 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.

- [136] 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.
- [137] S. Kullback and R. Leibler. On Information and Sufficiency. *Annals of Mathematical Statistics*, 22(1) :79–86, 1951.
- [138] D. Kuroopka, editor. *Modelle zur Repräsentation natürlichsprachlicher Dokumente- Information-Filtering und -Retrieval mit relationalen Datenbanken*, volume 10 of *Advances in Information Systems and Management Science*. Logos Verlag, Berlin, 2004.
- [139] N. D. Lane, D. Lymberopoulos, F. Zhao, and A. T. Campbell. Hapori : context-based local search for mobile phones using community behavioral modeling and similarity. In *Ubicomp'10 : Proceedings of the 12th ACM international conference on Ubiquitous computing*, pages 109–118. ACM, 2010.
- [140] K. Lang. Newsweeder : learning to filter netnews. In *Proceedings of the 12th International Conference on Machine Learning*, pages 331–339, San Mateo, CA, USA, 1995. Morgan Kaufmann publishers Inc.
- [141] B. Larsen, H. Lund, J. Andreasen, and P. Ingwersen. Using value-added document representations in inex. In *Proceedings of INEX'03 Workshop*, pages 67–72, 2003.
- [142] V. Lavrenko and W. Croft. Relevance-based language models. In *Proceedings of 24th annual International ACM SIGIR conference on Research and development in information retrieval*, pages 120–127. ACM Press, 2001.
- [143] E. Law, T. Klobucar, and M. Pipan. User effect in evaluating personalized information retrieval systems. In W. Nejdl and K. Tochtermann, editors, *EC-TEL*, pages 257–271. LNCS, vol. 4227, Springer-Verlag, 2006.
- [144] A. Le Parc-Lacayrelle, M. Gaio, and C. Sallaberry. La composante temps dans l'information géographique textuelle extraction et recherche d'information dans des fonds documentaires patrimoniaux numérisés. *Document numérique*, 10 :129–148, FEV 2007.
- [145] D. L. Lee. To find or to be found, that is the question in mobile information retrieval. In *SIGIR 2008 Workshop on MobIR'08*, pages 7–10. ACM, 2008.
- [146] J. H. Lee. Mining user similarity from semantic trajectories. In *Proceedings of the 20th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 267–276. ACM New York, 1997.
- [147] E. L. Lehmann and H. J. M. D'Abbrera, editors. *Nonparametrics : Statistical Methods Based on Ranks*. Prentice Hall, Englewood, Cliffs, NJ, USA, 1998.

- [148] K. Leung, D. Lee, and W. Lee. Personalized web search with location preferences. In *Proceedings of the 26th International Conference on Data Engineering (ICDE)*, pages 701–712. IEEE, 2010.
- [149] Y. Li and N. J. Belkin. A faceted approach to conceptualizing tasks in information seeking. *Information Processing and Management*, 44(6) :1822–1837, 2008.
- [150] H. Lieberman. Letizia : an agent that assists web browsing. In *IJCAI 95 : Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence.*, pages 924–929, 1995.
- [151] C. Lioma, M.-F. Moens, and L. Azzopardi. Collaborative annotation for pseudo relevance feedback. In *ESAIR 2008 : Workshop on Exploring Semantic Annotations in Information Retrieval at ECIR'08*, pages 25–35, Glasgow, UK, 2008.
- [152] 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.
- [153] T. S. Louise. Evaluation measures for interactive information retrieval. *Information Processing and Management*, 28(4) :503–516, 1992.
- [154] 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.
- [155] P. Malo and P. Siitari. A context-aware approach for user profiling with interactive preference learnig. Working paper, w-482, School of Economics, Aalto University, june 2010.
- [156] J. Mc-Gowan. *A multiple model approach to personalised information access*. Master thesis in computer science, Faculty of science, Université de College Dublin, February 2003.
- [157] A. Micarelli, F. Gasparetti, F. Sciarrone, and S. Gauch. Personalized search on the world wide web. In *P. Brusilovsky and A. Kobsa and W. Nedjl (Eds.) : The Adaptive Web*, pages 195–230. Springer-Verlag Berlin Heidelberg, 2007.
- [158] 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.
- [159] T. Mitchell. *Machine Learning*. McGraw-Hill Higher Education, 1997.
- [160] S. Mizzaro. Relevance : the whole history. *J. Am. Soc. Inf. Sci.*, 48(9) :810–832, 1997.
- [161] S. Mizzaro, E. Nazzi, and L. Vassena. Retrieval of context-aware applications on mobile devices : How to evaluate ? In *the second IiX Symposium*, pages 65–71. ACM, 2008.

- [162] B. Mobasher. Data mining for web personalization. In A. N. W. Brusilovsky, P. Kobsa, editor, *The Adaptive Web : Methods and Strategies of Web Personalization*, pages 90–135. LNCS, vol 4321, Springer Verlag Berlin, Heidelberg, 2007.
- [163] C. S.-D. Mohand Boughanem, Claude Chrisment. Query modification based on relevance back-propagation in an ad hoc environment. *Inf. Process. Manage.*, 35(2) :121–139, 1999.
- [164] 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.
- [165] D. Mountain and A. MacFarlane. Geographic information retrieval in a mobile environment : evaluating the needs of mobile individual. *Journal of Information Science*, 33(5) :515–530, 2007.
- [166] H. I. Myrhaug. Towards life-long and personal context spaces. In *Workshop on User Modelling for Context-Aware Applications*, <http://orgwis.gmd.de/gross/um2001ws/papers/myrhaug.pdf>, 2001.
- [167] M. Nakanishi and L. Cooper. Parameter estimates for multiplicative competitive interaction models : Least square approach. *Journal of Marketing Research*, 11 :303–311, 1974.
- [168] R. Navarro-Prieto, M. Scaife, and Y. Rogers. Cognitive strategies in web searching. In *Proceedings of the 5th Conference on Human Factors the Web*, pages 1–13. NIST, 1999.
- [169] I. Ounis, G. Amati, V. Plachouras, B. He, C. Macdonald, and D. Johnson. Terrier information retrieval platform. In *Proceedings of 27th European Conference on Information Retrieval*, pages 517–519. LNCS 3408, Springer, 2005.
- [170] D. Palacio, G. Cabanac, C. Sallaberry, and G. Hubert. Measuring effectiveness of geographic ir systems in digital libraries : Evaluation framework and case study. In A. R.-F. S. I. F. Mounia Lalmas, Joemon Jose, editor, *Proceedings of the 14th European Conference on Research and Advanced Technology for Digital Libraries (ECDL'10)*, pages LNCS Vol. 6273, 340–351. Springer-Verlag Berlin, 2010.
- [171] F. Pan. *Representing complex temporal phenomena for the semantic web and natural language*. Ph.d thesis, University of Southern California, Dec 2007.
- [172] C. Panayiotou, M. Andreou, G. Samaras, and A. Pitsillides. Time based personalization for the moving user. In *Proceedings of the 4th Int'l Conference on Mobile Business (ICMB'05)*, pages 128–136, 2005.
- [173] C. Paolo, M. V. Della, di Gaspero Luca, L. Raffaella, M. Danny, M. Stefano, N. Elena, S. Ivan, and V. Luca. Ai techniques in a context-aware ubiquitous environment. In A. Hassaniien, J. Abawajy, A. Abraham, and

- H. Hagrais (Ed.), Pervasive Computing : Innovations in Intelligent Multimedia and Applications. Series in Computer Communications and Networks*, pages 157–180. Springer-Verlag, 2009.
- [174] M. Papadogiorgaki, V. Papastathis, E. Nidelkou, SimonWaddington, B. Bratu, M. Ribiere, and I. Kompatsiaris. Two-level automatic adaptation of a distributed user profile for personalized news content delivery. *International Journal of Digital Multimedia Broadcasting*, 2008 :1–21, 2008.
- [175] T. K. Park. Toward a theory of user-based relevance : a call for a new paradigm of inquiry. *Journal of the American Society for Information Science*, 45(3) :135–141, 1994.
- [176] J. Pascoe. The stick-e note architecture : Extending the interface beyond the user. In *Proceedings of the 1997 Int'l Conference on Intelligent User Interfaces (IUI'97)*, pages 261–264. ACM, 1997.
- [177] G. Pass, A. Chowdhury, and C. Torgeson. A picture of search. In *proceedings of the 1st International Conference on Scalable Information Systems, Article 1*, Milton Keynes, 2006. ACM.
- [178] M. J. Pazzani, J. Muramatsu, and D. Billsus. Syskill & weber : Identifying interesting web sites. In *Proceedings of the 30th National Conference on Artificial Intelligence*, pages 54–61, Portland, 1996.
- [179] B. Pia. The concept of relevance in ir. *Journal of the American Society for Information Science*, 54(10) :913–925, 2003.
- [180] B. Pia. The iir evaluation model : A framework for evaluation of interactive information retrieval systems. *Journal of Information Research*, 8(3), 2003.
- [181] J. Platt. Fast Training of Support Vector Machines using Sequential Minimal Optimization. In C. B. B. Schoelkopf and e. A. Smola, editors, *Advances in Kernel Methods - Support Vector Learning*. MIT Press, 1998.
- [182] M. Porter. An algorithm for suffix stripping. *Program*, 14(3) :130–137, 1980.
- [183] S. Poslad, H. Laamanen, R. Malaka, A. Nick, P. Buckle, and A. Zipf. Crumpet : Creation of user-friendly mobile services personalised for tourism. In *3G 2001-Second Int'l Conference on 3G Mobile Communication Technologies*, pages 28–32, 2001.
- [184] A. J. Purnima Chandrasekaran. Mobileiq : A framework for mobile information access. In *Proceedings of the 3rd Int'l Conference on Mobile Data Management (MDM'02)*, pages 43–50, 2004.
- [185] J. Pustejovsky, K. Lee, H. Bunt, and L. Romary. Iso-timeml : An international standard for semantic annotation. In *Proceedings of the 7th conference on International Language Resources and Evaluation LREC10*. European Language Resources Association (ELRA), 2010.

- [186] F. Qiu and J. Cho. Automatic identification of user interest for personalized search. In *WWW'06 : Proceedings of the 15th int'l. conference on World Wide Web*, pages 727–736. ACM, 2006.
- [187] D. Quercia, N. Lathia, F. Calabrese, G. D. Lorenzo, and J. Crowcroft. Recommending social events from mobile phone location data. In *Proceedings of 2010 IEEE Int'l. Conference on Data Mining (ICDM 2010)*, pages 971–976, 2010.
- [188] R. Quinlan. *C4.5 : Programs for Machine Learning*. Morgan Kaufmann Publishers, San Mateo, CA, 1993.
- [189] 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.
- [190] L. Ratinov and V. Roth. Design challenges and misconceptions in named entity recognition. In *Proceedings of 13st Conference on Computational Natural Language Learning*, pages 147–155. Association for Computational Linguistics, 2009.
- [191] F. Ren, L. Fan, and J.-Y. Nie. Saak approach : How to acquire knowledge in an actual application system. In *IASTED International Conference on Artificial Intelligence and Soft Computing*, pages 136–140, 1999.
- [192] B. Rhodes. Using physical context for just-in-time information retrieval. *IEEE Transactions On Computers*, 52(8) :1011–1014, 2003.
- [193] J. Rieman. The diary study : a workplace-oriented research tool to guide laboratory efforts. In *INTERACT'93 and CHI'93 conference on Human factors in computing systems*, pages 321–326. ACM NY, 1993.
- [194] S. Robertson. The probability ranking principle in ir. In *Journal of documentation*, 33(4) :294–304, 1977.
- [195] S. Robertson. Comparing the performance of adaptive filtering and ranked output systems. *Information Retrieval*, 5 :257–268, 2002.
- [196] S. Robertson and M. Hancock-Beaulieu. On the evaluation of ir systems. *Information Processing and Management*, 28 :457–466, 1992.
- [197] S. E. Robertson, S. Walker, M. Hancock-Beaulieu, and M. Gatford. Okapi at trec-3. In *Text REtrieval Conference (TREC-3)*, pages 109–126, 1995.
- [198] J. J. Rocchio. Relevance feedback in information retrieval. In Salton [204].
- [199] H. Rode and D. Hiemstra. Using query profiles for clarification. In M. Lalmas, A. MacFarlane, S. M. Rüger, A. Tombros, T. Tsirikra, and A. Yavlinsky, editors, *ECIR*, volume 3936 of *Lecture Notes in Computer Science*, pages 205–216. Springer, 2006.
- [200] D. E. Rose and D. Levinson. Understanding user goals in web search. In *13th international conference on World Wide Web*, pages 13–19. ACM, 2004.

- [201] N. Ryan, J. Pascoe, and D. Morse. Enhanced Reality Fieldwork : the Context-Aware Archaeological Assistant. In V. Gaffney, M. van Leusen, and S. Exxon, editors, *Computer Applications in Archaeology*. British Archaeological Reports, 1997.
- [202] W. Ryen, I. Ruthven, J. Jose, and C. J. V. Rijsbergen. Evaluating implicit feedback models using searcher simulations. *ACM Transactions on Information Systems*, 23(3) :325–361, 2005.
- [203] G. Salton. A comparison between manual and automatic indexing methods. *Journal of American Documentation*, 20(1) :61–71, 1971.
- [204] G. Salton, editor. *The SMART Retrieval System - Experiments in Automatic Document Processing*. Prentice Hall, Englewood, Cliffs, New Jersey, 1971.
- [205] G. Salton and M. McGill. *Introduction to Modern Information Retrieval*. McGraw-Hill, New York, 1983.
- [206] G. Salton and C. Yang. On the specification of term values in automatic indexing. *Journal of documentation*, pages 351–372, 1973.
- [207] M. Sanderson. The Reuters collection. In *Proceedings of the 16th BCS IRSG Colloquium*. London : T. Graham, 1994.
- [208] T. Saracevic. Relevance reconsidered '96. In . N. P. P. Ingwersen, editor, *Proceedings of CoLIS 2, second international conference on conceptions of library and information science : Integration in perspective*, pages 201–218. Copenhagen : Royal School of Library, 1996.
- [209] 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, 1997.
- [210] T. Saracevic, H. Mokros, L. Su, and A. Spink. Interaction between users and intermediaries in online searching. In *Proceedings of the 12th Annual National Online Meeting*, pages 329–341, 1991.
- [211] B. Schilit, N. Adams, and R. Want. Context-aware computing applications. In *IEEE Workshop on Mobile Computing Systems and Applications*, pages 85–90, 1994.
- [212] B. Schilit and M. Theimer. Disseminating active map information to mobile hosts. *IEEE Network*, 8(5) :22–32, 1994.
- [213] A. Schmidt, M. Beigl, and H. Gellersen. There is more to context than location. *Computers & Graphics Journal*, 23(6) :893–902, 1999.
- [214] M. C.-J. H. M. K. Y. H. S.H. Lin, C.S. Shih. Extracting classification knowledge of internet documents with mining term associations : A semantic approach. In *Proceedings of ACM SIGIR '98*, pages 241–249. ACM, 1998.
- [215] L. Shamber. Relevance and information behavior. *ARIST*, 29 :3–48, 1994.

- [216] 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.
- [217] 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.
- [218] A. Sieg, B. Mobasher, and R. Burke. User's information context : Integrating user profiles and concept hierarchies. In *Proceedings of the 2004 Meeting of the International Federation of Classification Societies*, pages 28–40, 2004.
- [219] 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. ACM, 2007.
- [220] 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*.
- [221] C. Silverstein, M. Henzinger, H. Marais, and M. Moricz. Analysis of a Very Large Web Search Engine Query Log. *ACM SIGIR Forum*, 33(1) :6–12, 1999.
- [222] B. Smyth and E. Balfe. Anonymous personalization in collaborative web search. *Information retrieval*, 9(2) :165–190, 2006.
- [223] T. Sohn, K. A. Li, W. G. Griswold, and J. D. Hollan. A diary study of mobile information needs. In *the 26th annual SIGCHI conference on Human factors in computing systems*, pages 433–442. ACM, 2008.
- [224] D. Sonnenwald. Evolving perspectives of human information behavior : Contexts, situations, social networks and information horizons. In *Exploring the Contexts of Information Behavior : Proceedings of the Second International Conference in Information Needs*, pages 176–190, London, UK, 1999. Taylor Graham Publishing.
- [225] 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. IEEE Computer Society, 2005.
- [226] T. Strang and C. L. Popien. A context modeling survey. In *Workshop on Advanced Context Modelling, Reasoning and Management, UbiComp 2004 In - The 6th International Conference on Ubiquitous Computing*, September 2004.

- [227] J. Sun, H. Zeng, H. Liu, Y. Lu, and Z. Chen. Cubesvd : a novel approach to personalized web search. In *WWW'05 : Proceedings of the 14th int'l. conference on World Wide Web*, pages 382–390. ACM, 2005.
- [228] S. Sweeney, F. Crestani, and A. Tombros. Mobile delivery of news using hierarchical query-biased summaries. In *Proceedings of the 17th ACM Symposium on Applied Computing*, pages 634–639. Adrian David Cheok, Luca Chittaro (Eds.), 2002.
- [229] K. Takahashi. A mobile portal service to provide location dependent information. In *Proceedings from the Joint W3C-WAP Forum workshop on*.
- [230] 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.
- [231] L. Tamine, M. Boughanem, and M. Daoud. Evaluation of contextual information retrieval effectiveness : overview of issues and research. *Knowledge and Information Systems*, 24(1) :1–34, 2009.
- [232] 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.
- [233] L. Tamine, C. Chrisment, and M. Boughanem. Multiple query evaluation based on an enhanced genetic algorithm. *Information Processing and Management*, 39(2) :215–231, 2003.
- [234] 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.
- [235] L. Tamine-Lechani, M. Boughanem, and M. Daoud. Evaluation of contextual information retrieval effectiveness : Overview of issues and research. *KIS*, 24(1) :1–34, 2009.
- [236] 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.*, 6(5) :354–366, 2008.
- [237] F. Tanudjaja and L. Mui. Persona : A contextualized and personalized web search. In *Proc 35th Hawaii International Conference on System Sciences*, pages 1232–1240, 2002.
- [238] Y. Tao, N. Mamoulis, and D. Papadias. Validity information retrieval for spatio-temporal queries. In *SSTD : International symposium on advances in spatial and temporal databases*, pages 159–178, 2003.
- [239] taptu. Making search social : Unleashing search for the mobile generation. Taptu white paper, February 2008.

- [240] taptu. The state of the mobile touch web. Taptu report, www.taptu.com/metrics, Avril 2010.
- [241] 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.
- [242] T. Tezuka and K. Tanaka. Temporal and spatial attribute extraction from web documents and time-specific regional web search system. In *Web and Wireless Geographical Information Systems W2GI'04*, pages 14–25. Y.J. Kwon, A. Bouju, and C. Claramunt (Eds.), LNCS 3428, Springer-Verlag, 2005.
- [243] M. Timothy, T. Sherry, and M. Robert. Hypermedia learning and prior knowledge : domain expertise vs. system expertise. *Journal of Computer Assisted Learning*, 21(12) :53–64, 2005.
- [244] S. Toivonen, J. Kolari, and T. Laakko. Facilitating mobile users with contextualized content. In *Proceedings of the Artificial Intelligence in Mobile Systems (AIMS) 2003 Workshop held in conjunction with Ubi-comp 2003*, 2003.
- [245] A. Tombros, I. Ruthven, and J. M. Jose. How users assess web pages for information seeking. *Journal of the American Society of Information Science and Technology (JASIST)*, 56(4) :327–344, 2005.
- [246] F. S. Tsai, M. Etoh, X. Xie, W.-C. Lee, and Q. Yang. Introduction to mobile information retrieval. *Intelligent Systems*, 25(1) :11–15, 2010.
- [247] S. Vadrevu, Y. Zhang, B. Tseng, G. Sun, and X. Li. Identifying regional sensitive queries in web search. In *17th international conference on World Wide Web*, pages 507–514. ACM, 2008.
- [248] 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.
- [249] D. Vallet, M. Fernández, P. Castells, P. Mylonas, and Y. Avrithis. Personalized information retrieval in context. In *Proceedings of the 21st National Conference on Artificial Intelligence - 3rd International Workshop on Modeling and Retrieval of Context*. American Association for Artificial Intelligence, 2006.
- [250] C. van Rijsbergen. *Information Retrieval*. Butterworth & Co (Publishers) Ltd, second edition, London, 1979.
- [251] M. van Setten, S. Pokraev, and K. J. Context-aware recommendations in the mobile tourist application compass. In *Adaptive Hypermedia and Adaptive Web-Based Systems*, pages 515–548. Nejdil, W. & De Bra, P. (Eds.), LNCS 3137, Springer-Verlag, 2004.

- [252] V. Varma, N. Sriharsha, and P. Pingali. Personalized web search engine for mobile devices. In *Int'l Workshop on Intelligent Information Access, IIIA '06*, 2006.
- [253] W. Viana, M. Villanova-Oliver, J. Gensel, and H. Martin. Annotation contextuelle automatique avec photomap. In *Emmanuel Dubois, Jean-Marc Pierson (Eds.) : Actes des 4^{ième} journées francophones Mobilité et Ubiquité 2008, UBIMOB'08*, pages 89–90. ACM, 2008.
- [254] 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.
- [255] L. Wang, C. Wang, X. Xie, J. Forman, Y. Lu, W. Ma, and Y. Li. Detecting dominant locations from search queries. In *Proceedings of 28th annual international ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 424–431. ACM, 2005.
- [256] C. Watters and G. Amoudi. Geosearcher : location-based ranking of search engine results. *Journal of the American Society in Information Science and Technology*, 54(2) :140–151, 2003.
- [257] 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.
- [258] M. Welch and J. Cho. Automatically identifying localizable queries. In *Proceedings of 31st annual International ACM SIGIR conference on Research and development in information retrieval*, pages 1185–1186, Milton Keynes, 2008. ACM.
- [259] J.-R. Wen, N. Lao, and W.-Y. Ma. Probabilistic model for contextual retrieval. 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.
- [260] 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.
- [261] T. WILSON. On user studies and information needs. *Journal of Librarianship*, 37(1) :3–15, 1981.
- [262] Z. Wu and M. Palmer. Verb semantics and lexical selection. In *32nd Annual Meeting of the Association for Computational Linguistics*, pages 133–138, 1994.
- [263] 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.

- [264] J. XU. *Solving the word mismatch problem through automatic text analysis*. Ph.d thesis, Department of Computer Science, University of Massachusetts, Amherst, MA,USA, May 1997.
- [265] R. Yager. Prioritized aggregation operator. *International Journal of Approximate Reasoning*, 48(1) :263–274, 2008.
- [266] S. Yau, H. Liu, D. Huang, and Y. Yao. Situation-aware personalized information retrieval for mobile internet. In *27th Annual International Computer Software and Applications Conference (COMPSAC'03)*, 2003.
- [267] C. A. Yeung, N. Gibbins, and N. Shadbolt. Web search disambiguation by collaborative tagging. In *ESAIR 2008 : Workshop on Exploring Semantic Annotations in Information Retrieval at ECIR'08*, pages 48–61, Glasgow, UK, 2008.
- [268] J. Yi, F. Maghoul, and J. Pedersen. Deciphering mobile search patterns : A study of yahoo! mobile search queries. In *the 17th International Conference on World Wide Web*, pages 257–266, 2008.
- [269] X. Yi, H. Raghavan, and C. Leggetter. Discovering users' specific geo intention in web search. In *the 18th Int'l Conference on World Wide Web*, pages 481–490, 2009.
- [270] J. J.-C. Ying, E. H.-C. Lu, W.-C. Lee, T.-C. Weng, and V. S. Tseng. Mining user similarity from semantic trajectories. In *2nd ACM SIGSPATIAL Int'l. Workshop on Location Based Social Networks*. ACM, 2010.
- [271] H. Yoon, E. Kim, and M. Lee. A user context awareness model for mobile environment processing. In *Proceedings of the 5th Int'l. Conference on Software Engineering Research, Management and Applications*, pages 293–296. IEEE Computer Society, 2007.
- [272] 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.
- [273] 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.
- [274] J. Zien, J. Meyer, J. Tomlin, and J. Liu. Web query characteristics and their implications on search engines. In *Proceedings of the 10th International WWW Conference, Hong Kong*, 2001.