Chapitre 7

Recherche d'information contextuelle et *web*

7.1. Introduction

La recherche d'information, domaine déjà ancien, n'a cessé d'évoluer dans le but de rationaliser le processus complexe permettant l'identification, au sein de volumes de plus en plus importants d'informations, celles qui sont potentiellement intéressantes pour l'utilisateur. Cette évolution a été tout d'abord marquée par l'émergence d'approches pour la modélisation de l'accès à l'information, assujetties à des méthodologies d'évaluation de leur efficacité. Ensuite, des réflexions ont été menées dans le sens de l'introduction de la dimension cognitive dans l'approche de modélisation de la chaîne d'accès à l'information. Cette vision est résumée par Demey [DEM 77] :

« ... any processing of information, whether perceptual or symbolic, is mediated by a system of categories or concepts which, for the information processing device (human being or a machine), are a model of his (its) world ».

Cette expression, qui ne dit rien des méthodes à employer, ni des verrous à lever, pointe avec précision l'objectif : contextualiser le traitement de l'information. Cette initiative a été largement encouragée, ces dernières années, par l'explosion phénoménale du *web* et des nouveaux types de services qu'il offre. En effet, la situation est actuellement paradoxale : la masse d'informations est telle que l'accès à une information pertinente, adaptée aux besoins spécifiques d'un utilisateur donné, devient à la fois difficile et nécessaire. En clair, le problème n'est pas tant la disponibilité de l'information mais sa pertinence relativement à un contexte d'utilisation particulier. Les besoins sont ainsi énormes. Dans ce cadre, la RI contextuelle émerge comme un

---

Chapitre rédigé par Lynda TAMINE-LECHANI et Sylvie CALABRETTO.

domaine à part entière : sans remettre en cause ses origines, elle pose des problématiques nouvelles allant de la modélisation du contexte jusqu'à la modélisation de la pertinence cognitive en passant par la modélisation de l'interaction entre un utilisateur et un système de recherche d'information (SRI).

L'objectif de ce chapitre est d'apporter un éclairage sur ces problématiques, sur leurs origines, sur les verrous posés par la RI classique dans un tel cadre, ainsi que sur les solutions apportées dans la communauté.

L'organisation retenue pour ce chapitre est la suivante : nous rapportons tout d'abord, en section 7.2, les diverses motivations qui ont conduit à l'émergence de la RI contextuelle puis nous présentons, en section 7.3, les définitions des principaux concepts qui en traduisent les dimensions fondamentales. La section 7.4 est consacrée à la présentation des principaux travaux portant sur les approches, modèles et techniques de RI contextuelle, suivie en section 7.5, d'une présentation des systèmes opérationnels nés de ces travaux. La section 7.6 traite de la problématique de l'évaluation de l'efficacité des modèles de RI contextuelle puis en présente les nouveaux protocoles d'évaluation mis en œuvre. Enfin la section 7.7 conclut le chapitre.

7.2. De la recherche d'information orientée système à la recherche d'information orientée utilisateur

La revue de la littérature en RI depuis les années 1960 jusqu'à ce jour met indéniablement en exergue des avancées dans le domaine. Les années 1960 à 1990 ont marqué l'apparition de trois approches majeures en RI : RI orientée système, RI orientée utilisateur et RI cognitive.

L'approche orientée système, apparue les années 1960, a axé l'investigation sur la représentation de l'information, l'évaluation de requêtes ainsi que l'évaluation des performances de recherche. Les résultats phares obtenus sont de façon non exhaustive :

– le développement de modèles mathématiques de RI : modèle vectoriel [SAL 68] et modèle probabiliste durant la période 1960 à 1970 [ROB 77] puis modèles logiques durant les années 1980 [RIJ 86] ;

– le développement de méthodes de classification et de catégorisation de textes durant la période 1960 à 1970 [BEO 63] ;

– le développement de stratégies de RI adaptative : réinjection de la pertinence et d'expansion de requête [ROC 71] durant les années 1970 ;

– l'application de techniques d'analyse du langage naturel durant les années 1980 [CHU 89] ;

– le développement d'une méthodologie d'évaluation en RI basée sur le paradigme de Cranfield durant les années 1960 [CLE 67].

L'approche orientée utilisateur ainsi que l'approche cognitive, apparues en 1977, ont investi quant à elles, les aspects liés aux interactions entre un utilisateur et un système d'accès à l'information et ce, à travers les phases d'expression des besoins, de perception de l'information et de la définition de la pertinence. Ce courant de recherche est né de l'essor du *web* qui a remis l'utilisateur au centre d'un processus de recherche d'information. Les travaux s'intéressent alors à l'intégration de la perspective cognitive de l'utilisateur [BRO 77, JAR 86, ING 96] dans l'interprétation du concept information [BEL 78], de l'interprétation du besoin en information dans le cadre d'une tâche ou d'une situation, de l'interdépendance des éléments de l'environnement de l'utilisateur et leur impact sur sa perception de la pertinence [BEL 87].

Le début des années 1990 a marqué une période de profusion des travaux de recherche dans le domaine de la RI de manière générale. Cet essor a été encouragé d'une part par l'apparition de TREC [HAR 92a], et d'autre part, motivé par la généralisation du *web* qui a posé de nouvelles problématiques. Dans le cadre particulier des approches orientées utilisateur, la recherche s'est orientée vers une vision plus large et plus conséquente de l'utilisateur dans le processus d'accès à l'information. Plus précisément, les stratégies de RI adaptative qui ont germé de l'approche orientée système ont évolué, grâce aux fondements théoriques de l'approche cognitive, vers de nouveaux modèles théoriques et techniques d'interprétation de l'information (requête et/ou document) et de la pertinence utilisateur. Les résultats fondamentaux issus de cette perspective sont notamment :

– le développement de modèles cognitifs de l'interaction en RI [SAR 91, VAK 01] ;

– le développement de méthodes pour la multi-représentation de documents et requêtes représentant les différentes visions cognitives de leur contenu [ING 94, LAR 03] ;

– le développement de stratégies et de modèles d'accès contextuel à l'information [BEL 96, TEE 05].

Dans ce chapitre, nous nous focaliserons sur les approches et modèles d'accès contextuel à l'information. Comme préambule à la présentation des travaux associés, cette section développera les principaux facteurs qui ont conduit à l'émergence de la problématique liée à l'accès contextuel à l'information, puis identifiera les principaux verrous technologiques qui y sont inhérents.

7.2.1. *La recherche d'information adaptative : quel bilan ?*

7.2.1.1. *Large aperçu de la RI adaptative*

Il est communément admis, dans la communauté en RI, qu'une problématique majeure dans le domaine est la différence des univers de discours des utilisateurs et des

auteurs de documents. Ceci se traduit par la différence de vocabulaire utilisé d'une part pour l'expression des contenus des documents et, d'autre part, pour l'expression des besoins en information. Les modèles classiques de sélection de l'information pertinente étant basés sur l'appariement des descripteurs des documents et des requêtes, il s'ensuit ainsi un défaut d'appariement qui engendre une dégradation des performances de recherche. Ce constat est d'autant plus problématique quand on considère d'autres facteurs : requêtes courtes, volume d'information important, expression plus ou moins vague du besoin en information, etc. [XU 97]. La première direction des travaux ayant apporté des solutions à ce problème est 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.

On distingue principalement deux grandes classes de stratégies : (1) stratégies semi-automatiques guidées par l'utilisateur, appelées aussi stratégies de réinjection de la pertinence, (2) stratégies automatiques qui intègrent d'autres sources d'évidence que l'utilisateur lui-même.

On rappelle globalement dans ce qui suit le principe adopté ainsi que les principaux travaux du domaine. Une présentation large de ces travaux est rapportée dans [EFT 96].

– *Stratégies de reformulation semi-automatique* : la reformulation semi-automatique de requête, dite par réinjection de pertinence, est un processus itératif qui consiste à enrichir la requête initiale de l'utilisateur par ajout et/ou repondération de termes sur la base de la structure des documents retrouvés par le SRI et jugés explicitement pertinents ou non pertinents par l'utilisateur. La première formalisation de cette réécriture a été donnée par Rocchio [ROC 71]. D'autres travaux ont, dans ce cadre, exploré les algorithmes d'ordonnancement des termes d'expansion selon leur qualité [BUC 95, HAR 92b, BEA 97], ou proposé la combinaison de différents algorithmes d'injection de la pertinence [ING 94, LEE 98].

– *Stratégies de reformulation automatique* : le principe de reformulation de requête ne fait pas intervenir explicitement l'utilisateur. Il est basé dans ce cas sur l'utilisation d'autres sources d'évidence telles que :

1) les descripteurs des premiers documents extraits de la liste ordonnée retournée par le SRI [CRO 79]. En l'absence du jugement explicite de pertinence de l'utilisateur, cette stratégie pose essentiellement le problème de dérive du sujet de la requête (*query drift*) auquel des solutions ont été apportées notamment par [MIT 98, BUC 94] ;

2) l'utilisation de relations sémantiques entre termes de la requête et termes de la collection nécessitant un *thesaurus* [JIN 94] ou des ontologies [VOO 94, JAR 94] ;

3) l'exploitation des informations issues des interactions de l'utilisateur avec un système d'accès à l'information. Ces travaux plus récents [KEL 04] sont les précurseurs quant à l'évaluation implicite de la pertinence en vue d'un accès contextuel à l'information.

7.2.1.2. *Bilan*

Les travaux en RI adaptative ont certes apporté des solutions au problème du défaut d'appariement requête-document en permettant d'améliorer les performances du processus de recherche d'information [HAR 92b]. Cependant une analyse fine des résultats d'expérimentations rapportées dans la littérature montre que ces performances dépendent de nombreux facteurs *a priori* non contrôlés de manière inhérente au processus de réécriture adaptative de la requête. Ces facteurs, qui sont ainsi problématiques, peuvent être catégorisés selon trois principales dimensions : l'utilisateur, l'information portée par la requête et/ou le document, l'interaction entre l'utilisateur et le SRI. Nous présentons dans ce qui suit chacune de ces dimensions puis en dégageons les éléments précurseurs ayant déterminé les directions d'investigation de la RI contextuelle.

7.2.1.2.1. La dimension utilisateur

On y met en évidence les éléments suivants :

1) l'expression initiale du besoin en information de l'utilisateur (ce qu'il ne sait pas du sujet de la requête) dépend de ses centres d'intérêts (ce qu'il sait déjà du sujet de la recherche) et de ses buts (pourquoi savoir sur le sujet de la requête) [ING 96]. Cependant, ces éléments ne se déclinent pas dans le processus de réécriture de la requête initiale ;

2) [HSI 93] démontre qu'il existe une corrélation positive entre la familiarité de l'utilisateur avec le sujet de la requête et les performances de la stratégie de réinjection de la pertinence. De plus, le niveau d'expertise de l'utilisateur [RUT 03, WHI 03] a un impact sur les performances de recherche car des utilisateurs expérimentés effectuent de meilleurs choix quant à la qualité des documents et termes utilisés pour la réécriture de la requête, relativement à des utilisateurs novices ;

3) [FID 91, HEU 99] montrent que la discipline professionnelle de l'utilisateur n'est pas sans impact dans la perception de l'information et donc de la pertinence. Ceci influe directement sur les performances de recherche ;

4) la nature (utilité, intérêt, préférence) et la valeur du jugement de pertinence de l'utilisateur (peu pertinent, très pertinent, assez pertinent, etc.) dépend de nombreux facteurs : (1) de ses centres d'intérêts et ses buts [VAK 00] (2) de l'objet de la requête (ce qui est attendu à travers une requête : service, information, page de référence) [LOR 06, KAN 04], (3) de la complexité de la tâche de recherche qui est déterminée par la quantité d'information que doit traiter l'utilisateur pour atteindre l'information

pertinente [VAK 01]. Cependant, la RI adaptative exploite des jugements de pertinence binaire supposés ne dépendre que du contenu des documents.

7.2.1.2.2. La dimension information

On y met en évidence les éléments suivants :

1) le volume important de l'information sur le *web* engendre incontestablement une diversité importante du vocabulaire. Par conséquent, les algorithmes d'ordonnancement des termes d'expansion de requête en fonction de leur corrélation au sujet de la requête, sont peu performants [CRO 95] ;

2) les documents du *web* contiennent de nombreuses informations non directement liées au sujet du document telles que les liens de navigation, les informations ou images publicitaires, etc. Ces informations, même extraites des documents les mieux classés à l'issue d'une recherche initiale, engendrent du bruit lors d'un processus de réécriture de requête [YU 03] ;

3) les stratégies classiques de réinjection de pertinence sont peu capables de rappeler des documents traitant de différents sujets auxiliaires associés à un sujet fédérateur véhiculé par la requête [ZHA 03]. Le même problème est posé avec des documents traitant de nombreux sujets à la fois tels que les journaux [YU 03].

7.2.1.2.3. La dimension interaction

On y met en évidence les éléments suivants :

1) les processus de réinjection de la pertinence induisent une interaction qui est à l'origine d'une surcharge cognitive pour l'utilisateur. La valeur ajoutée de ces interactions dépend du degré de participation de l'utilisateur. De plus, des études ont montré [BEL 01, WHI 03] que les utilisateurs n'usent pas forcément de l'ensemble des possibilités offertes par le système quant à l'enrichissement de la requête et ce, pour une raison majeure : les utilisateurs n'en cernent pas le principe et le lien avec l'opération de sélection de l'information pertinente ;

2) la forme de présentation des documents (titre, résumé, texte plein) exploités pour la reinjection de la pertinence a un impact non négligeable sur le jugement de l'utilisateur [JAN 91] ;

3) l'utilité de la réinjection de pertinence est plus déterminante aux dernières itérations d'un processus de recherche d'information adaptative [WHI 03].

Ce bilan montre globalement que les stratégies de RI adaptative ne sont pas garantes de l'uniformité de la qualité des résultats d'un SRI dans des conditions d'utilisation différentes. Il en ressort que les éléments-clés à intégrer dans de telles stratégies dans le but d'en améliorer les performances, sont dépendants les uns des autres, liés cependant à différentes dimensions. Plus précisément, ce bilan suggère d'investir les directions suivantes :

– considérer, outre le contenu des documents, les facteurs descriptifs de l'utilisateur, dans le processus de sélection de l'information pertinente ;

– subordonner la notion de pertinence à un utilisateur spécifique dans une situation de recherche d'information donnée ;

– explorer le but de la recherche d'information conjointement, à travers la requête et l'utilisateur qui l'a émise.

Ces points constituent précisément une partie des enjeux fixés dans le domaine de la RI contextuelle.

7.2.2. *Le web et la recherche d'information contextuelle*

Le développement d'Internet au niveau mondial a profondément transformé la gestion des documents. Cette révolution technologique a engendré de nouvelles problématiques documentaires pour la RI [KOB 00] du fait de l'accroissement incessant et quasi exponentiel [LYM 03] du volume d'information disponible sur le web ainsi que du nombre d'utilisateurs inexpérimentés dans la manipulation des moteurs de recherche. Parmi ces problématiques, l'intégration du contexte constitue une priorité dans le traitement de grands volumes d'information. En effet, il est devenu difficile pour les utilisateurs de trouver des informations sur le *web* qui satisfont leurs besoins personnels alors que les ressources informationnelles du *web* ne cessent d'augmenter. Même dans cette configuration, les moteurs de recherche du *web* aident les utilisateurs à trouver des informations utiles sur la toile. Une étude gouvernementale américaine a montré que 85% des personnes utilisent les moteurs de recherche du *web* pour localiser des informations. Cependant, les moteurs de recherche du *web* possèdent des limitations significatives. Lorsqu'une même requête est soumise par différents utilisateurs, la plupart des moteurs de recherche retournent les mêmes résultats. En effet, les résultats pour une requête donnée sont identiques, indépendants de l'utilisateur, ou du contexte dans lequel l'utilisateur émet sa requête. Mais, en général, chaque utilisateur a des besoins d'information différents pour sa requête. Les documents dépendent du contexte de la recherche, par exemple : la formation, les intérêts, les précédentes expériences de l'utilisateur, et les informations sur la requête courante. L'utilisateur recherche-t-il une entreprise qui vend un produit donné ou des détails techniques sur le produit ? [LAW 00] Un autre exemple concerne la requête Java. Pour cette requête, certains utilisateurs sont intéressés par des documents en rapport avec le langage de programmation Java alors que d'autres utilisateurs souhaitent des documents relatifs au café [SUG 04]. Comme on peut noter un accroissement du nombre de personnes se connectant sur le *web* pour rechercher des informations, il est devenu indispensable de proposer de meilleures solutions pour les services de recherche. La prise en compte des informations du contexte de l'utilisateur constitue une piste d'amélioration intéressante. Les résultats de recherche du *web* doivent s'adapter aux utilisateurs possédant différents besoins en information. Les prochaines générations de moteurs

de recherche doivent faire un usage croissant du contexte [LIM 06], soit en utilisant des informations du contexte explicite ou implicite de l'utilisateur, soit en implémentant des fonctionnalités additionnelles au sein de contextes restreints [LAW 00]. Une meilleure utilisation du contexte dans les moteurs de recherche permet d'aider à accroître la compétition et la diversité sur le *web*. En outre, aux bases de données utilisées dans les SRI traditionnels, le *web* se différencie en termes de contenu et de structure. En effet, il se caractérise par une forte hétérogénéité des sources d'information : hétérogénéité des langues (on y compte actuellement plus de cent langues différentes sur le *web*), hétérogénéité des médias (texte, image, vidéo), hétérogénéité des structures, etc. La contextualisation de la recherche permet d'adapter le contenu et la structure du document au contexte de l'utilisateur [SUG 04], et de diminuer ainsi l'hétérogénéité des sources d'information. Enfin, l'émergence du domaine de la RI contextuelle provient également de l'évolution de l'environnement physique de l'utilisateur : PDA, téléphonie mobile, etc. Plusieurs études ont considéré les implications de ce nouvel environnement en RI [RHO 00b] [COP 03] [JON 00] [BRO 02]. Les conclusions de ces travaux ont montré la nécessité de prendre en compte le contexte physique dans les modèles de RI.

7.2.3. *Verrous technologiques*

En tant que thématique de recherche, l'accès contextuel à l'information est confronté à des verrous que l'on peut projeter sur trois niveaux : conception, évaluation et mise en œuvre.

– *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.

– *Evaluation :* les méthodologies d'évaluation classiques des systèmes d'accès à l'information devraient 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.

– *Mise en œuvre :* l'exploitation de systèmes d'accès personnalisé 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. Cependant, 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. La contextualisation de la RI doit donc s'accompagner d'une réflexion sur les architectures des systèmes de traitement d'informations et de la définition de techniques de protection de ces informations (méthode de cryptographie, modèles d'octroi ou révocation de droits).

7.3. Notions et vocabulaire de base

7.3.1. *Contexte*

Les notions de *contexte* et *situation* sont en amont de l'interaction utilisateur-SRI. Ces notions ont été initialement introduites, sans distinction de sens, par les travaux de Saracevic [SAR 97] et Ingerwersen [ING 96]. Le contexte (ou situation) y est défini(e) comme l'ensemble des facteurs cognitifs et sociaux ainsi que les buts et intentions de l'utilisateur au cours d'une session de recherche. Une tentative de distinction entre ces notions a fait l'objet d'autres travaux [ALL 97, SON 99] qui 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. C'est le sens générique du contexte qui a été largement exploré cette dernière décennie en RI contextuelle [ING 04].

Même si les auteurs ne convergent pas vers une même définition, on retrouve toutefois des dimensions descriptives communes telles que l'environnement cognitif, le besoin mental en information, l'interaction liée à la recherche d'information, la tâche de recherche, le type du besoin véhiculé par la requête, les ressources disponibles.

Vu sous l'angle de la RI, le contexte possède, selon N. Fuhr [FUH 00], trois principales dimensions : social, application et temps. La dimension sociale définit la composante d'appartenance de l'utilisateur : individuel, groupe ou communauté. La dimension application définit le contexte applicatif du besoin exprimé : recherche *ad hoc*, résolution de problème ou *workflow*.

La dimension temps permet de décrire la circonscription temporelle du besoin exprimé : temps passé *(batch)*, instant courant ou à court terme (interactive), intention ou long terme (personnalisation). Sous l'angle de la dimension temps, on distingue deux types de contexte avec des démarches de personnalisation appropriées. Le contexte courant ou à court terme décrit les besoins et préférences de l'utilisateur lors d'une session d'une recherche. Le contexte persistant décrit les besoins à long terme de l'utilisateur sur diverses sessions de recherche.

7.3.2. *Profil*

Le concept de profil est directement lié à l'utilisateur. L'utilisation de ce concept a été introduite par les travaux en filtrage d'information, pour décrire une structure

représentative de l'utilisateur, plus particulièrement de ses centres d'intérêts. Cette notion est réutilisée en RI contextuelle pour cibler les éléments du contexte dépendant directement de l'utilisateur : centres d'intérêts, familiarité avec le sujet de la recherche, domaine professionnel, expertise, etc.

7.3.3. *RI contextuelle ou RI personnalisée ?*

L'objectif de la RI contextuelle est de délivrer une information pertinente et appropriée au contexte de l'utilisateur qui a émis la requête. La RI contextuelle traduit précisément l'exploitation des éléments du contexte de la recherche dans l'une des principales phases de l'évaluation de requête : reformulation, calcul du score de pertinence de l'information, présentation des résultats de recherche. La RI personnalisée est, à notre connaissance, un type de RI contextuelle où l'accent est mis sur l'utilisation d'un modèle de l'utilisateur préalablement construit [LIU 04, SU 03]. La littérature du domaine ne fait pas état, cependant, d'une distinction franche entre RI contextuelle et RI personnalisée.

7.3.4. *Pertinence contextuelle*

La pertinence est incontestablement la question fondamentale posée en RI. Cette notion subjective, dépendant essentiellement du point de vue de l'utilisateur, a de nouveau été l'objet d'investigations dans le cadre de la RI orientée utilisateur de manière générale, de la RI contextuelle de manière particulière. Dans [BOR 03], on rapporte que la pertinence est un concept multidimensionnel. On distingue principalement, quatre types de pertinence :

– *pertinence algorithmique* : la pertinence est traduite par une mesure algorithmique dépendant des caractéristiques des requêtes d'une part et des documents d'autre part. C'est le seul type de pertinence qui est indépendant du contexte ;

– *pertinence thématique* : la pertinence traduit le degré d'adéquation de l'information à couvrir, en partie, le thème évoqué par le sujet de la requête. C'est le type de pertinence adressé par les assesseurs de la campagne d'évaluation TREC ;

– *pertinence cognitive* : c'est la pertinence liée au thème de la requête, « pondérée » par la perception ou les connaissances de l'utilisateur sur ce même thème ;

– *pertinence situationnelle* : c'est la pertinence liée à la tâche de recherche. Ce type de pertinence traduit essentiellement l'utilité de l'information relativement au but de recherche de l'utilisateur.

La RI contextuelle explore essentiellement la pertinence cognitive et la pertinence situationnelle.

7.4. Les modèles de recherche d'information contextuelle

Nous pouvons noter un manque de représentation du contexte dans les modèles de RI et les SRI. Les modèles classiques ont été définis en supposant un utilisateur unique, un besoin d'information pour chaque requête, une localisation, un temps, un historique et un profil.

Des techniques *ad hoc* pour capturer le temps, l'espace, l'historique et les profils sont injectées dans les modèles. Des capteurs pour récupérer la localisation, des fichiers *logs* pour implémenter l'historique, des méta-données pour décrire les profils, des horloges et des calendriers pour récupérer le temps sont alors nécessaires.

7.4.1. *Déclinaison du contexte dans les systèmes de recherche d'information*

Les approches et techniques de modélisation du contexte sont au cœur du processus d'accès contextuel à l'information. L'objectif du système d'accès contextuel à l'information est de délivrer une information pertinente en fonction du contexte de l'utilisateur. Le contexte peut être exploité à différentes phases du processus d'accès à l'information : dans la formulation de la requête, dans l'accès, dans l'ordonnancement des résultats. 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. Le contexte peut également intervenir dans la définition de la fonction de pertinence. Dans ce sens, [FAN 04] ont proposé l'adaptation des paramètres de la fonction de pertinence au contexte de l'utilisateur, en utilisant les techniques de programmation génétique. Jeh et Widom [JEH 03] ont proposé une variante personnalisée de l'algorithme PageRank en l'occurence PPV (*Personnalized PageRank Vector*). Son principe fondamental consiste à privilégier les pages reliées aux pages préférées de l'utilisateur ou pages citées par ces dernières au cours du processus de calcul des scores de sélection. L'ordonnancement des résultats constitue la phase ultime du processus d'accès à l'information. Cette phase peut également prendre en compte le contexte pour réordonner les résultats fournis par le processus de sélection. De ce fait, l'ordre final des documents à présenter à l'utilisateur est une combinaison de l'ordre produit par le processus de sélection et celui fourni par le contexte de l'utilisateur *via* un calcul de similarité [CHA 04] ou des jugements explicites de la pertinence [MCG 03].

7.4.2. *Etat de l'art des modèles de recherche d'information contextuelle*

Plusieurs modèles peuvent être utilisés pour prendre en compte le contexte en RI : basé sur des *logs* de requêtes, Wen [WEN 04] construit un modèle probabiliste par expansion des termes de la requête pour la recherche contextuelle. Melucci [MEL 05] propose une extension du modèle vectoriel intégrant le contexte de l'utilisateur. Rode

et Hiemstra [ROD 04] proposent un modèle basé sur un langage statistique pour incorporer des informations du contexte dans le processus de recherche. Finkelstein [FIN 02] développe un système de recherche qui utilise le contexte entourant le texte pour améliorer l'extension de requête. [PRE 99] exploitent des profils basés sur des ontologies pour des recherches personnalisées. [YU 03] étudient l'amélioration de la recherche en se basant sur des *web log data*. SearchPad [HUL 02] explicite des contextes de recherche sur le *web*.

Nous allons présenter plus en détail deux modèles de RI contextuelle qui sont des extensions des modèles classiques de RI, en l'occurrence les modèles probabiliste et vectoriel.

7.4.3. *Extension du modèle probabiliste*

Dans les travaux sur la RI contextuelle, requête, contexte et document sont représentés par le même type de composant : le mot. Il est ainsi aisé d'adapter les données du contexte dans le cadre de la RI textuelle. Dans de nombreux travaux, le contexte est simplement combiné avec la requête pour effectuer une expansion de requête. Cependant, le contexte peut contenir des informations non compatibles avec les requêtes et les documents d'où une impossibilité d'incorporer le contexte dans une requête. La méthode proposée dans [WEN 04] fait appel aux *logs* de requêtes. Ces *logs* de requête enregistrent les sessions de requêtes précédentes. Une session requête en RI contextuelle enregistre une séquence requête-contexte-document sous la forme suivante :

Session requête := <requête, contexte> [*clicked documents*]

Chaque session contient une requête, son contexte et un ensemble de documents que l'utilisateur a sélectionnés (par un *click*) (appelés *clicked* documents). L'idée générale du modèle est que si un ensemble de documents est souvent sélectionné pour des requêtes similaires dans des contextes similaires alors les termes de ces documents sont strictement liés aux termes des requêtes et aux éléments du contexte. De même, si des requêtes similaires et des contextes similaires sont fréquemment co-occurrents dans les *logs*, les termes de la requête sont bien corrélés aux éléments du contexte. Ainsi des corrélations probabilistes entre les termes de la requête, les éléments du contexte et les termes des documents peuvent être établies sur la base des *logs* de requêtes, comme on peut le constater sur la figure 7.1.

L'information mutuelle est utilisée pour déterminer les degrés de corrélation entre les termes de la requête, les éléments contextuels et les termes des documents (figure 7.2). L'information mutuelle est une mesure de dépendance statistique entre deux variables aléatoires basée sur l'entropie de Shannon. Il s'agit d'un score d'association de deux mots (x, y) noté IM qui permet de comparer la probabilité d'observer ces deux mots ensemble avec la probabilité de les observer séparément. Selon [CHU 90],

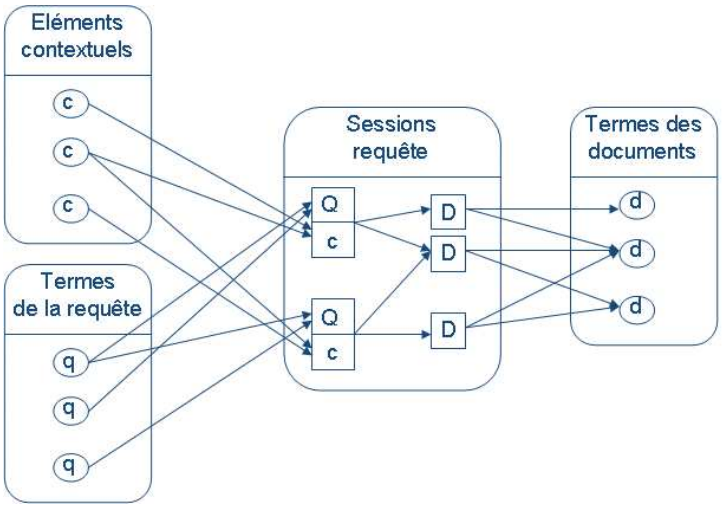

Figure 7.1. *Les sessions de requête construisent les corrélations entre les termes de la requête, les éléments contextuels et les termes des documents*

la définition du score IM est la suivante : $IM(x,y) = log_2(P(x,y)/(P(x)P(y)))$ où $P(x)$ et $P(y)$ sont les probabilités d'observer les mots x et y, et $P(x,y)$ est la probabilité de les observer simultanément. Si $IM(x,y)$ est fortement positive, cela signifie que x et y apparaissent très souvent ensemble. Si $IM(x,y)$ est proche de 0, alors x et y n'ont aucun rapport et enfin, si $IM(x,y)$ est fortement négative, alors x et y ont des distributions complémentaires.

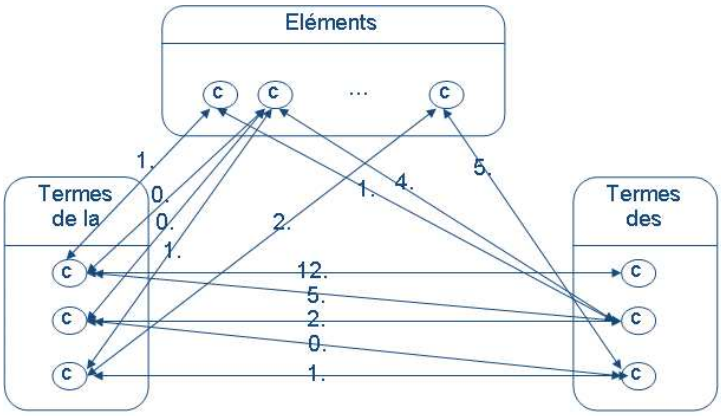

Figure 7.2. *Information mutuelle entre les termes de la requête, les éléments contextuels et les termes des documents*

Dans la suite et sur la base de ces corrélations, nous présentons quatre modèles probabilistes qui ont été proposés pour générer des expansions de termes pour la recherche contextuelle.

7.4.3.1. *Modèle orienté contexte*

Le premier modèle est assez simple et intuitif : les termes des documents bien corrélés au contexte sont choisis comme expansion de termes pour modifier la requête initiale :

$M_1(d \triangleleft Q, C >) = I(d, C) = \sum_i I(d, c_i)$
où $I(d, C)$ est l'information mutuelle entre le contexte et les termes du document. Du point de vue de la combinaison de la requête et du contexte, ce modèle est similaire à la solution traditionnelle en recherche contextuelle. Cependant, ici, les termes utilisés pour l'expansion de requête ne proviennent pas directement du contexte mais des documents corrélés au contexte.

7.4.3.2. *Modèle indépendant requête-contexte*

Le principal défaut du premier modèle est que les termes de l'expansion sont générés uniquement à partir du contexte et ne sont donc pas corrélés à la requête. Le deuxième modèle utilise la requête et le contexte pour contrôler le processus d'expansion de requête :

$M_2(d \triangleleft Q, C >) = I(d, < Q, C >)$

$= I(d, C) + I(d, Q)$

$= \sum_i I(d, c_i) + \sum_i I(d, q_i)$
où $I(d, Q)$ est l'information mutuelle entre la requête et les termes du document.

L'avantage du deuxième modèle par rapport au premier réside dans le fait qu'il affecte des poids plus élevés aux termes des documents qui sont corrélés à la fois à la requête et au contexte.

7.4.3.3. *Modèle dépendant requête-contexte*

Pour diminuer l'effet de l'hypothèse d'indépendance entre la requête et le contexte, un troisième modèle a été introduit pour prendre en compte les relations de dépendance entre la requête et le contexte. Ce troisième modèle est défini par :

$M_3(d \triangleleft Q, C >)$

$= I(d, < Q, C >)$

$= \sum_i I(d, c_i) + \sum_j I(d, q_j) + \sum_{ij} I(d, < q_j, c_i >)$
où $\sum_{ij} I(d, < q_j, c_i >)$ est l'information mutuelle entre un terme de document et

une paire requête-contexte. C'est ce facteur qui introduit une dépendance requête-contexte. Le paramètre est introduit pour ajuster le poids de ce facteur de dépendance requête-contexte.

7.4.3.4. *Modèle filtrant le contexte*

Un problème commun entre les trois premiers modèles est que le bruit dans un contexte n'est pas traité. Manifestement, le bruit du contexte peut facilement produire des expansions de termes hors de propos. Ainsi, le quatrième modèle utilise l'information mutuelle entre requête et contexte pour éliminer le bruit dans les éléments du contexte :

$M_4(d \triangleleft Q, C >) = I(d, < Q, C >) = I(d, < Q, C' >) = \sum_i I(d, c_i) + \sum_j I(d, q_j) + \sum_{ij} I(d, < q_j, c_i >)$
où $C' = \{c/c \in C, I(c, Q) \geq \alpha\}$.

$I(c, Q)$ est l'information mutuelle entre la requête et un élément contextuel. Elle est utilisée pour filtrer des éléments contextuels non corrélés à la requête courante, α représente le seuil pour le processus de filtrage.

La méthode d'expansion de requête est basée sur les modèles probabilistes décrits ci-dessus. Lorsqu'une nouvelle requête avec contexte est soumise, les termes des documents corrélés sont sélectionnés et classés selon leur probabilité conditionnelle obtenue par les modèles. Enfin les meilleurs termes classés peuvent être sélectionnés comme termes d'expansion.

7.4.4. *Extension du modèle vectoriel*

M. Melucci [MEL 05] propose de modéliser le contexte par un vecteur. Plus précisément, le contexte est modélisé par des bases d'espaces de vecteurs et son évolution est modélisée par des transformations linéaires d'une base à une autre. Chaque document ou requête peut être associé à une base distincte, qui correspond à un contexte.

Avant de présenter la contribution de M. Melucci, nous rappelons ci-dessous les éléments principaux du modèle vectoriel.

Dans le cadre du modèle vectoriel classique, un contexte unique est considéré pour le document et la requête. Soit $\{t_1, \ldots, t_n\}$ une liste d'autorités, une base T est modélisée telle qu'il existe un vecteur de T pour chaque terme. Chaque document et chaque requête est vectorisé par T : $Z = \sum_{i=1}^{n} a_i t_i$ où Z est le document ou la requête, t_i est le terme et a_i le poids de t_i pour générer Z. Le classement des documents trouvés est basé sur le calcul suivant où x représente un document, y une requête et $T^T \cdot T$ la matrice de corrélation :

$$x \cdot y = (T \cdot a)^T \cdot (T \cdot b) = a^T \cdot (T^T \cdot T) \cdot b$$

Les hypothèses du modèle vectoriel contextuel sont les suivantes : une base de vecteurs modélise un descripteur de document ou de requête. La sémantique d'un descripteur de document ou de requête dépend du contexte. Une base peut être dérivée d'un contexte. De ce fait, une base de vecteurs modélise le contexte.

En d'autres termes, chaque document et chaque requête sont associés à des bases distinctes et les liens d'une base à une autre sont gouvernés par une transformation linéaire. Par exemple, soit $n = 2$ et t_1, t_2 deux descripteurs, disons *Padoue* et *Venise*, respectivement. Si un utilisateur recherche des informations sur *Padoue* et voyage autour de cette ville, une requête q contient t_1 avec le poids 1 ; si les vecteurs représentant les descripteurs sont supposés orthonormaux, alors des vecteurs peuvent être utilisés et $q = [1, 0]^T$. Si une information contextuelle est disponible relativement à l'espace, par exemple, l'utilisateur s'est déplacé jusqu'à *Venise*, les descripteurs ne peuvent plus être considérés comme orthonormaux et l'ensemble de vecteurs $\{[1, 2]^T, [3, 1]^T\}$ peut être utilisé comme base, et ainsi $q = [1, 3]^T$. Toute base peut être projetée dans une autre par une transformation linéaire. Ainsi, le contexte influence non seulement le choix d'un descripteur mais aussi sa sémantique et la manière dont il se réfère aux autres descripteurs.

Le vecteur x du document x écrit dans son propre contexte est généré par la base T qui n'est pas nécessairement égal à la base U qui génère un vecteur requête y ou à la base T' qui génère un autre vecteur document. Ainsi, x est représenté par $x = T \cdot a$ alors que y est représenté par $y = U \cdot b$ où a et b sont les coefficients utilisés pour combiner les bases de vecteurs de T et U respectivement. Si la pertinence est estimée par le produit scalaire classique, les documents sont rangés grâce à la formule : $x^T \cdot y = a^T \cdot (T^T \cdot U) \cdot b$. Cette dernière formule montre que la relation entre les descripteurs utilisée pour exprimer les documents et la requête dépendent de deux contextes : le premier est impliqué dans l'indexation du document et le deuxième dans la formulation de la requête.

M. Melucci propose également un modèle pour l'évolution du contexte basé sur le principe suivant : tout changement de contexte peut être modélisé par des transformations linéaires d'une base vers une autre.

7.5. Les systèmes d'accès contextuel à l'information

Les moteurs de recherche du *web* traitent généralement des requêtes isolées. Les résultats pour une requête donnée sont identiques, indépendants de l'utilisateur, ou du contexte dans lequel l'utilisateur pose sa requête. Les informations du contexte peuvent être fournies par l'utilisateur sous la forme de mots-clés ajoutés à une requête, par exemple un utilisateur recherchant la page web personnelle d'un individu devrait

ajouter des mots-clés comme *home* ou *homepage* à la requête. Cependant, fournir un contexte sous cette forme est difficile et limitée. Une solution pour ajouter des informations bien définies du contexte à une requête consiste pour le moteur de recherche à demander spécifiquement de telles informations. On peut classer les systèmes de recherche d'information contextuelle selon que le contexte est explicitement demandé à l'utilisateur ou automatiquement inféré.

7.5.1. *Expression explicite du contexte*

Le projet INQUIRUS 2 [GLO 99], [GLO 00] requiert les informations du contexte sous la forme d'une catégorie d'informations désirées. Les utilisateurs doivent choisir une catégorie telle que « pages personnelles », « papiers de recherche », « événements actuels » ou « introduction générale » lorsqu'ils posent leur requête sous forme de mots-clés. INQUIRUS 2 est un méta-moteur de recherche qui exploite la requête exprimée explicitement par l'utilisateur ainsi qu'un ensemble de préférences pour réécrire la requête et identifier le moteur de recherche à utiliser. Les informations du contexte sont utilisées pour sélectionner les moteurs de recherche à qui envoyer les requêtes, pour modifier les requêtes et pour sélectionner la politique d'ordonnancement des résultats. Le contexte correspond donc à une requête reformulée, adaptée à la catégorie de recherche. Par exemple, une requête qui recherche des articles sur *machine learning* pourra envoyer des requêtes multiples à des moteurs de recherche. Les termes ajoutés à la requête si on recherche des papiers de recherche seront *abstract, keywords, introduction.*

7.5.2. *Génération automatique des informations du contexte*

INQUIRUS 2 a permis d'améliorer de manière satisfaisante la précision de la recherche, mais impose à l'utilisateur de rentrer explicitement les informations du contexte. L'objectif du projet Watson [BUD 00] est d'inférer automatiquement les informations du contexte. Watson vise à modéliser le contexte de l'utilisateur en se basant sur le contenu des documents édités par Microsoft Word, ou visualisés par Internet Explorer. Les documents que les utilisateurs éditent ou visualisent sont analysés à l'aide d'un algorithme de termes pondérés, qui vise à identifier les mots qui donnent une indication sur le contenu du document. Des informations comme la taille de la police sont également utilisées pour assigner un poids aux termes. Si un utilisateur exprime une requête explicite, Watson modifie la requête en se basant sur le contenu du document édité ou visualisé, et envoie la requête modifiée vers des moteurs de recherche, ajoutant ainsi automatiquement des informations du contexte à la recherche sur le *web*. Les autres projets similaires sont nombreux : Margin Notes [RHO 00a] qui réécrit les pages *web* pour intégrer des liens vers des fichiers personnels ; le projet *Haystack* [ADA 99] dont le but est de créer une communauté de *Haystack* interactif ou des entrepôts d'informations personnelles ; et le programme Automy's Kenjin

(www.kenjin.com) qui suggère automatiquement du contenu du *web* ou de fichiers locaux, basé sur les documents qu'un utilisateur lit ou édite. Nous pouvons citer également les systèmes suivants qui apprennent les profils d'intérêt de l'utilisateur pour recommander des pages *web* : Fab [BAL 97], Letizia [LIE 95], WebWatcher [ARM 95], Siskill and Weber [PAZ 96], Web Personae [MCG 03] et SIS [DUM 03].

7.5.3. *Deviner les besoins de l'utilisateur*

Une technique en croissance sur le *web* consiste à deviner le contexte de la requête de l'utilisateur. Les moteurs de recherche Excite (www.excite.com), Lycos (www.lycos.com), Google (www.google.com) et Yahoo (www.yahoo.com) intègrent des fonctionnalités spéciales pour certains types de requêtes. Par exemple, des requêtes posées à Excite et Lycos sur le nom d'un artiste ou une entreprise produisent des résultats additionnels qui orientent directement aux informations sur l'artiste ou l'entreprise. Yahoo a ajouté de telles fonctionnalités en 2000, et fournit des résultats spécialisés pour de nombreux types différents de requêtes - par exemple, des noms d'équipes de sport sont reliés à des information d'équipes ou de ligues. D'autres exemples avec Yahoo incluent des modèles de voiture, des célébrités, des musiciens, des villes importantes, des noms de médicaments ou de maladies, des signes du zodiaque, des lignes d'aviation, des shows de télévision et des parcs nationaux. Google identifie des requêtes qui ressemblent une adresse de rue américaine, et fournit des liens directs avec des plans de ville ou de quartier. Plutôt que de demander explicitement à l'utilisateur des informations telles que *je cherche un plan de ville* ou *je cherche le site web d'une équipe de football*, cette technique devine lorsque de tels contextes peuvent être pertinents. Les utilisateurs peuvent relativement aisément identifier des contextes d'intérêt. Cette technique est limitée aux cas où des contextes potentiels peuvent être identifiés à partir de la requête sous forme de mots-clés.

Une autre solution pour ajouter du contexte à une recherche sur le *web* est de restreindre le contexte des moteurs de recherche. Des milliers de ces moteurs de recherche existent déjà. Nous pouvons citer www.invisibleweb.com et www.completeplanete.com.

7.6. Evaluation des modèles et techniques de recherche d'information contextuelle

Les protocoles d'évaluation largement adoptés en RI sont empiriques et souvent basés sur une évaluation d'avantage quantitative que qualitative. En effet, les résultats obtenus sont issus de la comparaison de mesures et de métriques en termes de rappel-précision, sur les réponses fournies par le système relativement à celles issues des réponses attendues qui constituent le référentiel. L'introduction de la dimension utilisateur dans le processus d'accès à l'information pose de nouveau la question de

l'évaluation. Cette section présente la problématique de l'évaluation de l'efficacité des modèles et algorithmes d'accès contextuel à l'information puis présente un aperçu des protocoles adoptés par les travaux du domaine.

7.6.1. *Le modèle d'évaluation Cranfield : apports et limites*

Le modèle d'évaluation Cranfield [CLE 67] est incontestablement le modèle de référence adopté dans les campagnes d'évaluation en RI. Ce modèle fournit une base d'évaluation comparative de l'efficacité de différents algorithmes, techniques et/ou systèmes moyennant des ressources communes : collections de tests contenant des documents, des requêtes préalablement construites et des jugements de pertinence associés construits selon la technique de *pooling*, des métriques d'évaluation essentiellement basées sur le rappel-précision. L'émergence de la RI orientée utilisateur a cependant remis en cause la viabilité de ce modèle pour l'évaluation de systèmes interactifs ou de manière générale, les systèmes d'accès contextuel à l'information [ING 05]. Parmi les principales limites du modèle de Cranfield dans un tel cadre, on cite notamment les suivantes [ING 05, KEK 05] :

1) la construction des collections de tests est basée sur des jugements de pertinence thématique des assesseurs. Les autres types de pertinence liées au contexte de l'utilisateur (centres d'intérêts, familiarité, but ou tâche de la recherche, etc.) ainsi que les différents degrés de pertinence (très pertinent, pertinent, marginalement pertinent, etc.) ne sont pas pris en compte ;

2) les requêtes sont soumises en mode *batch*. L'interaction utilisateur-SRI n'est pas considérée dans l'évaluation de l'efficacité de la recherche ;

3) l'évaluation de la pertinence est fondée sur l'hypothèse d'indépendance des documents ; or la réapparition de documents pertinents mais similaires a un effet sur le jugement de l'utilisateur ;

4) les mesures de rappel et précision sont adaptées à l'évaluation de l'efficacité de la recherche du seul point de vue de la pertinence binaire et thématique ;

5) les mesures d'évaluation agrégées sur les requêtes ne considérent pas les variations des performances selon la dimension utilisateur.

7.6.2. *Vers un modèle d'évaluation adapté à la recherche d'information contextuelle ?*

Compte tenu des précédentes critiques du modèle Cranfield quant à l'évaluation de l'accès contextuel à l'information, de nombreux travaux ont initié de nouveaux protocoles d'évaluation basés sur des mesures d'évaluation, collections de tests et scénarios appropriés. Un large aperçu des résultats obtenus est présenté ci-dessous.

7.6.2.1. *Les prémices dans TREC*

Les premières initiatives de l'intégration des facteurs contextuels de l'utilisateur pour l'évaluation de l'efficacité de l'accès à l'information se sont soldées par des tâches dans TREC en l'occurrence *interactive track* et *hard track*.

– La tâche interactive [HAR 95]. Cette tâche a été menée dans TREC depuis 1995 (TREC-4) jusqu'en 2002 (TREC-10). Elle a eu pour double objectif le développement de méthodologies appropriées à l'évaluation de processus de recherche d'information interactive et la mesure de l'impact des différentes caractéristiques des utilisateurs dans l'évaluation de la pertinence des résultats. A cet effet, des questionnaires et des interviews sont établis au préalable pour décrire les utilisateurs participant effectivement à la campagne d'évaluation. Les principales caractéristiques recueillies concernent leur familiarité avec les sujets de la requête, leur expertise dans l'utilisation de moteurs de recherche, leur habilité à interagir avec le système, leur démarche pour mener une tâche de recherche d'information, etc. Les collections de documents n'ont pas subi de changements conséquents relativement aux campagnes TREC précédentes. Les mesures classiques de rappel-précision sont généralement utilisées pour mesurer l'effet de certaines variables isolées (liées aux caractéristiques citées ci-avant) sur les performances de recherche. A TREC-7 une nouvelle mesure est proposée, en l'occurrence le rappel au niveau instance, qui mesure pour chaque utilisateur le nombre d'instances de réponses correctes trouvées pour une question donnée sur un intervalle de temps déterminé (15 à 20 minutes).

– La tâche *HARD* [HAR 03]. Cette tâche a été menée dans TREC depuis 2003 (TREC-12) jusqu'en 2005 (TREC-14). Son objectif est d'atteindre les performances de haute précision pour des utilisateurs spécifiques. A cet effet, et à la différence de la tâche interactive, la tâche a introduit (1) l'utilisation de méta-données dans les documents et requêtes de la collection tests qui décrivent le contexte de recherche, on cite notamment : familiarité, genre et granularité (2) une pertinence graduelle (3) des mesures d'évaluation qui considèrent les niveaux de pertinence pour l'évaluation des performances de recherche ; d'autres informations additionnelles sont éventuellement demandées aux utilisateurs participant à la campagne d'évaluation à l'aide de formulaires de clarification. L'objectif de l'évaluation est alors de mesurer l'efficacité d'un système à fournir les réponses précises en fonction du contexte de la recherche.

7.6.2.2. *Les mesures d'évaluation*

Dans ce qui suit, une présentation des principales mesures d'évaluation ayant émergé dans les travaux de référence sur l'évaluation de systèmes d'acès interactif à l'information.

1) La mesure RR (*Relative Relevance*). La mesure RR [BOR 98] a pour objectif de considérer différents types de pertinence (pertinence non binaire, voir paragraphe 7.3.4.) dans l'évaluation de l'efficacité d'un système d'accès contextuel à l'information. Cette mesure quantifie le degré de concordance entre les types de jugement de

pertinence émis dans le cas de deux ensembles de jugements (soit R_1 et R_2) associés à une même liste de documents qui constitue les résultats d'une session de recherche. En pratique, R_1 correspond généralement aux scores de pertinence algorithmique retournés par un SRI et R_2 à des scores de pertinence contextuelle correspondant à un type de pertinence donné : situationnelle si elle est exprimée par un utilisateur, thématique si elle est exprimée par un assesseur, etc. La valeur de corrélation entre R_1 et R_2 est généralement calculée en utilisant une mesure du cosinus ; elle quantifie globalement la capacité du système à prédire le type de pertinence contextuelle considéré. A la différence de la mesure classique de précision, cette mesure permet de considérer les différents types de pertinence ; néanmoins, elle pose un problème lors de l'évaluation comparative entre différents algorithmes de recherche, voire entre différents SRI [BOR 03]. En effet les scores de pertinence algorithmique ne sont pas étalonnés à la même échelle entre différents SRI, ce qui rend la comparaison de mesures RR non significative.

2) La mesure RHL (*Ranked Half-Life*).
La mesure RHL [BOR 98] est une mesure orientée position, apportant une solution au problème de l'évaluation comparative posée par la mesure RR. L'idée de base est exploitée en bibliométrie [EGG 90] où la mesure connue sous le nom de *cited half-life*, mesure le temps mis pour que la motié des citations référencent un document donné. Dans le cas de l'évaluation d'un processus de RI contextuelle, le rang du document se substitue à la dimension temps. L'objectif de cette mesure est alors de déterminer, pour un SRI donné, le rang à partir duquel la moitié du nombre total de documents pertinents (ou de la quantité d'information pertinente dans le cas d'une pertinence graduelle) est présentée à l'utilisateur. Plus la mesure RHL est petite, plus les documents les plus pertinents sont en début de liste et donc, plus le SRI est efficace, en ce sens que l'ordre des documents retournés s'accorde avec l'ordre de pertinence contextuelle de l'utilisateur. La formule de calcul de la mesure RHL est la suivante :

$$RHL = R_m + \left(\frac{P\_tot/2 - \sum_{R<R_m} P(R)}{P(R_m)} \right) \qquad [7.1]$$

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

La mesure RHL affine la mesure de précision classique en ce sens qu'elle considère en plus, le degré de pertinence (pertinence graduelle) des documents retournés par un SRI et par conséquent le degré de correspondance de la pertinence algorithmique avec la pertinence exprimée effectivement par l'utilisateur. Par ailleurs les mesures RR et RHL sont complémentaires : la première couvre la pertinence horizontalement en en considérant différents types, tandis que la seconde couvre la pertinence verticalement en en considérant différents degrés [BOR 03].

3) Les mesures CG *(Cumulative Gain)* et DCG *(Discount Cumulative Gain)*. Les mesures CG et DCG [JAR 00, JAR 02] sont des mesures orientées position définies dans le contexte d'une pertinence graduelle et dont l'objectif est d'estimer le gain de l'utilisateur en termes de pertinence cumulée en observant les documents situés jusqu'à un rang donné. Ces mesures sont définies comme suit :

$$CG[i] = \begin{cases} G[1], si\ i = 1 \\ CG[i-1] + G[i],\ sinon \end{cases} \quad [7.2]$$

où $G[i]$ est la valeur de pertinence associée au document de rang i :

$$CG[i] = \begin{cases} G[1], si\ i = 1 \\ CG[i-1] + G[i]/log\ i,\ sinon \end{cases} \quad [7.3]$$

Comparativement à la mesure CG, la mesure DCG permet d'atténuer le gain de pertinence apporté par un document en fonction du rang associé. Ceci rejoint en effet l'hypothèse évidente que plus le rang d'un document est élevé, moins il est probable que l'utilisateur l'examine et donc moins il est à l'origine d'un gain effectif de pertinence. Relativement à la mesure RHL, les mesures CG et DCG offrent l'avantage de donner une estimation du gain de pertinence à n'importe quel rang, indépendamment de la base de rappel ; cette dernière doit être en effet préalablement fixée, dans le cas de la mesure RHL, pour définir le rang de la pertinence médiane [JAR 00].

4) La mesure GRP *(Generalised Recall and Precision)*. La mesure GRP [JAR 02] est également une mesure orientée position qui généralise les mesures classiques de rappel-précision en considérant une pertinence graduelle. Le rappel généralisé (GR) et précison généralisée (GP) sont calculées comme suit :

$$gP = \sum_{d \in R} r(d) / |R| \quad [7.4]$$

$$gR = \sum_{d \in R} r(d) / \sum_{d \in D} r(d) \quad [7.5]$$

où R est l'ensemble des documents retournés par le SRI, D est l'ensemble des documents de la collection, $r(d)$ est la valeur de pertinence graduelle associée au document d. De manière analogue aux mesures classiques de rappel-précison, ces mesures offrent la possibilité d'être agrégées pour plusieurs requêtes ou plusieurs niveaux de rappel et donnent ainsi la possibilité de tracer des courbes de performances.

7.6.2.3. *Les collections de tests*

Hormis les collections de tests construites pour la tâche *HARD* de TREC, aucune collection de tests standard n'a été construite à notre connaissance pour évaluer l'efficacité de l'accès contextuel à l'information. De telles collections contiendraient divers éléments du contexte liés à l'utilisateur directement (historique des recherche, centres d'intérêts, expertise, etc.) ou à la session de recherche (but de la recherche, tâche, etc.). Un tel défi n'est pas encore levé à ce jour.

La littérature (relativement récente) fait état de deux principaux protocoles de construction de collections de tests :

1) réutilisation des collections de tests de TREC (documents, requêtes et jugements de pertinence) puis leur augmentation par des éléments du contexte. Ces éléments, tels que l'historique des interactions, sont extraits à partir des interactions d'utilisateurs effectifs interrogeant la base TREC à l'aide de requêtes TREC [SHE 05a]. Le référentiel d'évaluation étant disponible, les mesures agrégées de rappel-précision sont alors exploitées pour évaluer les différences de performances entre les scénarios de recherche basique (ne tenant pas compte du contexte) et scénario de recherche contextuelle ;

2) construction de collections de tests en menant une campagne d'évaluation : c'est le protocole adopté par la plupart des travaux. Un ensemble d'utilisateurs est identifié et un ensemble de requêtes est construit. On collecte un volume de données-tests issu du croisement des jugements de pertinence pour chaque utilisateur spécifique et chaque requête. Les scénarios d'évaluation consistent alors, de manière classique, à comparer les performances de recherche d'un moteur de recherche classique et d'un moteur de recherche augmenté par la technologie basée sur l'exploitation des contextes des utilisateurs assesseurs de jugements [LIU 04, GAU 03]. Dans le cas de l'utilisation de mesures agrégées de rappel-précision, un référentiel est généralement construit sur la base de l'ensemble des documents pertinents jugés par l'ensemble des utilisateurs pour chaque requête. Autrement, l'utilisation de mesures orientées rang évite l'utilisation d'un tel référentiel.

7.7. Conclusion

Dans ce chapitre, nous avons fait le point de l'état de l'art sur la RI contextuelle sur le *web*. Il s'agit d'un domaine de recherche récent dont l'émergence provient d'une part de l'évolution des supports physiques (PDA, téléphonie mobile) et d'autre part de l'accroissement du volume d'information sur le *web*. Les travaux de recherche actuels portent sur la modélisation du contexte, l'intégration du contexte dans les modèles de RI, le développement de SRI contextuelle et les méthodes d'évaluation de ces systèmes.

Nous pouvons d'ores et déjà établir un certain nombre de conclusions :

– il existe une multitude de définitions du contexte dans la littérature. Nous avons proposé une synthèse de ces travaux et montré la différence entre contexte et profil ;

– nous avons montré les liens entre RI adaptative, RI personnalisée et RI contextuelle ;

– de nombreux modèles de RI contextuelle ont été définis ces dernières années dont des extensions des modèles classiques (vectoriel, probabiliste) ;

– les moteurs de recherche du *web* utilisent depuis plusieurs années la notion de contexte. En particulier, Google propose une solution personnalisée pour accéder aux documents du *web* (*Google Personalized Search*) ;

– des propositions existent déjà pour évaluer les systèmes de RI contextuelle. En particulier, la tâche interactive de TREC et la tâche *HARD* constituent des premières initiatives pour intégrer des facteurs contextuels de l'utilisateur. Il existe également des mesures d'évaluation adaptée à la RI contextuelle dont la mesure RR (*Relative Relevance*) et la mesure RHL (*Ranked Half-Life*).

Les efforts à poursuivre dans le domaine concernent prioritairement l'évaluation des SRI contextuelle. Aucune collection de tests standard n'a été construite pour évaluer l'efficacité de tels systèmes. Un autre point à approfondir porte sur la modélisation du contexte. Une définition standardisée des éléments du contexte pourrait être proposée. Une troisième perspective concerne les modèles de RI contextuelle. La plupart des propositions ne s'intéressent qu'à une dimension du contexte et il s'agit le plus souvent des centres d'intérêts de l'utilisateur. Les nouveaux modèles doivent prendre en compte toutes les dimensions du contexte (préférence utilisateur, temps, espace, etc.). Les extensions des modèles classiques permettent déjà la prise en compte de plusieurs éléments contextuels.

Les résultats présentés dans ce chapitre montrent que la RI contextuelle constitue actuellement un domaine de recherche très actif. D'ailleurs, le Workshop IRiX (*Information Retrieval in Context*) est organisé conjointement à la conférence internationale SIGIR depuis l'année 2004. De nombreuses propositions très prometteuses y sont présentées et visent l'amélioration de la RI sur le *web*.

7.8. Bibliographie

[ADA 99] ADAR E., KARGER D., ANDREA STEIN L., « Haystack : Per-user information environments », *Proceedings of the 1999 Conference on Information and Knowledge Management (CIKM)*, 1999.

[ALL 97] ALLEN B., « Information seeking in context », *Proceedings of an International Conference on Research in needs, seeking and use in different context*, p. 111-122, 1997.

[ARM 95] ARMSTRONG R., FREITAG D., JOACHIMES T., MITCHELL L., « WebWatcher : A learning apprentice for the World Wide Web », *AAAI Spring Symposium on Information Gathering from Heterogeneous, distributed environments*, 1995.

[BAL 97] BALABANOVIC M., SHOHAM Y., « Fab : content-based, collaborative recommendation », *Communications of the ACM*, vol. 40, n° 3, p. 66-72, 1997.

[BRO 77] BROOKES B. C., « The developping cognitive viewpoint in information science », *De Mey M. et al. (dir.)*, p. 195-203, septembre 1977.

[BEA 97] BEAULIEU M., « Experiments with interfaces to support query expansion », *Journal of documentation*, vol. 53, n°1, p. 8-19, 1997.

[BEL 78] BELKIN N., « Information concepts for information science », *Journal of documentation*, vol. 34, p. 55-85, 1978.

[BEL 87] BELKIN N., « Knowledge elicitation using discourse analysis », *International Journal of Man-Machine studies*, vol. 27, p. 127-144, 1987.

[BEL 96] COOL C., PARK S., BELKIN N.J., KOENEMANN J., « Information seeking behaviour in new searching environments », *Ingwersen, P., Pors*, p. 403-416, 1996.

[BEL 01] BELKIN N.J., COOL C., KELLY D., LIN S.J, PARK S.Y., PEREZ-CARBALLO J., SIKORA C. , « Iterative exploration, design and evaluation of support of query reformulation in interactive information retrieval », *Information Processing and Management*, vol. 37, n° 3, p. 404-434, 2001.

[BEO 63] BEORKO H., BERNICK M., « Automatic document classification », *Journal of the Association for Computing Machinery*, p. 151-161, 1963.

[BOR 98] BORLUND P., INGWERSEN P., « Measures of relative relevance and ranked half-life : performance indicators in interactive IR », *Croft W.B et al.(dir.), Proceedings of the* 21^{st} *ACM SIGIR International Conference on Research and Development*, p. 324-331, août 1998.

[BOR 03] BORLUND P., « The IIR evaluation model : A framework for evaluation of interactive information retrieval systems », *Journal of Information Research*, vol. 8, n° 3, p. 152-179, 2003.

[BRO 02] BROWN P.J., JONES G., « Exploiting Contextual Change in Context-Awware Retrieval », *Proceedings of the* 17^{th} *ACM Symposium on Applied Computing (SAC2002)*, Madrid, p. 650-656, 2002.

[BUC 94] BUCKLEY C., SALTON G., JAMES A., SINGHAL A., « Automatic query expansion using SMART : TREC-3 », *Proceedings of the* 3^{rd} *Text REtrieval Conference*, p. 69-80, 1994.

[BUC 95] BUCKLEY C., « New retrieval approaches using SMART : TREC 4 », *Proceedings of the* 4^{th} *Text REtrieval Conference*, p. 25-48, 1995.

[BUD 00] BUDZIK J., HAMMOND K., « User interactions with everyday applications as context for just-in-time information access », *Intelligent User Interface*, p. 41-51, 2000.

[CHA 04] CHALLAM V. K. R., Contextual information retrieval using ontology based user profiles, Master's Thesis, 2004.

[CHU 89] CHURCH K., GALE W., HANKS P., HINDLE D., « Parsing, word associations and typical predicated-argument relations », *Proceedings of the 1989 DARPA speech and natural language Workshop*, 1989.

[CHU 90] CHURCH K.W., HANKS P., « Word Association Norms, Mutual Information and Lexicography », *Computational Linguistics*, p. 22-29, 1990.

[CLE 67] CLEVERDON C., « The Cranfield test on index language devices », *Aslib*, p. 173-194, 1967.

[COP 03] COPOLLA P., DELLA MEA V., DI GASPERO L., MIZZARO S., « The Concept of Relevance in Mobile and Ubiquitous Information Access », *Proceedings of the Mobile HCI*

2003 Information Workshop on Mobile and Ubiquitous Information Access, Udine, p. 1-10, 2003.

[CRO 79] CROFT W, HARPER D., « Using probabilistic models for information retrieval without relevance information », *Journal of documentation*, vol. 35, n° 4, p. 285-295, 1979.

[CRO 95] CROFT B., « What do people want from information retrieval », *D-Lib Magazine*, 1995.

[DEM 77] DE MEY M., « The cognitive viewpoint : Its development and its scope », *International Workshop on the cognitive viewpoint*, p. 285-295, 1977.

[DUM 03] DUMAIS S., CUTTRELL E., DUMAIS J.J., JANCKE G., SARIN R., ROBBINS D.C., « Stuff I've seen : a system for a personal information retrieval and re-use », *Proceedings of the* 26^{th} *ACM SIGIR*, Toronto, p. 72-79, juillet 2003.

[EFT 96] EFTHIMIADIS E., « Query expansion », *Annual Review of Information Science and Technology (ARIST)*, vol. 31, p. 121-187, 1996.

[EGG 90] EGGHE L., ROUSSEAU R., « *Introduction to infometrics : Quantitative methods in library, documentation and information science* », Elsevier, 1990.

[FAN 04] FAN W., GORDON M., PATHAK P., « Discovery of context specific ranking functions for effective information retrieval using genetic programming », *IEEE Transactions on knowledge and data engineering*, vol. 16, n°4, p. 523-527, 2004.

[FID 91] FIDEL R., « Searchers'selection of search keys », *Journal of the American Society of Information Science (JASIS)*, vol. 34, 1991.

[FIN 02] FINKELSTEIN L., « Placing Search in Context : The Concept Revisited », *ACM Transactions on Information System*, vol. 20, n° 1, p. 116-131, 2002.

[FUH 00] FUHR N., *Information Retrieval : introduction and survey*, Post-Graduate course on Information retrieval, University of Duisburg-Essen, Allemagne, 2000.

[GAU 03] GAUCH S., CHAFFE J., PRETSCHNER P., « Ontology based user profiles for search and browsing », 2003.

[GLO 99] GLOVER E.J., LAWRENCE S., GORDON M.D., BIRMINGHAM W.P. ,LEE GILES C., « Architecture of a metasearch engine that supports user information needs », *Proceedings of CIKM'1999*, p. 210-216, novembre 1999.

[GLO 00] GLOVER E.J., LAWRENCE S., « Web search - your way », *CACM*, 2000.

[HAR 92a] HARMA D., « Overview of the the 1^{st} text retrieval conference (TREC-1) », *Proceedings of the* 1^{st} *text retrieval conference (TREC-1)*, National Institute of Standards and Technology, NIST special publication, p. 1-20, 1992.

[HAR 92b] HARMAN D., « Relevance feedback revisited », *Proceedings of the* 15^{th} *ACM SIGIR International Conference on Research and Development*, N. Belkin, P. Ingwersen, A. Mark Pejtersen (dir), p. 1-10, 1992.

[HAR 95] HARMA D., « Overview of the the 1st text retrieval conference (TREC-4) », *Proceedings of the* 1^{st} *text retrieval conference (TREC-4)*, National Institute of Standards and Technology, NIST special publication, p. 1-24, 1995.

[HAR 03] HARMA D., « Overview of the the 1^{st} text retrieval conference (TREC-12) », *Proceedings of the 1^{s} text retrieval conference (TREC-12)*, National Institute of Standards and Technology, NIST special publication, p. 1-12, 2003.

[HEU 99] HEUER R. J., « Psychology of intelligence analysis », *Center for the study of intelligence, Central intelligence agency*, 1999.

[HSI 93] HSIEH-YEE I., « Effects of the searcher and experience and subject knowledge on the search tactics of novice and exprienced searchers », *Journal of the American Society for Information Science (JASIS)*, vol. 44, n°3, p. 161-174, 1993.

[HUL 02] HULL R., KUMAR B., SAHUGUET A., MING X., « Have It Your Way : Personalization of Network-Hosted Services », *BNCOD*, p. 1-10, 2002.

[ING 94] INGWERSEN P., « Polyrepresentation of information needs and semantic entities : Elements of a cognitive theory of information retrieval and interaction », Croft W.B, Van Risjsbergen, C.J (dir.), *Proceedings of the 17^{th} ACM SIGIR International Conference on Research and Development*, p. 101-111, août 1994.

[ING 96] INGWERSEN P., « Cognitive perspectives of information retrieval interactions : Elements of a cognitive IR Theory », *Annual review of information science and technology*, vol. 52, n° 1, p. 3-50, 1996.

[ING 04] INGWERSEN P., JARVELIN K., « Information retrieval in context », *In Proceedings of the 27^{th} ACM SIGIR Worshop on information retrieval in context*, p. 6-8, juillet 2004.

[ING 05] INGWERSEN P., JARVELIN K., *The turn Integration of information seeking and retrieval in context*, Springer, 2005.

[JAN 91] JANES J., « Relevance judgements and the incremental presentation of document representation », *Information Processing and Management*, vol. 27, n° 6, p. 629-646, 1991.

[JAR 94] JARVELIN K., KRISTENSEN J., « A deductive data model for query expansion », *Proceedings of the 19^{th} annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Frei, HP. et al. (dir.), p. 235-249, 1994.

[JAR 00] JARVELIN K., KEKALAINEN J., « IR evaluation methods for highly relevant documents », *Proceedings of the 23^{rd} annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Belkin et al., p. 41-48, 2000.

[JAR 02] JARVELIN K., KEKALAINEN J., « Cumulative gain-based evaluation of IR techniques », *ACM Transactions on Information Systems (ACM TOIS)*, vol. 20, n°4, p. 422-446, 2002.

[JEH 03] JEH G., WIDOM J., « Scaling personalized Web search », *Proceedings of the 12^{th} International World Wide Web Conference*, 2003.

[JIN 94] JING Y., CROFT W., An association thesaurus for information retrieval, Technical report TR-1994-17, University of Massachussets, Dept of computer sciense, 1994.

[JON 00] JONES G.J, BROWN P., « Information access for context-aware appliances », *Proceedings of SIGIR'2000*, p. 382-384, 2000.

[JAR 86] JÄRVELIN K., « On information, information technology and the development of society : An information science perspective », Ingwersen P., Kajberg L., Pejtersen A.M

(dir.), p. 35-55, 1986.

[KAN 04] KANG I., KIM G., « Integration of multiple evidences based on a qury type for web search », *Information Processing and Management (IPM)*, vol. 40, n° 3, p. 459-478, 2004.

[KEK 05] KEKALAINEN J.H., JARVELIN K., « Evaluating information retrieval systems under the challenges of interaction and multidimensional dynamic relevance », *Proceedings of the* 4^{th} *Conference on conceptions of Librairy and Information Science (COLIS)*, H. Bruce et al. (dir.), p. 253-270, 2005.

[KEL 04] KELLY N. J., « Understanding implicit feedback and document preference : a naturalistic study », *PHD dissertation*, Ritgers University, Etats-unis, janvier 2004.

[KOB 00] KOBAYASHI M., TAKEDA K., « Information Retrieval on the Web », *ACM Computing Surveys*, vol. 32, n° 2, p. 143-173, juin 2000.

[LAR 03] LARSEN B., LUND H., « Using value-added document representations in INEX », *INEX Workshop proceedings*, p. 67-72, 2003.

[LAW 00] LAWRENCE S., « Context in Web Search », *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering*, 2000.

[LEE 98] LEE J. H., « Combining the evidence of different relevance feedback methods for information retrieval », *Information Processing and Management (IPM)*, vol. 34, n° 6, p. 681-691, 1998.

[LIE 95] LIEBERMAN H., « Letizia : An Agent That Assists Web Browsing », *Proceedings of the International Joint Conference IJCAI*, p. 924-929, août 1995.

[LIM 06] LIMBU D.K., CONNOR A., PEARS R., STEPHEN M., « Contextual Relevance Feddback in Web Information Retrieval », *Information Interaction in Context, IIiX*, Copenhagen Denmark, p. 138-143, 2006.

[LIU 04] LIU F., YU C., « Personalized Web search for improving retrieval effectiveness », *IEEE Transactions on knowledge Data Engineering*, vol. 16, p. 28-40, 2004.

[LOR 06] LORIGO L B., PAN H., HEMBROOKE H., GRANKA L., GERY G., « The influence of task and gender on search and evaluation behavior using Google », *Information Processing and Management (IPM)*, vol. 42, p. 1123-1131, 2006.

[LYM 03] LYMAN P., « How much informations 2003 », octobre 2003.

[MCG 03] MC GOWAN J., The Turn : Integration of Information Seeking and Retrieval in Context, Thesis of Master in Computer Science, Faculty of Science, University College Dublin, 2003.

[MEL 05] MELUCCI M., « Context modeling and discovery using vector space bases », *Proceedings of CIKM 2005*, p. 808-815, 2005.

[MIT 98] MITRA M., SINGHAL A., « Improving automatic query expansion », *Proceedings of the* 21^{st} *annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, p. 206-214, 1998.

[PAZ 96] PAZZANI M.J., MURAMATSU J., BILLSUS D., « Placing Search in Context : The Concept Revisited », *AAAI/IAAI*, vol. 1, p. 54-61, 1996.

[PRE 99] PRETSCHNER A., GAUCH S., « Ontology Based Personalized Search », *Proceedings of the* 11^{th} *IEEE International Conference on Tools with Artificial Intelligence (ICTAI)*, novembre 1999.

[RHO 00a] RHODES B., Just-in-time Information Retrieval, PhD Thesis, Massuchessetts Intitute of Technology, 2000.

[RHO 00b] RHODES B.J., MAES P., « Margin Notes - Building a Contextually Aware Associative Memory », *Proceedings of the International Conference on Intelligent User Interfaces*, New Orleans, LA, 2000.

[RIJ 86] VAN RIJSBERGEN C., « A non-classical logic for Information Retrieval », *The computer journal*, vol. 29, n° 6, p. 481-485, 1986.

[ROB 77] ROBERTSON S., « The probability ranking principle in IR », *Journal of documentation*, vol. 33, n° 4, p. 294-304, 1977.

[ROC 71] ROCCHIO J., *Relevance feedback in Information retrieval*, Prentice Hall, 1971.

[ROD 04] RODE H., HIEMSTRA D., « Conceptual Language Models for Context-Aware Text Retrieval », *Proceedings of TREC-13, NIST Special Publication*, 2004.

[RUT 03] RUTHVEN I., LALMAS M., « A survey on the use of relevance feedback for information access systems », *Kowledge engeneering review*, vol. 18, n° 2, p. 95-145, 2003.

[SAL 68] SALTON G., *Automatic information organisation and retrieval*, McGraw-Hill, New York, 1968.

[SAR 91] SARACEVIC T., SPINK A., WU M.H., « Nature of interaction between users and intermediaries in online searching », Williams M. (dir.), p. 329-341, 1991.

[SAR 97] SARACEVIC T., « The stratefied model of information retrieval interaction : extension and applications », *Proceedings of the* 60^{th} *annual meeting of the American Society for Information Science*, Medford, Etats-Unis, p. 313-327, 1997.

[SHE 05a] SHEN X., TAN B., ZHAI. C., « Context sensitive information retrieval using implicit feedback », *Proceedings of the* 28^{th} *Annual ACM Conference on Research and Development in Information Retrieval SIGIR*, p. 43-58, août 15-19 2005.

[SON 99] SONNENWALD D. H., « Evolving perspectives of human behaviors : contexts, situation, social networks and information horizons », *Exploring the contexts of information behaviour : Proceedings of the* 2^{nd} *international conference on research in information needs, seeking and use in different contexts*, p. 176-190, 1999.

[SU 03] SU J., LEE M., « An exploration in personalized and context-sensitive search », *Proceedings of the* 7^{th} *annual UK special interest group for computatonal linguists research colloquium*, 2003.

[SUG 04] SUGIYAMA K., HATANO K., YOCHIKAWA. M., « Adaptive Web Search Based on User Profile Constructed without Any Effort from Users », *WWW'2004*, p. 675-684, 2004.

[TEE 05] TEEVAN J., DUMAIS S., HORVITZ E., « Personalizing search via automated analysis of interests and activities », *Proceedings of the* 28^{th} *Annual ACM Conference on Research and Development in Information Retrieval SIGIR*, p. 449-456, août 15-19 2005.

[VAK 00] VAKKARI P., « Relevance and contributing information types of searched documents », *Proceedings of the* 23^{rd} *annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, p. 2-9, août 2000.

[VAK 01] VAKKARI P., « A theory of the task-based information retrieval process : a summary and generalisation of a longitudinal study », *Journal of documentation*, vol. 57, n°1, p. 44-60, 2001.

[VOO 94] VOORHEES E., « Query expansion using lexical-semantic relations », *Proceedings of the* 17^{th} *annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Croft, W.B and Van Rijsberegen, C.J Eds, p. 206-214, 1994.

[WEN 04] WEN J.R., LAO N., MA. W., « Probabilistic Model for Contextual Retrieval », *Proceedings of the* 27^{th} *annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, p. 57-63, 2004.

[WHI 03] WHITE R., RUTHVEN I., JOSE J., « A study of factors affecting the utility of implicit relevance feedback », *Proceedings of the* 28^{th} *annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Marchionini G., Moffat A. Tait J., Baeza-Yates R., Ziviani N. (dir.), p. 15-19, août 2003.

[XU 97] XU J., *Solving the word mismatch problem through automatic text analysis*, Ph.D. Thesis, Department of Computer Science, University of Massachusetts, Amherst, MA, USA, May 1997, 1997.

[YU 03] YU S., CAI D., WEN D., MA W., « Improving pseudo-relevance feedback in web information retrieval using web page segmentation », *Proceedings of the World Wide Web conferennce*, 2003.

[ZHA 03] ZHAI C., COHEN A., « Beyond independent relevance : Methods and evaluation metrics for subtopical retrieval », *Proceedings of the* 27^{th} *annual international ACM SIGIR Conference on Research and development in Information retrieval*, p. 10-17, août 2003.

Index