

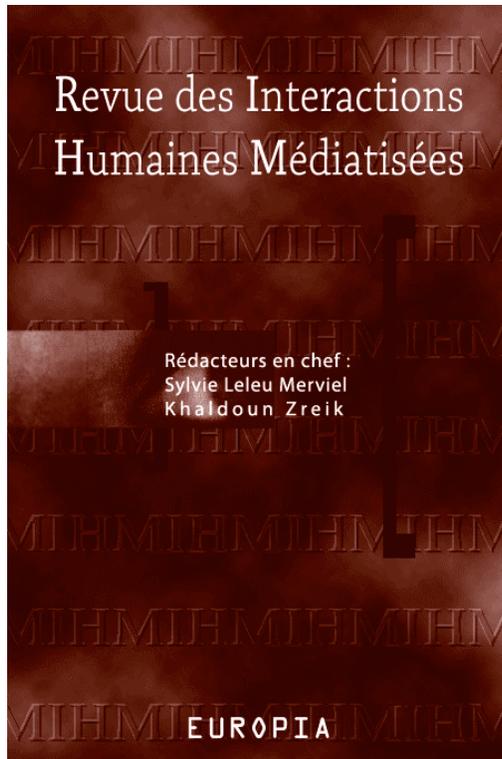
# Revue des Interactions Humaines Médiatisées

Journal of Human Mediated Interactions

## Rédacteurs en chef

Sylvie Leleu-Merviel & Khaldoun Zreik

Vol 15 - N° 2 / 2014



© europia, 2015  
15, avenue de Ségur,  
75007 Paris - France  
Tel 33 1 45 51 26 07  
<http://europia.org/RIHM>  
[rihm@europia.org](mailto:rihm@europia.org)

# Revue des Interactions Humaines Médiatisées

*Journal of Human Mediated Interactions*

## Rédacteurs en chef / *Editors in chief*

- Sylvie Leleu-Merviel, Université de Valenciennes et du Hainaut-Cambrésis, Laboratoire DeVisu
- Khaldoun Zreik, Université Paris 8, Laboratoire Paragraphe

## Comité éditorial / *Editorial Board*

- Thierry Baccino (Université Paris8, LUTIN - UMS-CNRS 2809, France)
- Karine Berthelot-Guiet (CELSA- Paris-Sorbonne GRIPIC, France)
- Pierre Boulanger (University of Alberta, Advanced Man-Machine Interface Laboratory, Canada)
- Jean-Jacques Boutaud (Université de Dijon, CIMEOS, France )
- Aline Chevalier (Université Paris Ouest Nanterre La Défense, CLLE-LTC, France)
- Yves Chevalier (Université de Bretagne Sud, CERSIC -ERELLIF, France)
- Didier Courbet (Université de la Méditerranée Aix-Marseille II, Mediasic, France)
- Viviane Couzinet (Université de Toulouse3, LERASS, France)
- Milad Doueïhi (Université de Laval - Chaire de recherche en Cultures numériques, Canada)
- Pierre Fastrez (Université Catholique de Louvain, GReMS, Belgique)
- Pascal Francq (Université Catholique de Louvain, ISU, Belgique)
- Bertrand Gervais (UQAM, Centre de Recherche sur le texte et l'imaginaire, Canada)
- Yves Jeanneret (CELSA- Paris-Sorbonne GRIPIC, France)
- Patrizia Laudati (Université de Valenciennes, DeVisu, France)
- Catherine Loneux (Université de Rennes, CERSIC -ERELLIF, France)
- Marion G. Müller (Jacobs University Bremen, PIAV, Allemagne)
- Marcel O'Gormann (University of Waterloo, Critical Média Lab, Canada)
- Serge Proulx (UQAM, LabCMO, Canada)
- Jean-Marc Robert (Ecole Polytechnique de Montréal, Canada)
- Imad Saleh (Université Paris 8, CITU-Paragraphe, France)
- André Tricot (Université de Toulouse 2, CLLE - Lab. Travail & Cognition, France)
- Jean Vanderdonckt (Université Catholique de Louvain, LSM, Belgique)
- Alain Trognon (Université Nancy2, Laboratoire InterPsy, France)

# Revue des Interactions Humaines Médiatisées

Journal of Human Mediated Interactions

Vol 15 - N° 2 / 2014

## Sommaire

### Editorial

Sylvie LELEU-MERVIEL, Khaldoun ZREIK (Rédacteurs en chef) 1

### Construction de sens et pratiques informationnelles chez les chefs militaires

*Commanders' sense making and information practices*

Anna LEZON RIVIERE, Madjid IHADJADENE 3

### Interactions communicationnelles à travers les dispositifs socio-techniques d'enregistrements audiovisuels de témoignages

*Communicative interactions with technological and relationship devices for testimony records*

Alain LAMBOUX-DURAND 31

### Implicite et incomplétude dans des requêtes adressées à un moteur de recherche Web

*Implicit and incompleteness in query sent to Web search engine*

Talal ZOUHRI 65

### Créativité architecturale. Inscription de traces informationnelles au service du sense-making urbain

*Design tools and creative thinking*

Hafida BOULEKBACHE-MAZOUZ 93

## Editorial

R.I.H.M., *Revue des Interactions Humaines Médiatisées*, est qualifiante en sciences de l'information et de la communication, puisqu'elle figure sur la liste AERES-CNU de 71<sup>ème</sup> section. Elle est également indexée par la base internationale EBSCO Publishing. Assumant sa vocation interdisciplinaire, elle croise volontiers les regards disciplinaires sur des objets partagés. Toutefois, les quatre articles de ce numéro émanent d'auteurs de la discipline de référence.

En effet, le premier article porte sur les pratiques informationnelles et la construction de sens chez les chefs militaires. S'appuyant sur les théories de la construction de sens, notamment celles de Weick et de Dervin, une étude empirique fondée sur vingt-sept entretiens individuels avec des officiers généraux et officiers supérieurs montre l'importance du facteur humain, de la confiance, de la structure et des réseaux dans les pratiques informationnelles des chefs militaires.

Le deuxième article étudie les interactions communicationnelles à travers les dispositifs sociotechniques qui permettent l'enregistrement audiovisuel de témoignages. Il montre que le dispositif mis en place pour le recueil influence sur le témoignage, son potentiel à l'exploitation et sa réception. L'analyse est menée à partir de cas expérimentaux relatifs à la sauvegarde audiovisuelle de témoignages autour d'objets patrimoniaux liés aux deux guerres mondiales, ainsi que leurs usages en contexte de médiation muséale.

Le troisième article examine l'activité de recherche d'information sous l'angle du lien entre les termes formulés dans la requête et l'intention de recherche de l'énonciateur. 61 étudiants ont été observés en situation de recherche d'information via un moteur de recherche Web et interrogés sur leur intention de recherche. Il apparaît que les requêtes recèlent une part importante d'implicite et qu'elles sont généralement incomplètes, ce qui ouvre de nouvelles voies d'amélioration via le lien entre intention et formulation de la recherche.

Enfin, le dernier article questionne la créativité architecturale en termes d'inscription de traces informationnelles au service du sense-making urbain. Fondée sur les outils de pensée, les systèmes de signes, les traces et les tracés, cette réflexion tout à fait originale sur l'architecture et l'urbain montre que de tels travaux peuvent s'ancrer résolument en sciences de l'information et de la communication à travers l'angle d'analyse et le type de regard porté sur ces objets.

Nous vous souhaitons à toutes et à tous une très bonne lecture et nous vous remercions de votre fidélité.

Sylvie LELEU-MERVIEL et Khaldoun ZREIK  
Rédacteurs en chef

# Implicite et incomplétude dans des requêtes adressées à un moteur de recherche Web

*Implicit and incompleteness in query sent to Web search engine*

Talal ZOUHRI

Laboratoire de recherche lyonnais des sciences de l'information et de la communication ELICO-EA 4147. Université Claude Bernard Lyon 1  
talal\_zouhri@yahoo.fr

**Résumé.** Les études sur l'activité de recherche d'information apportent une meilleure compréhension du comportement de recherche d'information de l'utilisateur. Pour autant, peu d'études ont examiné le lien entre les termes formulés dans la requête et l'intention de recherche de son énonciateur. Nous avons observé 61 étudiants en situation de recherche d'information via un moteur de recherche Web et nous les avons interrogés sur leur intention de recherche. Nous avons constaté que les étudiants n'expriment que partiellement leur intention de recherche dans la requête et laissent certaines informations implicites.

**Mots-clés.** Besoin d'information, implicite, incomplétude, but, intention de recherche, requête, moteur de recherche Web, recherche d'information.

**Abstract.** The studies in field of information retrieval gives a better understanding of behavior of information seeking by user. However, few studies examined the link between the terms formulated in the query and the search intent of the user. We notably analyzed 61 students' speech about their search intent. We noticed that the students express only partially their search intent in their queries and leave certain implicit information.

**Keywords.** Information need, implicit, incompleteness, goal, search intent, query, Web search engine, information seeking.

## 1 Introduction

Le présent article porte sur l'activité de recherche d'information sur le Web via un moteur de recherche. La recherche d'information sur le Web a fait l'objet de travaux nombreux et la communication entre un usager et un système d'information retient toujours l'intérêt de beaucoup de chercheurs. Ces travaux soulignent les progrès que les systèmes d'information ont atteint ces dernières décennies, mais aussi les limites (Boubée & Tricot, 2010 ; Lewandowski *et al.*, 2012) qu'ils doivent dépasser afin d'améliorer la communication avec l'utilisateur.

Il reste nécessaire de mieux comprendre ce qui, dans cette situation de communication homme/machine, limite l'interaction entre les deux parties. Une des difficultés rencontrées dans ce type de communication est que le système

d'information ne dispose que d'une représentation incomplète de la demande de l'utilisateur, de ses besoins, de ses objectifs, du contexte... autant de facteurs dont on a pu montrer qu'ils ont une incidence sur la qualité des résultats obtenus (Pitkow *et al.*, 2002 ; Allan *et al.*, 2003 ; Deveaud & Bellot, 2012). Mais si la requête ne représente qu'une explicitation partielle du besoin d'information, ceci implique que, pour des raisons variables, certaines informations ne sont pas formulées.

Pourrait-on, au moins dans une certaine mesure, les identifier ou les déduire à partir des énoncés explicites des requêtes, ou des types de requêtes, ou des stratégies employées ?

La recherche proposée ici porte sur la préparation des requêtes et leur formulation/reformulation. Nous chercherons à cerner les informations qui restent non explicitées dans la requête et à comprendre la manière dont l'absence de ces informations influe sur la qualité des résultats et sur la stratégie de recherche mise en œuvre.

D'abord, nous montrerons qu'en recherche d'information, les questions de l'implicite et de l'incomplétude soulèvent des difficultés d'approches tant sur le plan de la définition que sur celui des méthodes employées. Puis, nous présenterons une étude menée auprès de 61 étudiants, la méthode et les résultats. Enfin, nous discuterons des résultats obtenus et proposerons quelques éléments pour améliorer la prise en charge des requêtes par les moteurs de recherche Web.

## 2 Implicite, incomplétude et recherche d'information

### 2.1 Implicite et incomplétude

« Implicite » et « incomplétude » sont des termes employés régulièrement pour évoquer le fait que tout n'est pas exprimé lors d'un échange entre des personnes ou des personnes et une machine et que l'énoncé porte sur des significations plus larges que ce qui est explicité.

Ainsi, selon *Le Petit Robert* (2004), le contenu *implicite* est à déduire ou à induire : « implicite : [...] → impliquer ; plier (encadré) – Qui est virtuellement contenu dans une proposition, un fait, sans être formellement exprimé, et peut être tiré par déduction, induction. [...] » (*Le Petit Robert*, 2004).

Dans le domaine de la linguistique pragmatique, Kerbrat-Orecchioni (1986) considère que l'énoncé formulé par un émetteur est interprété par le destinataire suivant deux niveaux de contenus : un contenu explicite et un contenu implicite.

S'agissant du contenu explicite, Raeber (2011) estime qu'il « est constitué de l'ensemble des conditions qui *doivent nécessairement* être remplies dans le monde pour que l'énoncé soit vrai. On considère également que le contenu explicite, par sa nature encodée, est l'ensemble des contenus de sens sur lesquels le locuteur s'engage nécessairement ».

Quant au contenu implicite, il correspond à une inférence effectuée par l'interlocuteur et qui l'amène à re-construire un contenu différent de celui de l'énoncé. Ce contenu implicite peut être un présupposé ou un sous-entendu (Kerbrat-Orecchioni, 1986 ; Ducrot, 1984).

Ainsi, contenus explicites et implicites sont liés et complémentaires. Ensemble, ils forment le message tel que compris par le destinataire. La présence de l'implicite dans la communication interpersonnelle est utile car elle permet à l'émetteur de ne pas tout dire, tout expliquer : « trop peu, la communication n'est pas viable, trop elle n'est pas féconde » (Kerbrat-Orecchioni, 1986).

La notion d'incomplétude représenterait « un état de ce qui n'est pas complet, de ce qui n'est pas achevé [...] » (TLFi, 1994). Utilisée plus souvent par les

recherches en mathématiques et en psychologie, elle met l'accent sur le manque dans un message d'une ou plusieurs informations qui aurai(en)t permis de préciser le message. L'incomplétude ne s'oppose pas à l'implicite. Dans les deux cas, il y a un manque d'information.

Lorsque la recherche d'information se déroule dans le cadre de la communication interpersonnelle, usager et bibliothécaire se trouvent dans une situation de négociation. Au cours de cette négociation, le bibliothécaire<sup>1</sup> amène l'utilisateur à expliciter les informations manquantes et à préciser son besoin d'information avant de répondre à sa demande d'information (Le Coadic, 1998).

En revanche, une recherche d'information menée via un système d'information (un moteur de recherche web compris) ne pourra pas bénéficier de cette possibilité d'ajustement : « [...] le système d'information n'étant généralement pas capable de modifier ses connaissances à propos de celle ou celui qui cherche et de la situation (*i.e.* le motif de la recherche d'information) » (Boubée & Tricot, 2010), le système conduit la recherche sur les seuls termes explicites de la requête.

## **2.2 L'expression du besoin d'information, implicite et incomplétude**

Bien que le concept de « besoin d'information » soit discuté par de nombreux chercheurs, il demeure une notion centrale dans le domaine de la recherche d'information et a fait l'objet de nombreux travaux (Boubée & Tricot, 2010). Selon Le Coadic (1998 : 14) : « Dans le cas d'un chercheur, le besoin d'information traduit l'état de connaissances dans lequel il se trouve lorsqu'il est confronté à l'exigence d'une information qui lui manque, d'une information qui lui est nécessaire pour poursuivre un travail de recherche. Il naît donc d'une impulsion d'ordre cognitif ».

Les demandes d'information restent incomplètes pour différentes raisons. Tout d'abord, Belkin *et al.* (1982) avait déjà montré que l'utilisateur éprouve généralement des difficultés pour exprimer son besoin d'information et, Taylor (1968) avait observé que même dans les cas où l'utilisateur élabore une représentation précise de ce besoin, cela ne le conduit pas nécessairement à en faire une description fidèle à son interlocuteur (bibliothécaire) ou à la machine (système documentaire).

Ensuite, Ingwersen (1992 : 178) a également montré dans une étude empirique l'importance du phénomène « d'étiquetage » (*label effect*) : il s'agit d'un phénomène dans lequel les formulations de demandes peuvent se composer d'un ou de plusieurs concepts qui sont plus génériques que le besoin d'information et/ou qui sont exprimés hors du contexte qui constitue ce besoin d'information. Qui plus est, dans le cadre de recherche d'information sur le Web via les moteurs de recherche, on constate que les requêtes sont généralement brèves et comportent en moyenne 2,35 termes (Silverstein *et al.*, 1999).

Or, le système ne peut fournir que des réponses associées aux termes des requêtes, aussi Brini, Boughanem et Dubois (2004) considèrent que « la problématique majeure de la Recherche d'Information (RI) consiste à extraire à partir d'une collection de documents, ceux qui répondent à un besoin utilisateur en se basant souvent sur des informations pauvres ».

La requête ne constitue donc pas une représentation complète du besoin d'information que Simonnot (2006) caractérise comme étant l'inconnue des systèmes d'information.

---

<sup>1</sup> Voir le diagnostic du besoin d'information par Taylor (1968) et par Le Coadic (1998).

### 2.3 Améliorer la requête : différentes solutions proposées

Le manque d'informations dans la requête peut être désigné par le terme « implicite » (Deveaud & Bellot, 2012) ou par le terme « incomplétude » (Bouchard & Nie, 2006 ; Brini, Boughanem & Dubois, 2004).

Partant de l'idée que des connaissances additionnelles sont parfois nécessaires pour compléter la description manquante du besoin d'information dans la requête de l'utilisateur, différentes solutions ont été proposées pour enrichir automatiquement la requête initiale en s'appuyant sur les ressources du système d'information (Harman, 1992) ou sur des ressources externes. Ainsi, Deveaud et Bellot (2012) utilisent un modèle génératif probabiliste, l'allocation de Dirichlet latente (LDA), pour détecter les concepts implicites de la requête. A l'instar d'autres chercheurs (Li *et al.*, 2007 ; Suchanek *et al.*, 2007), Deveaud et Bellot s'appuient sur quatre ressources générales externes : Wikipédia comme source encyclopédique, le corpus LDC du New York Times et le corpus LDC du GigaWord comme sources d'articles de presse, et la catégorie B du ClueWeb09 comme source de pages web.

Deveaud et Bellot (2012) estiment que la méthode employée obtient les meilleurs résultats lorsque sont combinés les concepts implicites identifiés à partir de ces quatre ressources. Ainsi, l'hétérogénéité des ressources sur lesquelles s'appuie le système d'information pourrait représenter un élément important dans la détermination de ce que les chercheurs nomment « concepts implicites de qualité ».

On peut considérer qu'un tel enrichissement de la requête initiale peut apporter à l'utilisateur un nouveau vocabulaire en lien avec sa requête, et représenter une aide à la recherche documentaire. Toutefois, il reste difficile de vérifier si les concepts implicites estimés automatiquement permettent d'explicitier, au moins partiellement, la part d'implicite que recèle la requête de l'utilisateur. Pour Deveaud et Bellot (Deveaud et Bellot, 2012), ces concepts implicites sont liés au besoin d'information de l'utilisateur et leur explicitation vise à améliorer la représentation de la requête.

Bouchard et Nie (2006) proposent de compléter la requête en s'appuyant sur le domaine d'intérêt de l'utilisateur. Le modèle du domaine ainsi construit est exploité pour compléter le modèle initial de la requête : le domaine est caractérisé par un ensemble de documents et il est complété par un modèle de langue. Ce modèle est utilisé pour exploiter les dépendances lexicales du domaine afin d'étendre la requête et pour réordonner les documents préalablement retrouvés. Dans les expériences qu'ils ont menées sur des collections TREC, l'utilisateur formule une requête puis identifie un domaine pour son besoin d'information, en choisissant une des catégories d'une hiérarchie préétablie, par exemple « Finance », « Histoire », « Informatique », etc. Ces expériences conduites sur les collections TREC ont montré d'une part des améliorations de la performance de recherche et d'autre part, que la prise en compte du domaine d'intérêt de l'utilisateur peut, dans une certaine mesure, participer à diminuer l'ambiguïté de la requête en la complétant et en apportant une meilleure représentation du besoin d'information de l'utilisateur.

Une autre approche visant à améliorer la requête est celle qui consiste à repérer l'objectif de l'utilisateur à partir des requêtes exprimées.

En effet, dès les années 2000, Broder (2002) considère que la requête n'exprime que partiellement le besoin d'information et que l'intention de l'utilisateur est absente de la formulation de la requête. Aussi, Broder propose donc de catégoriser les requêtes en trois groupes d'intentions selon la forme de la requête :

- une requête navigationnelle correspondrait à l'intention de l'utilisateur d'accéder à un site particulier. Par exemple, si la requête est « /Greyhound Bus/, une entreprise de transport », alors Broder considère que la cible probable serait une adresse telle que <http://www.greyhound.com> ;
- une requête informationnelle correspondrait à l'intention de l'utilisateur de trouver des informations sur un ou plusieurs site(s) Web. Dans ce genre de requêtes, Broder estime que l'utilisateur préfère disposer de plusieurs liens pertinents sur le sujet plutôt que d'un lien unique. Dans ce cas, les requêtes informationnelles auraient une forme telle que « /cars/, /San Francisco/, /normocytic anemia/ ».
- une requête transactionnelle : l'intention de l'internaute serait de chercher à localiser un site avec lequel il interagirait. Broder considère que l'internaute formule ce type de requête lorsqu'il souhaite faire des achats, télécharger différents types de fichiers (images, chansons, etc.), accéder à des bases de données (par exemple : Les Pages Jaunes) ou à des serveurs pour jouer en ligne, etc.

Pour élaborer cette catégorisation, Broder a mené une étude empirique dans laquelle il a utilisé deux méthodes complémentaires : d'une part, un questionnaire qui s'affichait directement sur la page Web de l'internaute sous la forme d'une fenêtre pop-up et, d'autre part, une catégorisation des requêtes à partir des *logs*<sup>2</sup> du moteur de recherche AltaVista.

On peut regretter que Broder n'ait pas explicité les critères qu'il a retenus pour la catégorisation des requêtes à partir des *logs*. En outre, la catégorie des requêtes informationnelles semble être considérée comme une catégorie par défaut dans la mesure où le chercheur comptabilise, dans cette catégorie, les requêtes qui ne sont ni transactionnelles, ni navigationnelles<sup>3</sup> (Broder, 2002 : 7).

## **2.4 Identifier l'implicite dans les requêtes : les limites méthodologiques**

L'une des limites méthodologiques les plus fréquentes que l'on rencontre en recherche d'information est que les analyses portent sur les requêtes isolées ou sur des *logs*.

### ***L'analyse des requêtes isolées***

Lorsqu'il s'agit d'inférer l'intention de l'utilisateur à partir de la requête, Lewandowski *et al.* (2012) ont montré les limites méthodologiques de cette approche. Au cours de deux études, ils ont fait réaliser deux catégorisations manuelles de requêtes selon l'intention de l'auteur :

- Dans la première étude, deux juges experts dans le *Crowdsourcing* ont procédé à la catégorisation d'un corpus de requêtes isolées. Comme dans Broder (2002), ils ont utilisé les catégories de requêtes

---

<sup>2</sup> Selon Ihadjadene et Chaudiron (2008), « Les *logs* sont des fichiers de type journaux où sont enregistrées les opérations effectuées sur un serveur par des utilisateurs ou des processus. Outre la date et l'heure de connexion, ils donnent l'adresse IP des internautes et le texte de leurs commandes ».

<sup>3</sup> « From the remaining set the first 400 queries were inspected. Queries that were neither transactional, nor navigational, were assumed to be informational in intent » (Broder, 2002 : 7).

informationnelles, navigationnelles et transactionnelles. Ils ont aussi rajouté deux autres catégories : la catégorie des requêtes commerciales (une requête est considérée comme commerciale si l'utilisateur, en la formulant, peut être intéressé par des offres commerciales (Lewandowski *et al.*, 2012 : 1773) et la catégorie des requêtes locales (requêtes qui seraient formulées par l'utilisateur pour rechercher une information dans la zone géographique d'où il réalise sa recherche (Lewandowski *et al.*, 2012 : 1773). Les juges ne disposaient pas d'une liste exhaustive de critères mais d'exemples illustrant chacune des catégories.

- Dans la seconde étude, lorsqu'un utilisateur du portail de recherche allemand *T-online* formulait une requête identique à celle qui se trouvait déjà dans le corpus de l'étude 1, une fenêtre pop-up s'affichait sur l'écran et invitait l'utilisateur à classer lui-même sa requête. L'utilisateur devait répondre à trois questions. La première le questionnait sur ce qu'il souhaitait trouver avec sa recherche, une liste de choix à cocher était alors proposée. La deuxième question portait sur les raisons qui ont amené l'utilisateur à réaliser cette recherche (là aussi, il disposait d'une liste de propositions à cocher). La troisième question était ouverte et invitait l'utilisateur à exprimer avec ses propres mots quels résultats il espérait trouver avec sa recherche.

Ces deux études aboutissent à des résultats différents :

- Lorsque les experts catégorisent les requêtes, ce sont les requêtes navigationnelles et commerciales qui sont les plus nombreuses.
- Lorsque ce sont les utilisateurs du portail de recherche qui catégorisent leurs requêtes, alors les requêtes transactionnelles sont les plus nombreuses.

Lewandowski *et al.* (2012) en concluent que la dérivation d'intention de recherche à partir de requêtes isolées ne peut pas être considérée comme fiable.

### ***Evolution du besoin d'information et pertinence***

Outre la brièveté des requêtes, « le besoin d'information » de l'utilisateur est d'autant plus difficile à cerner qu'il est évolutif, et qu'il est indissociable des jugements de pertinence émis par l'utilisateur (Borlund, 2003). Ainsi, selon Schamber *et al.* (1990), la pertinence est une notion : a- « dynamique » dans la mesure où les critères de jugements de pertinence de l'utilisateur sont amenés à évoluer au cours de la recherche, b- « situationnelle » car liée au problème informationnel à résoudre et c- mesurable car il est possible de dénombrer les critères de pertinence appliqués par les utilisateurs.

La dimension temporelle représente donc un élément central dans l'analyse de l'activité de recherche d'information car pendant le déroulement de cette activité, le besoin d'information évolue au fil de la recherche (Bates, 1989) : dans toute démarche de recherche d'information, un document jugé pertinent peut, non seulement, faire évoluer le besoin d'information, mais aussi conduire, dans le cadre de la reformulation, à supprimer un (ou plusieurs) terme(s) ou bien à en introduire de nouveaux. De même, des résultats jugés peu pertinents peuvent amener l'utilisateur à davantage expliciter son besoin d'information. L'analyse des requêtes isolées ne peut pas tenir compte de la dimension temporelle inhérente à la recherche d'information, de l'évolution du besoin de l'utilisateur, ni de ses objectifs.

Si l'analyse limitée à la requête ne convient pas, il reste à définir la session de recherche. En recherche d'information, de très nombreux travaux s'appuient sur les

*logs*. Or, des chercheurs ayant analysé des *logs* de moteurs de recherche Web ont montré qu'il n'est pas possible de déterminer avec exactitude, uniquement à partir des données disponibles dans les *logs*, la session de recherche réelle de l'utilisateur (Jansen et Spink, 2006 : 254). À défaut d'une session de recherche réelle de l'utilisateur, les chercheurs ayant réalisé ces travaux proposent, en alternative, une session de recherche dont la durée est fixée artificiellement. Par exemple, Chau, Fang et Yang (2007 : 1048) considèrent que chaque adresse IP (Internet Protocol) représente un seul utilisateur et fixent la durée des sessions à 30 minutes.

Compte tenu d'une part, de la nécessité de tenir compte du déroulement complet d'une recherche sur une ou plusieurs requêtes, et d'autre part de cerner les intentions de l'utilisateur, nous proposons ici la démarche suivante :

- en partant de l'idée que l'utilisateur sait quand il démarre une session de recherche d'information et quand il a terminé celle-ci, nous choisissons l'observation empirique directe de l'activité de recherche d'information ; la durée d'une session est donc définie par l'utilisateur,
- sachant que la seule forme de la requête ne permet pas d'inférer l'intention de l'utilisateur, nous proposons de compléter l'observation directe par un entretien avec l'utilisateur,
- enfin, nous considérons ici que le terme « d'implicite » désigne toute information absente de la requête alors qu'elle est présente dans l'objectif de recherche de l'utilisateur.

### **3 Etude empirique : analyse des requêtes formulées par des étudiants en sciences**

#### **3.1 Objectif de recherche et hypothèses**

L'observation expérimentale que nous proposons a pour objectif de réaliser une étude qualitative de la phase de recherche d'information sur le web par des étudiants en sciences. D'une part, elle vise à recueillir les requêtes formulées lors de sessions de recherche et, d'autre part, elle vise à recueillir les justifications orales fournies par les étudiants ayant formulé ces requêtes.

Nous partons des hypothèses suivantes :

- h1 : toute requête est incomplète et certaines informations absentes sont implicites.
- h2 : cet implicite peut être utilisé dans les tactiques de recherche via des formulations/reformulations.
- h3 : c'est la session de recherche (*i.e.* l'enchaînement des requêtes lors d'une recherche d'information donnée) qui permet d'approcher l'intention de l'auteur.

#### **3.2 Méthodologie**

Au cours du mois de mai 2012, un mois de révision pour les étudiants, nous avons mené une observation dans des salles informatiques en libre accès de l'université Claude Bernard Lyon 1.

##### ***Les sujets***

Ont participé à l'étude 61 étudiants (21 femmes et 40 hommes) inscrits en cursus scientifique à l'université Claude Bernard Lyon 1. Les sujets étaient volontaires et non rémunérés. Ils étaient âgés de 18 à 32 ans avec une moyenne d'âge de 22 ans. Excepté le sujet de 32 ans qui venait de finir une thèse en chimie, ils étaient tous inscrits en licence ou en master.

Pour ces sujets, l'ordinateur est un instrument qu'ils connaissent et qu'ils utilisent de manière autonome quotidiennement. Tous ont bénéficié, dès leur première année universitaire, d'une formation à la méthodologie de la recherche documentaire, dispensée par le service commun de la documentation de leur université.

### ***Matériels utilisés***

Les sujets étaient installés dans une salle informatique en libre accès où ils se rendent régulièrement pour réaliser leurs propres recherches. Cette salle était notre lieu d'observation. Les entretiens ont tous été enregistrés sur un dictaphone. En outre, nous avons utilisé un carnet sur lequel nous avons noté, pour chaque sujet, le numéro de l'enregistrement sur le dictaphone, la ou les requête(s) formulée(s) et des informations correspondant à son sexe, son âge, sa discipline et son niveau d'étude.

### ***Déroulement de l'observation***

Cette observation a été réalisée en deux temps consécutifs : a) observation et recueil des requêtes b) directement après la phase (a) de l'observation, nous avons mené des entretiens avec chacun des étudiants à propos des requêtes formulées.

Tout d'abord, nous observions les écrans des étudiants. Lorsqu'un étudiant utilisait un moteur de recherche Web, nous lui demandions, après nous être présenté, s'il souhaitait participer à la recherche expérimentale que nous réalisions dans le cadre de notre parcours universitaire. Lorsqu'il nous donnait son accord, nous lui demandions l'autorisation de relever la requête qu'il venait de formuler.

Ensuite, dans le cadre d'un entretien, nous le questionnions sur cette requête en l'interrogeant sur son intention de recherche : « Avec cette requête que voulez-vous chercher ? »

Après cela, nous lui demandions s'il avait saisi une autre requête avant celle-ci. S'il nous répondait positivement, alors nous lui demandions de nous indiquer cette requête et si possible d'utiliser l'historique du navigateur Web afin de l'afficher. Dans un premier temps, nous relevions cette requête et, dans un second temps, nous l'interrogeons sur son intention de recherche.

### ***Corpus recueillis***

Différentes données qualitatives ont été recueillies :

- Dans la mesure où nous étions aux côtés de chaque sujet lors du déroulement de son activité de recherche et donc lors de la formulation/reformulation de chacune des requêtes, nous avons recueilli la ou les requête(s) formulée(s), de même que nous avons pu noter l'éventuelle démarche de navigation, le ou les lien(s) cliqué(s).
- En plus de ces informations, nous avons explicitement interrogé chaque sujet sur son intention de recherche et chaque sujet, lors de cet entretien, a donc justifié de la formulation/reformulation de sa (ses) requête(s).

À l'issue de chaque session, nous disposions donc pour chaque sujet :

- De la session complète de recherche, c'est-à-dire de la requête formulée et, lorsqu'il y en avait plusieurs, de la suite ordonnée de ces requêtes et des actions des sujets.
- Des verbalisations associées à ces requêtes.

### ***Les requêtes formulées***

Nous avons relevé au total 101 requêtes lors de cette observation formulées dans le cadre de 94 sessions de recherche.

### *Les verbalisations associées aux requêtes*

61 entretiens correspondant aux verbalisations recueillies *a posteriori* ont été enregistrés avec un dictaphone. En moyenne, la durée des entretiens réalisés est de deux minutes et l'entretien le plus long a duré sept minutes. Ces entretiens ont été retranscrits dans leur intégralité.

### **Traitement des données recueillies**

#### *Analyse des requêtes, des informations recherchées et des justifications des étudiants*

Nous cherchons à catégoriser les requêtes selon l'intention de l'utilisateur et en utilisant le modèle de Broder (2002) mais contrairement à Broder, notre analyse n'est pas fondée sur la requête isolée de son contexte, mais sur l'association de chaque requête aux justifications fournies par le sujet.

Ainsi, le corpus de 101 requêtes est analysé en mobilisant les notes issues de notre observation de l'activité de recherche des étudiants ainsi que de leur justification.

#### *Analyse des sessions de recherche*

L'expression « session de recherche » désigne ici l'ensemble des requêtes formulées en vue de réaliser une tâche de recherche d'information donnée. L'expression « tâche de recherche » est plus large : elle englobe à la fois la session et les verbalisations afférentes aux différentes requêtes composant la session. Chaque sujet est susceptible de mener une seule tâche de recherche ou plusieurs.

Les sessions de recherche sont caractérisées selon leur composition :

- la première catégorie contient les sessions de recherche composées d'une seule requête,
- la seconde catégorie contient les sessions de recherche composées de plus d'une requête (nous avons tenu compte dans ce cas de la dimension temporelle, les différentes reformulations représentant une évolution de la formulation de la requête initiale).

#### *Recherche de l'implicite*

Notre analyse porte sur les sessions de recherche. Pour chacune de ces sessions :

- nous comparons les termes de chaque requête avec les termes des verbalisations la justifiant.
- nous examinons l'évolution des termes dans les requêtes formulées/reformulées dans une même session tout en nous appuyant sur les verbalisations afférentes à ces requêtes.

## **4 Résultats**

### **4.1 Catégorisation des requêtes selon l'intention**

En associant chaque requête aux justifications données par le sujet, nous avons classé les 101 requêtes selon les trois groupes de Broder (2002). Pour chaque catégorie, nous donnerons ci-dessous quelques exemples :

- Une requête navigationnelle : la requête est formulée en vue d'accéder au lien d'un site que l'étudiant connaît déjà. L'étudiant utilise dans ce cas le moteur de recherche comme un annuaire. Le tableau 1 ci-après représente quelques exemples de requêtes navigationnelles accompagnées des justifications de leur énonciateur.

	<b>Requête</b>	<b>Justifications fournies par le sujet</b>
<b>Sujet 21</b>	insa lyon (requête 55)	« je cherchais le site de l'INSA »
<b>Sujet 46</b>	youtube (requête 79)	« accéder au site de <i>Youtube</i> »
<b>Sujet 56</b>	spiral connect (requête 95)	« Je veux entrer dans mon site, le site de la fac »

**Tableau 1.** Exemples de requêtes navigationnelles

- Une requête informationnelle : une requête est formulée afin de trouver une ou plusieurs information(s). Dans ce cas, le sujet cherche à obtenir une information nouvelle ou à en vérifier une dont il n'est plus sûr. Le tableau 2 ci-après représente des exemples de requêtes informationnelles accompagnées des justifications de leur énonciateur.

	<b>Termes des requêtes</b>	<b>Justifications fournies par le sujet</b>
<b>Sujet 17</b>	matrice inverse 2x2 (requête 28)	« la réponse à une question en math »
<b>Sujet 28</b>	alanine (requête 43)	« trouver plus d'informations sur une molécule, une protéine, un acide aminé »
<b>Sujet 32</b>	tommy tallarico (requête 52)	« tous les produits auxquels ce compositeur de musique a participé »

**Tableau 2.** Exemples de requêtes informationnelles

- Une requête transactionnelle : une requête formulée par le sujet afin de localiser un site avec lequel il pourra interagir. Le tableau 3 représente quelques exemples de requêtes transactionnelles accompagnées des justifications de leur énonciateur.

	<b>Termes des requêtes</b>	<b>Justifications fournies par le sujet</b>
<b>Sujet 94</b>	windows 7 professionnelle (requête 55)	« à télécharger Windows 7 » « je dois aller sur le site de Windows en fait pour le télécharger et le graver sur un CD »
<b>Sujet 34</b>	test domino (requête 57)	« c'est pour faire des exemples de tests domino parce que j'ai un entretien psychotechnique maintenant en fait là je recherche des tests domino pour s'habituer à faire des tests de domino »

**Tableau 3.** *Exemples de requêtes transactionnelles*

Les requêtes ne sont pas réparties uniformément dans les trois catégories. Ainsi les requêtes navigationnelles sont les plus nombreuses (50,49%), suivies des requêtes informationnelles (43,56%) et des requêtes transactionnelles (5,94%).

Le nombre de termes par requête est également variable. Nous désignons par « terme » un ou plusieurs caractère(s) séparé(s) d'un espace. Par exemple, nous considérons que la requête 14 /spiral univ lyon 1/ se compose de 4 termes.

Sur l'ensemble du corpus de requêtes, le nombre de termes moyen par requête pour l'ensemble du corpus de requêtes est de 2,11. Mais ce nombre moyen varie selon la catégorie des requêtes : les requêtes transactionnelles sont celles qui comportent le nombre moyen de termes par requête le plus élevé (3,16 termes par requête), suivi des requêtes informationnelles (2,52 termes par requête). Enfin, c'est dans les requêtes navigationnelles que le nombre de termes moyen par requête est le plus bas (1,64 terme par requête).

## **4.2 Des particularités observées dans le corpus des requêtes**

### *Sur le nombre de termes par requête*

La moyenne de termes par requête dans notre corpus (2,11 termes par requête) est similaire à celle observée dans les résultats d'études sur les moteurs de recherche Web (2 termes par requêtes) (Jansen & Pooch, 2001). Nous constatons que les requêtes formulées pour accéder à un site (requêtes navigationnelles) sont plus courtes que celles formulées pour rechercher de l'information (requêtes informationnelles) ou pour interagir avec un site Web (requêtes transactionnelles). En outre, les requêtes informationnelles apparaissent comme un cas à part, puisque c'est dans ce groupe qu'on trouve les requêtes avec le plus grand nombre de termes (de 1 à 8 terme(s)).

### *Sur les catégories de Broder*

Nous avons pu constater que certaines requêtes échappent aux trois catégories de Broder (2002) :

#### *a) Détournement de la fonction Expansion des requêtes du moteur de recherche*

Lors de la classification des requêtes par catégorie, nous avons considéré la requête /persévère/ comme une requête informationnelle. Le sujet 39 justifiait sa requête en expliquant qu'il souhaitait connaître la bonne orthographe du terme

« persévérant » pour un travail d'écriture en cours (un compte-rendu). Or, ce sujet a formulé cette requête mais ne l'a pas envoyée au moteur de recherche Web. Il s'est contenté d'examiner les expansions de requêtes fournies par le moteur de recherche pour en retenir celle comportant l'orthographe qui lui semblait pertinente. Ensuite, il a repris la rédaction de son compte-rendu sur un logiciel de traitement de texte. Le sujet 39 n'a donc pas mené de recherche d'aucune sorte, il s'est contenté de détourner la fonction Expansion des requêtes du moteur de recherche.

*b) Utilisation des résultats du moteur de recherche pour vérifier une expression ou l'orthographe d'un terme*

Il s'agit ici du cas du sujet 16. En effet, celui-ci formule la requête 25 /capacité mnésique/ afin de vérifier si cette expression existe, comme le montre le dialogue sujet/enquêteur ci-après :

- Sujet 16 : « *En fait, là je rédige un compte-rendu et avec /capacité mnésique/ j'ai voulu vérifier si cette expression existe ou pas. Vous comprenez, avant de mettre une expression, il faut d'abord vérifier si elle existe. Moi, je pense qu'elle existe mais je ne suis pas sûre. J'ai une astuce, je regarde dans les résultats de Google si elle existe. Donc vous pouvez voir que sur le premier lien de Wikipédia on peut lire : « amnésiques » ou au contraire « doués d'une capacité mnésique exceptionnelle ». Donc, je peux utiliser cette expression dans mon compte-rendu.* »

- Enquêteur : « *vous procédez souvent comme ça ?* »

- Sujet 16 : « *oui, à chaque fois que j'ai un doute.* »

- Enquêteur : « *et vous regardez que le lien de Wikipédia ?* »

- Sujet 16 : « *non, des fois le lien de Wikipédia n'est pas affiché donc je regarde au niveau des autres liens. Pour l'orthographe, je tape le mot qui me pose problème et je rajoute le nom d'un grand journal. Par exemple hier j'ai tapé /concurrent l'Express/ puis je regarde dans les articles de l'Express comment s'écrit le mot. Par exemple, j'ai vite vu qu'il faut deux "r" à "concurrent" ».*

Contrairement, au sujet précédent (numéro 39), celui-ci envoie la requête saisie, puis utilise les résultats du moteur pour procéder à cette vérification.

*c) Utilisation du moteur de recherche pour vérifier la conjugaison d'un verbe*

Le sujet 16 formule la requête 23 /distraire/ afin d'accéder à la fiche de conjugaison de ce verbe sur le site *laconjugaison.com*. Le sujet explique procéder ainsi à chaque fois qu'il souhaite vérifier la conjugaison d'un verbe « Sujet 16 : [...] *on peut vérifier la conjugaison des verbes et vu que j'avais besoin de celui-ci, je voulais voir comment il se conjugue [...]* ». Seul le sujet 16 a formulé ce type de requête.

On remarquera ici que ces trois requêtes (sujets 23, 25 et 39) ont toutes été formulées pendant l'accomplissement d'une tâche d'écriture. Ainsi, lors de la réalisation de travaux écrits, le moteur de recherche Web semble servir de support à la rédaction.

### 4.3 Analyse des sessions de recherche

Les 94 sessions de recherche se répartissent de la manière suivante : 87 sessions de recherche composées d'une seule requête et 7 sessions de recherche composées de plus d'une requête (tableau 6).

	<b>Session de recherche composée d'une seule requête</b>	<b>Session de recherche composée de plus d'une requête</b>
<b>Requêtes informationnelles</b>	33 (35,10%)	5 (5,31%)
<b>Requêtes navigationnelles</b>	48 (51,06%)	1 (1,06%)
<b>Requêtes transactionnelles</b>	6 (6,38%)	0
<b>(une session de recherche composée de requêtes informationnelles et navigationnelles)</b>		2 (2,12%)
<b>Total</b>	87 (91,48%)	7 (8,49%)

**Tableau 6.** *Catégorisation des requêtes selon la complexité de la session*

91,48% des sessions de recherche réalisées comportaient une seule requête et 8,49% en comportaient au moins deux.

Sur l'ensemble des sessions de plus d'une requête, une seule session est composée de 3 requêtes. Les autres sessions sont composées d'une requête initiale suivie d'une reformulation. Seul un individu (le sujet 6) est l'auteur de deux sessions.

Lorsque la session de recherche comporte deux requêtes, celles-ci sont principalement des requêtes informationnelles (tableau 7).

	Requête	justifications
<b>Sujet 6-session 1</b>	pulse B&K (requête 8)	« j'aimerais chercher des caractéristiques par rapport au logiciel pulse[...] »
	logiciel pulse B&K (requête 9)	« J'ai rajouté le mot logiciel et après je suis tombé sur un truc qui m'intéressait »
<b>Sujet 6-session 2</b>	doppelganger (requête 10)	« alors j'aimerais chercher tout ce qui va par rapport au mythe sur doppelganger [...] »
	legende doppelganger (requête 11)	« [...] les résultats ont été moins, ça parlait quand même de doppelganger mais c'était pas exactement ce que je recherchais principalement. J'ai plus ciblé principalement sur l'histoire et la légende [...] »
<b>Sujet 13</b>	windows live (requête 19)	« c'était un bug récent qui m'intéressait ».
	msn bug (requête 20)	« une solution à un problème, MSN marchait pas je voulais savoir comment faire pour qu'il remarque »
<b>Sujet 28</b>	acides aminés (requête 42)	« je voulais voir, en fait là, comment était formée la glycine »
	alanine (requête 43)	« trouver plus d'informations sur une molécule, une protéine, un acide aminé »
<b>Sujet 54</b>	time val (requête 54)	« c'est pour du C++, c'est pour la programmation, c'est pour récupérer une valeur temporelle, en fait »
	clock c fonction (requête 93)	« du coup, c'est quelqu'un d'autre qui m'a dit ce qu'il fallait chercher. J'ai pu trouver comme ça [...] il m'a dit ce qu'il fallait que je tape et je l'ai fait [...] c'était pour la même chose. Ça permet de récupérer l'heure du système en millisecondes »

**Tableau 7.** Sessions de recherche composées de requêtes informationnelles

Le tableau 7 ci-dessus montre que dans ces sessions de recherche, la reformulation d'une requête informationnelle porte sur le contenu de la requête. Le thème de recherche est identique mais les sujets ont introduit un changement dans la requête en passant du général au spécifique.

Une seule session de recherche comporte deux requêtes navigationnelles. La reformulation porte sur le changement du nom d'un site commercial /ecotour/ par un autre site proposant des services similaires /opodo/. Le sujet 41 considère ce dernier comme le concurrent du site *ecotour.com* et a réalisé ses recherches dans le cadre de la préparation d'un voyage. Le tableau 8 ci-après montre la session de recherche formée de ces deux requêtes navigationnelles.

	Session	Requête	justifications
<b>Sujet 41</b>	Requête 69	ecotour	« j'ai saisi la requête /ecotour/ pour accéder au site concurrent d'opodo »
	Requête 70	opodo	« [...] j'y suis allée sur opodo après pour comparer les tarifs proposés sur les deux portails, dans le but d'organiser un voyage [...] »

**Tableau 8.** Session de recherche comportant deux requêtes navigationnelles

Mais la session de recherche n'est pas nécessairement uniforme : des requêtes qui se suivent ne relèvent pas nécessairement de la même catégorie. Le tableau 9 ci-après montre deux exemples :

- Dans le premier cas, le sujet 17 formule d'abord une requête informationnelle pour savoir comment calculer une matrice, ensuite il formule une seconde requête navigationnelle afin d'accéder au site de son enseignant de mathématiques, puis reformule une troisième requête pour compléter sa requête initiale.
- Dans le second cas, la session comporte une requête navigationnelle suivie d'une requête informationnelle. Le sujet 19 a d'abord saisi la requête navigationnelle /recrutement galaxie/ afin de se rendre sur un site recensant des offres d'emploi de postes d'attaché temporaire dans l'enseignement et la recherche. Ensuite, il a formulé une requête informationnelle /ater 2012/ afin de chercher d'autres annonces sur le Web.

	Requête	justifications	Catégorie Broder
<b>Sujet 17</b>	matrice inverse (requête 26)	« comment calculer une matrice inverse »	informationnelle
	malbos (requête 27)	« la page personnelle du prof de math » [remarque : le professeur est l'auteur de l'exercice]	navigationnelle
	matrice inverse 2x2 (requête 28)	« la réponse à une question en math »	informationnelle
<b>Sujet 19</b>	recrutement galaxie (requête 30)	« là, je tombe sur le portail galaxie des enseignants chercheurs et après, je passe accès galaxie recrutement et après j'entre mon code et après j'ouvre ma page »	navigationnelle
	ater 2012 (requête 31)	je cherche les offres de postes d'ater qui sont disponibles sur les différents établissements	informationnelle

**Tableau 9.** Sessions de recherche contenant des requêtes de catégories différentes

Cette analyse des sessions de recherche montre que des informations figurant dans les justifications sont absentes des requêtes. Ainsi, les justifications des sujets sont nécessaires pour comprendre d'une part ce que les sujets recherchent comme information et d'autre part ce qui les amène à recourir à la reformulation.

#### 4.4 L'implicite dans les requêtes

Dans cette sous-section sera examinée la part d'implicite que recèlent les requêtes de notre corpus classées selon les trois catégories de Broder (2002). Cet implicite est évalué en comparant les termes de la requête avec les justifications fournies par les sujets.

Nous distinguons deux niveaux d'analyse :

- un niveau sémantique correspondant au thème de recherche et représenté par des syntagmes nominaux simples ou complexes,
- et un niveau pragmatique représenté par :
  - o un but : le cadre dans lequel s'inscrit la recherche. Par exemple, préparer un examen, un rapport, un entretien d'embauche,
  - o un but et un sous-but. Par exemple, préparer un compte-rendu en informatique (but) et obtenir une définition (sous-but).

#### *L'implicite dans les requêtes informationnelles*

##### *a) Implicite sémantique dans les requêtes informationnelles*

Les requêtes informationnelles peuvent être distinguées selon l'importance de l'implicite qu'elles comportent au niveau sémantique :

- Dans un premier groupe, les requêtes contiennent tous les termes explicités dans les verbalisations et décrivant le thème de recherche, ceci représente 48% des requêtes informationnelles.
- Dans un second groupe, les requêtes ne contiennent que certains termes explicités dans les verbalisations et décrivant le thème de recherche, c'est le cas pour 52% des requêtes informationnelles. Dans ce cas, sont absentes des requêtes les précisions et/ou des caractéristiques sur le thème de recherche. Des termes précis sont absents de la requête alors qu'ils sont explicités, voire répétés, dans les verbalisations.

Le tableau 10 ci-dessous présente des exemples de requêtes contenant une part plus ou moins importante d'implicite au niveau sémantique.

	<b>Requêtes</b>	<b>Extraits des verbalisations</b> ( <i>les termes décrivant le thème de la recherche et laissés implicites dans la requête sont surlignés</i> )	<b>Type de requête</b>
<b>Sujet 24</b>	equation de clapeyron en environnement (requête 88)	« c'est pour savoir la <b>résolution</b> et les <b>conditions initiales</b> quoi », « le principe dans quel cas on peut utiliser cette équation, c'est pour déterminer la <b>température</b> ou bien la <b>pression</b> », « et les différents cas d'application dans le cadre de la <b>pression</b> et de la <b>température</b> ».	Informationnelle
<b>Sujet 38</b>	pénicilline (requête 61)	« c'était pour connaître l'opinion de Wikipédia concernant les <b>découvreurs</b> de la pénicilline <b>Fleming</b> et <b>Howard</b> je sais plus qui »	Informationnelle
<b>Sujet 12</b>	ethylendiamine (requête 17)	« je cherche des informations sur l'éthylènediamine, la <b>densité</b> , la <b>masse</b> , comment les <b>configurations</b> , ben tout »	Informationnelle

**Tableau 10.** *Requêtes informationnelles avec implicite au niveau sémantique*

Ainsi on peut constater que dans le cas des requêtes informationnelles, l'implicite au niveau sémantique est important puisque la formulation des requêtes reste très incomplète ou d'un degré de précision très inférieur en comparaison des justifications orales des sujets.

*b) Implicite pragmatique dans les requêtes informationnelles*

La confrontation entre les termes des requêtes et les justifications fournies par les sujets montrent que d'une manière très générale, les requêtes informationnelles ne comportent pas d'indications sur les buts et sous-buts des sujets, ils restent à l'état d'implicite. Les deux exemples donnés dans le tableau 11 ci-après, montrent que ces requêtes informationnelles ne comportent aucune indication sur les objectifs réels et pratiques de l'utilisateur, tels que rédiger un compte-rendu, trouver le corrigé d'un exercice.

	Requête	Extraits des verbalisations	But
<b>Sujet 18</b>	effet d'un traitement sur la masse des rongeurs (requête 29)	« les résultats qui concernent un exercice ou un sujet de maths », « je fais l'exercice, ensuite je cherche le corrigé s'il existe ».	Rechercher le corrigé d'un exercice de mathématiques
	Requête	Extraits des verbalisations	But et sous-but
<b>Sujet 16</b>	capacite mnésique (requête 25)	« En fait, là je rédige un compte-rendu et avec /capacité mnésique/ j'ai voulu vérifier si cette expression existe ou pas. Vous comprenez, avant de mettre une expression, il faut d'abord vérifier si elle existe. Moi, je pense qu'elle existe, mais je ne suis pas sûre », « j'ai une astuce, je regarde dans les résultats de <i>Google</i> si elle existe. Donc vous pouvez voir que sur le premier lien de <i>Wikipédia</i> on peut lire « amnésiques » ou au contraire « doués d'une capacité mnésique exceptionnelle ». Donc, je peux utiliser cette expression dans mon compte-rendu ».	<b>But :</b> Rédiger un compte-rendu <b>Sous-but :</b> Vérifier si cette expression existe

**Tableau 11.** *Implicite pragmatique dans les requêtes informationnelles*

Une seule requête informationnelle ne correspond pas au schéma décrit ci-dessus et explicite le but du sujet. L'exemple en est donné dans le tableau 12 ci-dessous.

	Requête	Extraits des verbalisations	But
<b>Sujet 18</b>	mettre un axe secondaire excel (requête 88)	« c'est pour mettre un axe secondaire sur un graphique », « c'est pour réussir à faire un graphique correctement ».	Mettre un axe secondaire sur un graphique réalisé avec <i>Excel</i>

**Tableau 12.** *Requête informationnelle sans implicite pragmatique*

## L'implicite dans les requêtes navigationnelles

### a) Implicite sémantique dans les requêtes navigationnelles

Une analyse des requêtes navigationnelles montre que ce qui est formulé par les sujets est le nom du site et non son adresse URL. Le nom du site peut pour certaines adresses URL correspondre au nom de domaine du deuxième niveau (requête n° 1 /youtube/) ou au nom des domaines de deuxième et de premier niveau (requête n° 81 /Laposte.net/).

Certaines requêtes navigationnelles comportent une description incomplète du nom du site, un certain contenu est laissé implicite dans la requête mais explicité dans les verbalisations. Par exemple :

- la requête 66 /lyon 1/ avec laquelle le sujet 39 souhaitait chercher « l'adresse de l'université Claude Bernard »,
- la requête 4 /spiral/ avec laquelle le sujet 4 souhaitait « accéder au site de spiral connect ».

### b) Implicite pragmatique dans les requêtes navigationnelles

En outre, l'ensemble des requêtes navigationnelles comporte de l'implicite au niveau pragmatique. Dans ces requêtes, les sujets ne formulent pas leur but et sous-but.

- Dans l'exemple ci-dessous (tableau 13), le sujet dit chercher à accéder au site Youtube (sous-but) afin de regarder des vidéos de musique (but). Pourtant, on peut remarquer qu'il ne précise ni son but, ni son sous-but dans la requête.

	Requête	Extrait des verbalisations	Sous-But	But
Sujet 46	youtube (requête 79)	« accéder au site de Youtube »	Accéder au site Youtube	« regarder des vidéos de musique »

**Tableau 13.** But et sous-but implicites dans une requête navigationnelle

- Dans l'exemple suivant (tableau 14), le sujet ne mentionne pas au cours de l'entretien son sous-but qui est d'accéder au service de messagerie hotmail. Il mentionne directement son but qui est de lire son courriel. Dans sa requête, le but et le sous-but demeurent à l'état d'implicite.

	Requête	Extrait des verbalisations	But
Sujet 1	hotmail (requête 1)	« je voulais lire mes mails »	Lire des mails

**Tableau 14.** But explicité dans les verbalisations mais pas dans la requête

- Dans ce dernier exemple, le sujet cherche à accéder à un site dans le but d'utiliser un service (tableau 15). Dans les justifications, le sujet verbalise son sous-but (la localisation du site) et son but (utilisation

des services proposés par ce site). En revanche, dans la requête, le sujet n'explique pas son but et son sous-but.

	Requête	Extraits des verbalisations	But	Sous-but
<b>Sujet 39</b>	lyon 1 (requête 66)	« L'adresse de l'université Claude Bernard », « dans le but de trouver mes cours et de réviser ».	Réviser ses cours	Chercher l'adresse de l'université Claude Bernard Lyon 1

**Tableau 15.** *But et sous-but explicités dans les verbalisations mais pas dans la requête*

### *L'implicite dans les requêtes transactionnelles*

Sur les 6 requêtes transactionnelles de notre corpus, une seule requête comporte une part d'implicite au niveau sémantique (par rapport aux verbalisations afférentes à cette requête) et cinq requêtes comportent de l'implicite au niveau pragmatique. Nous présentons dans le tableau 16 ci-dessous des exemples de ces requêtes.

	Requête	Niveau pragmatique	Niveau sémantique
<b>Sujet 34</b>	test domino (requête 57)	le sujet n'a pas explicité dans sa requête son but qui est s'exercer à des tests psychotechniques	des termes en lien avec la recherche ne sont pas exprimés dans la requête et sont verbalisés dans les justifications : « tests logiques » et « tests psychotechniques »
<b>Sujet 55</b>	windows 7 professionnelle (requête 94)	le sujet n'explique pas dans sa requête son but qui est de télécharger un logiciel	le sujet n'apporte pas d'autres termes en lien avec le niveau sémantique de la recherche dans ses justifications.
<b>Sujet 4</b>	location de chapiteau (requête 5)	le sujet explicite dans sa requête son but qui est de louer un chapiteau mais n'explique pas son sous-but qui est de comparer les prix.	le sujet n'explique pas dans sa requête le terme « prix » alors qu'il le verbalise dans les justifications.

**Tableau 16.** *Exemples d'implicite dans des requêtes transactionnelles*

#### 4.5 Implicite et stratégies de recherche : analyse des sessions de recherche composées de plus d'une requête informationnelle

Dans cette sous-section, sont considérées les sessions de recherche composées de plus d'une requête informationnelle.

##### *Une tactique de reformulation de la requête : introduire de la précision*

Le sujet 6, après avoir formulé une requête initiale, l'a complétée en rajoutant un nouveau terme « logiciel ». Comme illustré dans le tableau 17, ce terme était implicite dans la requête initiale alors qu'il était exprimé par le sujet pour justifier cette même requête. Non satisfait des résultats proposés, il choisit d'introduire ce terme dans sa requête, en vue d'apporter de la précision à sa requête. Le tableau 17 ci-après présente la session de recherche réalisée par le sujet 6.

	Requête initiale n°8	Extrait des verbalisations	Reformulation de la requête 9	Extrait des verbalisations
Sujet 6	pulse B&K	« j'aimerais chercher des caractéristiques par rapport au logiciel pulse qui est fait par B&K qui est une boîte allemande. Je vais chercher des caractéristiques par rapport à ce logiciel par rapport à des fréquences pour un TP »	logiciel pulse B&K	« J'ai rajouté le mot logiciel et après je suis tombé sur un truc qui m'intéressait »

**Tableau 17.** Session de recherche réalisée par le sujet 6

Un autre exemple de cette tactique de reformulation par l'apport de précisions, est la session de recherche réalisée par le sujet 13 ; celui-ci avait pour objectif de dépanner un logiciel sur son ordinateur. Ce sujet a formulé deux requêtes dans sa session (tableau 18). D'abord, le sujet a formulé la requête initiale /windows live/ mais il n'a pas jugé la réponse satisfaisante. À ce niveau, et si on se réfère uniquement à la requête formulée, il n'est pas possible de savoir ce que le sujet a souhaité trouver comme informations sur ce logiciel. Ensuite, il a reformulé sa requête en tapant /msn bug/ :

- il a renommé le nom du logiciel en substituant /windows live/ par /msn/,
- il a rajouté le terme /bug/ exprimé dans la justification de la requête initiale.

Le tableau 18 ci-après présente la session de recherche réalisée par le sujet 13.

	Requête initiale n° 8	Extrait des verbalisations	Reformulation (requête 9)	Extrait des verbalisations
<b>Sujet 13</b>	windows live	« c'était un bug récent qui m'intéressait »	msn bug	« une solution à un problème, MSN marchait pas je voulais savoir comment faire pour qu'il remarque » [le sujet considère que MSN et Windows Live représentent « deux noms différents de la même chose »]

**Tableau 18.** Session de recherche réalisée par le sujet 13

Nous observons que lorsque la requête initiale ne permet pas d'obtenir des résultats satisfaisants, les sujets 6 et 13 ont réduit la part d'implicite dans leur requête en explicitant des termes décrivant leur besoin d'information.

**Des tactiques d'interrogation « interrogation par la classe » et « interrogation par le type »**

La session de recherche (tableau 19) réalisée par le sujet 28 représente la mise en œuvre de deux tactiques d'interrogation.

Session de recherche réalisée par le sujet n° 28	
requête n° 42	requête n° 43
/acides aminés/	/alanine/

**Tableau 19.** Session de recherche réalisée par le sujet 28

Le sujet 28 a interrogé le moteur de recherche en formulant une requête correspondant à la classe lexicale « acides aminés ». Pourtant, ses justifications montrent qu'il souhaitait savoir comment est formé un type d'acide aminé bien précis : la glycine<sup>4</sup>. Ainsi, le sujet n'a pas utilisé le terme exact qui l'intéressait et l'a donc laissé volontairement à l'état implicite dans sa première requête (requête 42). Ensuite, le sujet a formulé la deuxième requête (requête 43) /alanine/. Cette fois, il a choisi d'explicitier directement ce terme qui correspond précisément au type d'acide aminé qui l'intéresse.

Les verbalisations du sujet nous permettent d'en savoir davantage sur l'amont de cette formulation. Avec cette requête, celui-ci souhaitait savoir si l'acide aminé alanine était « naturellement une molécule en forme alpha ou bêta » et connaître « sa disposition dans l'espace ».

Nous constatons ainsi qu'un besoin d'information bien défini et caractérisé par des termes précis ne conduit pas nécessairement à la formulation d'une requête

<sup>4</sup> Cette information est apportée par l'exploitation des verbalisations associées à la requête 42

représentant clairement le besoin du sujet. Ici, seul le terme principal du besoin d'information a été formulé et les autres termes sont restés implicites.

#### ***Une tactique d'anticipation des résultats du moteur de recherche***

L'implicite n'est pas uniquement lié aux aspects sémantiques et pragmatiques de la demande d'information des usagers. La représentation du fonctionnement du moteur de recherche que se forgent les usagers reste à l'état d'implicite et vient influencer leurs tactiques dans la formulation/reformulation de la requête.

Si on prend pour exemple le sujet 38, celui-ci formule la requête /pénicilline/ et il dit souhaiter obtenir « l'opinion de *Wikipédia* sur les découvreurs de la pénicilline ».

Il souhaite donc atteindre le site de *Wikipédia*, et dans ce site, obtenir des informations sur un thème de recherche qui l'intéresse. Il saisit une seule requête qui amène le moteur de recherche Web à l'orienter directement vers la page traitant du thème de recherche qui l'intéresse dans *Wikipédia*. Ainsi, le sujet 38 savait que pour cette requête le moteur de recherche allait lui proposer un lien de *Wikipédia*. Cette représentation que le sujet s'est construit du moteur de recherche est implicite et l'a amené à anticiper les résultats du moteur de recherche Web. Autrement dit, cette requête remplit à la fois la fonction d'une requête navigationnelle et celle d'une requête informationnelle.

On peut aussi rappeler l'exemple du sujet 16 qui propose la requête /distraire/ et souhaite disposer de la conjugaison du verbe « distraire » sur un site précis « laconjugaison.com ». Ainsi, ce qui est visé est l'accès à la page du site contenant la conjugaison de ce verbe. Là encore, le sujet, de par sa propre connaissance du fonctionnement du moteur de recherche, en vient à anticiper les résultats et formule une requête remplissant la fonction d'une requête navigationnelle et d'une requête informationnelle. Là encore, la représentation que le sujet s'est construite du moteur de recherche est implicite et elle a influé sur sa stratégie de recherche.

## **5 Discussion**

Lors de cette étude empirique, nous n'avons pas pu interroger plus longuement les étudiants sur leur propre recherche, ceci afin de limiter la perturbation induite par notre observation pendant que les sujets réalisaient leur propre activité de recherche. Cependant, cette observation de la recherche des sujets en situation réelle nous a permis de disposer d'un corpus de requêtes riche. Ces requêtes ont été formulées afin de répondre à des besoins réels d'informations et non à des tâches de recherches prescrites (Aula, 2003), (Qu & Furnas, 2008).

Les requêtes produites sont assez courtes avec en moyenne deux termes par requête, ce qui représente un résultat similaire à celui observé dans d'autres travaux (Silverstein *et al.*, 1999 ; Jansen & Pooch, 2001). Mais ce chiffre moyen occulte des variations notables : ainsi, nous avons pu observer que la moyenne de termes par requête dans notre corpus varie selon la catégorie de la requête : ce sont les requêtes navigationnelles qui sont généralement plus courtes que les requêtes informationnelles et transactionnelles.

Il apparaît également que les requêtes navigationnelles et informationnelles sont de loin les plus nombreuses par rapport au faible nombre des requêtes transactionnelles. Ceci peut s'expliquer par le fait que l'observation s'est déroulée en période de révisions pour les étudiants, les sujets ayant participé à cette étude ont donc réalisé des recherches en lien avec leurs révisions et la préparation de dossiers et de travaux écrits.

Le fait de tenir compte à la fois de la requête et des justifications énoncées par le sujet montre d'une part, que la seule requête ne permet pas systématiquement d'inférer les intentions réelles des sujets et d'autre part que les catégories de Broder ne permettent pas de décrire l'ensemble des intentions et objectifs recueillis. Nous avons montré que dans le cas où les sujets utilisent le moteur de recherche comme un annuaire (requête navigationnelle), l'accès au site ne représentait que leur sous-but et l'utilisation d'un service précis proposé par ce site représentait leur but. Par exemple, le sujet 58 souhaitait localiser un site de transport (sous-but) afin d'accéder à la fonctionnalité itinéraire que propose ce site (but).

Nos résultats montrent donc que les requêtes comportent une part d'implicite, c'est-à-dire une part d'information manquante tant au niveau sémantique qu'au niveau pragmatique. Cependant, il semble exister des différences entre les trois catégories de requêtes : ainsi les requêtes informationnelles et transactionnelles peuvent receler une part d'implicite au niveau sémantique et au niveau pragmatique, alors que les requêtes navigationnelles recèlent de l'implicite plutôt au niveau pragmatique et certaines de ces requêtes ne formulent qu'une description incomplète du nom du site visé. De plus, cette part d'implicite dans les requêtes varie lorsque la session de recherche comporte au moins deux requêtes. Ainsi, nous avons observé que des sujets ont davantage explicité leur besoin d'information lors de la reformulation.

Il nous semble que l'apport de Broder (2002) réside moins dans sa catégorisation des requêtes en trois catégories informationnelle, navigationnelle et transactionnelle que dans l'idée de considérer que les requêtes peuvent être rassemblées dans des catégories plus larges, même si elles sont motivées par des intentions de recherche différentes. Cela permet d'introduire une démarche qualitative dans l'analyse des corpus de requêtes. Cependant, il nous semble qu'une telle démarche ne peut être possible sans une phase d'entretien amenant les sujets à exprimer leurs intentions, leurs objectifs et leurs besoins.

Les sujets observés ont principalement réalisé des sessions de recherche avec une requête unique (91,48% des sessions réalisées sont composées d'une seule requête) et ils étaient satisfaits de leur recherche et des résultats obtenus. Ceci représente peut-être une caractéristique de la population observée, néanmoins les sessions comportant plus d'une requête suggèrent quelques remarques :

- Sur la production des requêtes : les sujets ne rajoutent pas systématiquement des termes supplémentaires lors de leur reformulation : le nombre de termes par requête semble compter moins que la qualité des termes utilisés. Par exemple, la requête initiale du sujet 13 est */windows live/*. Cette première requête était trop large, le sujet a, par la suite, formulé une seconde requête également composée de deux termes */msn bug/* mais décrivant mieux son problème informationnel.
- Sur la session de recherche, considérée dans sa totalité : ce cadre d'analyse tient compte à la fois de la pertinence-usager et de la dimension temporelle durant l'activité de recherche d'information ; il permet d'approcher la manière dont le sujet adapte son interrogation en passant d'une requête initiale à une requête suivante : l'évolution du besoin d'information (Bates, 1989 ; Le Coadic, 1998) se répercute dans l'écriture de la requête suivante en termes de modifications, reformulations totales ou partielles, simples précisions, changements complets.

- S'agissant de la catégorisation des requêtes selon l'intention du sujet, nos résultats montrent que les catégories de Broder ne permettent pas de cerner la diversité des usages et des pratiques actuels des moteurs de recherche sur le web. Une meilleure catégorisation des requêtes selon l'intention des sujets ne peut reposer que sur un corpus de sessions de recherche plus étendu. Il serait utile d'étendre cette étude soit en visant le même public mais à un autre moment de l'année ou dans un environnement différent, soit encore en observant un public différent.

## **6 Conclusion**

La dimension temporelle inhérente à une activité de recherche d'information, l'évolutivité du besoin d'information et le rôle des jugements de pertinence émis par l'utilisateur ne sont abordables que par l'analyse des sessions de recherche.

La catégorisation des requêtes s'appuyant à la fois sur les termes de la requête et les justifications de l'utilisateur montre des usages, des tactiques plus diversifiés que le modèle de Broder.

Les requêtes sont courtes, incomplètes et ne formulent pas l'intention de recherche de l'utilisateur, même lorsque ce dernier a une représentation précise de son besoin d'information. Ce sont les justifications apportées par l'utilisateur et le déroulement de la session qui montrent l'étendue de l'implicite dans les requêtes, implicite que l'utilisateur peut être amené à expliciter plus ou moins selon le développement de la session de recherche, en fonction de la pertinence des résultats obtenus.

Nous avons montré que les requêtes recèlent une part importante d'implicite et qu'elles sont généralement incomplètes. Dans une situation d'interaction Homme-machine, le besoin d'information de l'utilisateur reste hors de la portée du moteur de recherche. La pauvreté informationnelle de la requête pourrait être en partie diminuée avec la généralisation de la recherche à facette. Celle-ci offre à l'utilisateur la possibilité d'affiner les résultats obtenus suite à une requête initiale en sélectionnant des critères supplémentaires tels que la date de publication, la source des résultats, la langue et le type de documents.

Il semble également important que les bibliothèques universitaires poursuivent leurs efforts de formation à la recherche documentaire et continuent à montrer la richesse de leur fonds et son adéquation avec le niveau d'étude et la spécialité des étudiants. Une telle démarche aurait l'intérêt de mettre en avant des outils académiques pouvant répondre à des besoins d'information d'étudiants en lien avec le programme universitaire suivi.

## **Bibliographie**

Allan, J., Aslam, J., Belkin, N., Buckley, C., Callan, J., Croft, B., Dumais, S., Fuhr, N., Harman, D., Harper, D.J., Hiemstra, D., Hofmann, T., Hovy, E., Kraaij, W., Lafferty, J., Lavrenko, V., Lewis, D., Liddy, L., Mammatha, R., McCallum, A., Ponte, J., Prager, J., Radev, D., Resnik, P., Robertson, S., Rosenfield, R., Roukos, S., Sanderson, M., Schwartz, R., Singhal, A., Smeaton, A., Turtle, H., Voorhees, E., Weischedel, R., Xu, J. & Zhai, C. (2003). Challenges in information retrieval and language modeling: Report of a workshop held at the center for intelligent information retrieval, *SIGIR Forum*, New-York: ACM, vol. 37, 31-47.

- Aula, A. (2003). Query formulation in web information search. In *International Conference WWW/Internet*, 403–410.
- Bates, M. (1979). Information search tactics. *Journal of the American Society for Information Science*, vol. 30, num. 4, 205–214.
- Bates, M. (1989). The design of browsing and berrypicking techniques for the online search interface. *Online Information Review*, vol. 13, num. 5, 407–424.
- Belkin, N., Oddy, R. & Brooks, H. (1982). ASK for information retrieval. Part II: results of a design study. *Journal of Documentation*, vol. 38, num. 3, 145–164.
- Borlund, P. (2003). The concept of relevance in IR. *Journal of the American Society for Information Science and Technology*, vol. 54, num. 10, 913-925.
- Boubée, N. & Tricot, A. (2010). *Qu'est-ce que rechercher de l'information ?* Presses de l'ENSSIB, Lyon.
- Bouchard, H. & Nie, J-Y. (2006). Modèles de langue appliqués à la recherche d'information contextuelle. In *Actes de la Conférence en Recherche d'Information et Application*, CORIA, Lyon.
- Brini, A.H., Boughanem, M. & Dubois, D. (2004). Une Approche Possibiliste pour la Recherche d'Information. In *Veille Stratégique Scientifique et Technologique*, VSST'2004.
- Broder, A. (2002). A taxonomy of web search. *SIGIR Forum*, vol. 36, num. 2, 3–10.
- Chau, M., Fang, X. & Yang, C. (2007). Web searching in chinese: A study of a search engine in Hong Kong. *Journal of the American Society for Information Science and Technology*, vol. 58, num. 7, 1044–1054.
- Deveaud, R & Bellot, P. (2012). Combinaison de ressources générales pour une contextualisation implicite de requêtes. In *Actes de la conférence conjointe JEP-TALN-RECITAL'2012*, TALN, Grenoble, vol. 2, 479–486.
- Ducrot, O. (1984). *Le dire et le dit*. Les Éditions de Minuit, Paris.
- Harman, D. (1992). Relevance feedback revisited. In *Proceedings of the 15th annual international ACM SIGIR*, Conference on Research and development in information retrieval, SIGIR'92.
- Ihadjadene, M. & Chaudiron, S. (2008). L'étude des dispositifs d'accès à l'information électronique : approches croisées. In *Problématiques émergentes dans les sciences de l'information*, Papy, F (Eds.), *Traité des sciences et techniques de l'information*, Lavoisier, Paris, 183–207.
- Ingwersen, P. (1992). *Information Retrieval Interaction*. Taylor Graham, London.
- Jansen, B. J. & Pooch, U. (2001). A Review of Web Searching Studies and a Framework for Future Research. *Journal of the American Society for Information Science and Technology*, vol. 52, num. 3, 235-246.
- Jansen, B. & Spink, A. (2006). How are we searching the world wide web? a comparison of nine search engine transaction logs. *Information Processing & Management*, vol. 42, num. 1, 248–263.
- Kerbrat-Orecchioni, C. (1986). *L'implicite*. Armand Colin, Paris.

- Le Coadic, Y.-F. (1998). *Le besoin d'information Formulation, négociation, diagnostic*. Sciences et techniques de l'information, ADBD, Paris.
- Le Petit Robert (2004). Dictionnaires Le Robert. Paris.
- Lewandowski, D., Drechsler, J. & Von Mach, S. (2012). Deriving query intents from web search engine queries. *Journal of the American Society for Information Science and Technology*, vol. 63, num. 9, 1773–1788.
- Li, Y., Luk, W. P. R., Ho, K. S. E. & Chung, F. L. K. (2007). Improving weak ad-hoc queries using wikipedia as external corpus. In *Proceedings of the 30th annual international ACM SIGIR*. Conference on Research and development in information retrieval, SIGIR '07.
- Pitkow, J., Schütze, H., Cass, T., Cooley, R., Turnbull, D., Edmonds, A., Adar, E. & Breuel, T. (2002). Personalized Search, *Communications of ACM*, vol. 45, 50-55, New-York : ACM.
- Raeber, T. (2011). Explicite/Implicite/Sous-entendu. Heurisis, Disponible à : [www.heurisis.ch/explicite-implicite-sous-entendus](http://www.heurisis.ch/explicite-implicite-sous-entendus).
- Schamber, L., Eisenberg, M.B. & Nilan, M.S. (1990). A re-examination of relevance: Toward a dynamic, situational definition. *Information Processing & Management*, vol. 26, 755–775.
- Silverstein, C., Henzinger, M., Marais, H. & Moricz, M. (1999). Analysis of a very large web search engine query log. *ACM SIGIR Forum*, vol. 33, num. 3, 6-12.
- Simonnot, B. (2006). Le besoin d'information : principes et compétences. In *Journée d'étude Thémat'IC « Information : besoins et usages »*, Strasbourg.
- Suchanek, F., Kasneci, G. & Weikum, G. (2007). Yago: a core of semantic knowledge. In *Proceedings of the 16th international conference on World Wide Web*, WWW'07, 697–706.
- Taylor, R.S. (1968). Question negotiation and information seeking in libraries. *Journal of the American Society for Information Science and Technology*, vol. 29, num. 3, 178-194.
- TLFi, X. (1994). Incomplétude. CNRTL, Disponible à : [www.cnrtl.fr/lexicographie/incomplétude](http://www.cnrtl.fr/lexicographie/incomplétude).
- Qu, Y. & Furnas, G.W. (2008). Model-driven formative evaluation of exploratory search : A study under a sensemaking framework. *Information Processing and Management*, vol. 44, num. 2, 534–555.