

Comment les spécialistes scientifiques interrogent-ils les bases de données bibliographiques en ligne ? Une étude exploratoire de la recherche de références par des chercheurs en neurosciences

Nicolas VIBERT

Laboratoire de Neurobiologie des Réseaux Sensorimoteurs, CNRS UMR 7060 – Université René Descartes Paris 5, Paris. Courriel : niv@ccr.jussieu.fr

Jean-François ROUET, Christine ROS, Mélanie RAMOND

Laboratoire Langage, Mémoire et Développement Cognitif, CNRS UMR 6215 – Université de Poitiers, Poitiers. Courriels : jean-francois.rouet@univ-poitiers.fr, christine.ros@univ-poitiers.fr, melanieramond@yahoo.fr

Introduction

Les systèmes documentaires en ligne ont connu un fort développement depuis l'avènement de la micro-informatique (Marchionini, 1995). Le recours à des bases de données et logiciels de recherche pour accéder aux références bibliographiques devient courant dans tous les domaines scientifiques et d'ingénierie (Dalglish & Hall, 2000 ; De Groote & Dorsch, 2003). Les études existantes montrent que la recherche bibliographique et documentaire mobilise des connaissances et processus cognitifs spécifiques, en partie indépendants des connaissances du domaine (Borgman, 1986 ; Marchionini, 1995 ; Marchionini, Dwiggins, Katz, et Lin, 1993). L'efficacité de la recherche d'informations, en particulier sur le Web, serait notamment liée à l'expertise d'utilisation du système, qui inclut la connaissance des procédures et outils documentaires, mais aussi celle des heuristiques et stratégies à utiliser (Colombi & Baccino, 2003 ; Hölscher & Strube, 2000 ; Sutcliffe, Ennis, et Watkinson, 2000). Rouet, Macedo-Rouet, Bouin, Deshouillères, et Menu (2004) ont montré que des étudiants spécialisés en documentation évaluent plus rapidement et correctement que les autres la qualité du contenu et la crédibilité de documents multimédias, même si les non-documentalistes ont une connaissance préalable du domaine dont relèvent les documents.

Qu'en est-il de l'expertise de recherche documentaire chez des publics possédant une très forte spécialisation, comme les chercheurs en neurosciences intégratives ? Peu de publications se sont intéressées aux spécialistes de ce niveau (Ondrusek, 2004). Pour tenter de répondre à cette question, nous avons d'abord mené une enquête par questionnaires et entretiens auprès d'une population de 64 doctorants et chercheurs en neurosciences. Les résultats (Rouet, Vibert, Ros, Ramond, et Deshoullières, 2005 ; Vibert, Rouet, Ros, Ramond, et Deshoullières, 2006) confirment le caractère familier des outils et procédures de recherche documentaire informatisés pour ces spécialistes, et suggèrent que l'usage de l'ordinateur rend quasiment obsolètes les moyens traditionnels de représentation écrite de l'information. Un autre résultat intéressant est que la base de données d'accès public et gratuit PubMed (qui regroupe plus de 16 millions de références dans le domaine des Sciences de la Vie, <http://www.pubmed.gov>) peut être qualifiée d'outil de recherche bibliographique standard pour les spécialistes en neurosciences. Toutefois, l'utilisation de telles bases de données ne va pas toujours de soi. D'après les entretiens, les difficultés rencontrées dans l'exploitation de PubMed ou des autres outils spécialisés semblent largement liées à l'absence de formation des chercheurs à ces outils. Les contraintes temporelles liées à l'activité professionnelle apparaissent déterminantes dans le choix et le mode d'utilisation d'un outil particulier, car doctorants et chercheurs privilégient les outils de recherche d'information faciles et rapides aussi bien à prendre en main qu'à utiliser ensuite quotidiennement.

Suite à cette enquête, nous avons mené une étude exploratoire sur les comportements de 12 spécialistes en neurosciences dans des tâches de recherche de références bibliographiques dans leur champ d'expertise, en utilisant la technique des protocoles verbaux concomitants (think aloud protocol, voir ci-dessous). Nous souhaitons connaître les procédures et stratégies qu'ils utilisent lorsqu'ils recherchent des références bibliographiques. Les participants devaient employer la base de données PubMed, mais étaient libres d'utiliser toutes les fonctionnalités proposées par l'interface. Cette interface fonctionne selon le principe usuel d'interrogation de bases de données par insertion de mots-clés dans une barre de requête, avec un ensemble de possibilités de limitations du champ de recherche (par exemple, restriction aux articles de synthèse). L'interface retourne des listes de références, dont les résumés peuvent être ouverts un par un d'un simple clic.

Méthode

Les 12 spécialistes incluait 3 enseignants-chercheurs (1 professeur des universités et 2 maîtres de conférence), 5 chercheurs (1 directeur de recherche et 4 chargés de recherche), 2 doctorants, 1 post-

doctorant et 1 ingénieur d'étude inscrit en doctorat. Ils se répartissaient en 8 hommes et 4 femmes, d'âge moyen = 36,7 ans ($ET = 5,9$). Ces 12 spécialistes utilisaient PubMed depuis 4 à 11 ans ($M = 8,4$, $ET = 2,5$). En dehors des 3 doctorants et post-doctorant, l'expertise de PubMed des 9 autres spécialistes était très similaire, puisque tous utilisaient cette base de données depuis 8 à 11 ans. Aucun des spécialistes testés n'avait reçu de formation à l'usage des outils informatiques de recherche bibliographique. Le niveau d'expertise en neurosciences des 12 spécialistes n'a pas formellement été utilisé comme variable indépendante, car nous n'avions pas d'hypothèse a priori concernant l'effet des différences d'expertise entre ces spécialistes sur l'efficacité de leur recherche de références. Quelques analyses descriptives ont cependant été menées en fonction des grades des participants, en attribuant l'expertise du domaine la plus faible aux 4 doctorants, post-doctorant et ingénieur d'étude inscrit en doctorat (« chercheurs en cours de formation »), l'expertise la plus forte au professeur des universités et au directeur de recherche (2 « chercheurs seniors »), et une expertise intermédiaire aux 6 chargés de recherche et maîtres de conférence (« chercheurs juniors »)

Les expérimentations ont eu lieu sur rendez-vous, au lieu de travail de chaque participant, et ont été réalisées par un seul expérimentateur. Chaque participant devait réaliser les mêmes tâches de recherche et de sélection de références bibliographiques en utilisant PubMed, tout en verbalisant ses activités mentales et en explicitant ses actions avec l'interface (think aloud protocol). L'interaction de l'expérimentateur se limitait à des relances lorsque le sujet était silencieux. Le temps consacré à chaque tâche était limité à 15 minutes. Après 2 tâches d'entraînement pour vérifier la bonne compréhension des consignes et habituer le sujet à verbaliser ses activités, les 5 tâches de recherche de références présentées à la suite de ce paragraphe (T1 à T5) étaient proposées dans un ordre aléatoire spécifique pour chaque participant. Ces tâches ont été choisies selon la spécificité de leur vocabulaire évaluée par un expert du domaine (*courant, plutôt spécifique* aux neurosciences, ou *spécifique* aux neurosciences), et le nombre de mots-clés du libellé (soulignés ci-dessous). Dans la mesure où il s'agissait d'une étude exploratoire et du fait du temps limité que les spécialistes pouvaient consacrer à notre expérience, nous n'avons pas mis en place un plan expérimental rigoureux pour ces deux variables.

T1 : « Trouver 1 article de synthèse traitant des mécanismes d'action cellulaire de l'alcool sur les neurones du système nerveux central » (vocabulaire courant, 8 mots-clés + article de synthèse).

T2 : « Trouver 2 articles de 2004 portant sur les liens entre neurogenèse et maladie d'Alzheimer » (vocabulaire plutôt spécifique, 4 mots-clés + année de publication).

T3 : « Trouver 1 article traitant des mécanismes neuronaux du pointage visuo-guidé du bras » (vocabulaire plutôt spécifique, 5 mots-clés).

T4 : « Trouver 2 articles de synthèse décrivant des modèles animaux de dyskinésies tardives » (vocabulaire spécifique, 4 mots-clés + article de synthèse).

T5 : « Trouver 1 article abordant l'utilisation chez l'animal des vecteurs viraux pour injecter dans le cerveau le gène du neuropeptide Y » (vocabulaire spécifique, 7 mots-clés).

L'expérimentateur enregistrait en audio les verbalisations concomitantes à la recherche des références, et recueillait la ou les références sélectionnées par chaque sujet en réponse à la question initiale. Le libellé de chaque tâche était présenté en français et anglais afin d'éviter les erreurs de traduction. La connaissance préalable de la terminologie utilisée en anglais n'a pas été vérifiée, mais peut être considérée comme bonne pour l'ensemble des participants du fait de leur haut niveau de spécialisation dans le domaine des neurosciences, domaine où les publications originales se font quasiment toutes dans cette langue. Le temps de lecture du libellé de la tâche n'était pas contraint, mais ce libellé n'était pas présenté en permanence. Lorsque le sujet avait fini de lire, l'expérimentateur retournait la feuille et la tâche débutait. Au cours de la recherche, le sujet pouvait demander à relire le libellé de la tâche à tout moment, mais devait alors s'interrompre jusqu'à ce que le libellé soit de nouveau caché. La pertinence des références sélectionnées comme réponse(s), a été évaluée pour chaque tâche et chaque sujet par une note allant de 0 à 5. Cette évaluation a été menée de manière indépendante par deux juges experts en neurosciences, qui ne faisaient pas partie des spécialistes testés. Sur l'ensemble des tâches et des spécialistes, une corrélation positive a été obtenue entre les notes attribuées par chaque juge ($r = 0,90$, $p < 0.001$).

Seule la variable indépendante « tâches » a été utilisée pour les analyses statistiques. Les variables dépendantes étaient le temps de lecture initiale du libellé de la tâche, le fait ou non de sélectionner au moins une référence comme réponse dans le temps imparti, le temps total requis pour la sélection des références (en secondes), le nombre de relectures du libellé de la tâche au cours de la recherche des références, et les notes évaluant la pertinence des références sélectionnées comme réponse (calculées en faisant la moyenne des scores attribués par les deux juges).

Parmi les nombreuses données extraites des protocoles verbaux, encore en cours de dépouillement, nous ne présenterons ici que les résultats concernant les modes de formulation des requêtes soumises à PubMed et le choix des mots-clés utilisés pour ces requêtes.

Une fois les tâches accomplies, chaque sujet remplissait un questionnaire sur ses connaissances de l'interface de PubMed et de ses fonctionnalités. La première question portait sur la définition des 4 icônes présentes à gauche des références dans les listes retournées par l'interface, qui indiquent la disponibilité de chaque publication (accès depuis PubMed au résumé seul, au texte complet, etc...), et était notée de 0 à 2. Les 8 questions suivantes, notées de 0 à 3, portaient chacune sur une

fonctionnalité distincte de l'interface. Pour chacune, les spécialistes devaient dire s'ils la connaissaient, expliquer son utilisation, puis faire eux-mêmes une démonstration de son utilisation avec l'interface. Le score sur 26 obtenu en additionnant les scores de ces 9 questions a été utilisé comme variable dépendante pour évaluer le niveau de connaissance des fonctionnalités de PubMed par chaque participant.

Résultats

Performances des spécialistes dans les tâches de recherche d'information bibliographique

Le temps moyen de lecture initiale du libellé de chaque tâche (en secondes) par les 12 spécialistes variait de $M = 15,75$, $ET = 5,66$ pour T2 à $M = 29,00$, $ET = 7,66$ pour T5. L'ANOVA a montré une différence de réussite significative entre les tâches ($F(4,44) = 7,97$, $p < 0,001$). Les tests post-hoc de Newman-Keuls ont indiqué que le temps de lecture du libellé de T5 était significativement supérieur à celui de toutes les autres tâches, et que le temps de lecture de T3 ($M = 22,08$ secondes, $ET = 7,90$) était supérieur à celui de T4 ($M = 16,67$ secondes, $ET = 6,98$). Les temps de lecture initiale des libellés s'ordonnent suivant la séquence T2 (4 mots-clés) < T4 (4 mots-clés) < T3 (5 mots-clés) < T1 (8 mots-clés) < T5 (7 mots-clés), ce qui suggère une corrélation positive avec le nombre de mots-clés.

Onze des 12 spécialistes ont réalisé toutes les tâches sans dépasser le temps maximum de 15 minutes. Un seul des participants (un chargé de recherche) n'a pu donner de réponse pour une des tâches (T1).

Le temps moyen de réalisation de chaque tâche (en secondes) allait de $M = 283$, $ET = 141$ pour T5 à $M = 493$, $ET = 193$ pour T1. L'ANOVA a montré une différence significative entre les tâches ($F(4,40) = 3,34$, $p = 0,02$). Les tests post-hoc de Newman-Keuls ont indiqué que le temps de réalisation de T1 était significativement supérieur à celui de T5 ($p = 0,01$). Les temps de réalisation de chaque tâche s'ordonnent suivant la séquence T5 (7 mots-clés, vocabulaire spécifique) < T3 (5 mots-clés, vocabulaire plutôt spécifique) < T4 (4 mots-clés, vocabulaire spécifique) < T2 (4 mots-clés, vocabulaire plutôt spécifique) < T1 (8 mots-clés, vocabulaire courant), et ne semblent donc pas dépendre de la spécificité du vocabulaire utilisé ou du nombre de mots-clés du libellé. La diminution presque continue du temps de réalisation de chaque tâche de T1 à T5 ne résulte pas d'un effet d'apprentissage, puisque l'ordre de présentation des tâches était aléatoire.

Le nombre moyen de relecture du libellé de chaque tâche pendant sa réalisation variait de $M = 0,92$, $ET = 1,08$ pour T5 à $M = 1,50$, $ET = 0,90$ pour T1. L'ANOVA n'a montré aucune différence

significative entre les tâches ($F(4,40) = 1,16, p = 0,34$). En fonction des participants, le nombre de relectures du libellé variait de 0 pour toutes les tâches à 2 pour T2, 3 pour T1, T3, et T4, et 4 pour T5. Sur les 5 tâches et pour chacun des spécialistes, le nombre total de relectures des libellés variait de 1 à 11 ($M = 5,58, ET = 3,42$). Si l'on sépare les participants selon les trois niveaux d'expertise définis dans la section Méthode, on obtient un nombre moyen de relectures similaires pour les chercheurs en cours de formation ($M = 6,00, ET = 3,16$), les chercheurs juniors ($M = 5,17, ET = 3,06$) et les chercheurs seniors ($M = 6,00, ET = 7,07$). Le nombre de relectures ne semble donc pas dépendre du niveau d'expertise en neurosciences de chaque participant.

Les notes moyennes de pertinence des références sélectionnées obtenues par les 12 participants sont indiquées dans le tableau 1.

	Tâche 1	Tâche 2	Tâche 3	Tâche 4	Tâche 5	Total (sur 25)
moyenne	4,2	4,5	4,4	3,7	3,9	20,7
écart-type	1,4	0,7	1,0	1,4	1,6	3,5

Tableau 1 : Moyennes et écarts-types des notes de pertinence des références sélectionnées obtenues par les 12 spécialistes pour chacune des 5 tâches proposées.

L'ANOVA n'a montré aucune différence de réussite significative entre les tâches ($F(4,44) = 0,99, p = 0,43$). Cette réussite ne serait donc pas liée à la spécificité du vocabulaire utilisé ou au nombre de mots-clés inclus dans les libellés des tâches. Dans l'ensemble, les spécialistes obtiennent de très bons résultats, avec une note moyenne par tâche de 4,14 sur 5. La note globale sur 25 obtenue par chaque participant varie de 14,25 à 24,5. Si l'on sépare les participants selon leur niveau d'expertise, on obtient des notes moyennes très similaires pour les chercheurs en cours de formation ($M = 21,9, ET = 3,8$), les chercheurs juniors ($M = 19,8, ET = 3,2$) et les chercheurs seniors ($M = 20,9, ET = 5,1$). L'efficacité de la recherche de références ne semble donc pas dépendre du niveau d'expertise en neurosciences des participants.

Questionnaire d'évaluation des connaissances de l'interface de PubMed

Le nombre d'icônes reconnues par les 12 spécialistes était très variable, puisque 4 n'en connaissaient aucune (2 chercheurs en formation, 1 chercheur junior, 1 chercheur senior) et 1 ne connaissait qu'1 icône, tandis que 4 spécialistes connaissaient 3 icônes et que les 3 derniers (2 chercheurs en cours de formation et 1 chercheur junior) connaissaient les 4 icônes. Le niveau global de connaissance des fonctionnalités de PubMed variait lui aussi largement et sur l'ensemble du questionnaire, les 12 spécialistes ont obtenu des scores allant de 6,5 à 24 (sur 26) pour une moyenne

de 13,9 ($ET = 5,5$). Nous avons vu que malgré cette connaissance très variable de l'outil, tous ont obtenu de bons résultats dans les tâches de recherche de références. De fait, la corrélation positive observée entre les scores de réussite aux tâches et le niveau de connaissance des fonctionnalités de PubMed n'est pas significative ($r = 0,46$, $p = 0,13$). Si l'on sépare les participants selon leur niveau d'expertise, on obtient des niveaux de connaissance de PubMed moyens (sur 26) très similaires pour les chercheurs en cours de formation ($M = 15,4$, $ET = 7,0$), les chercheurs juniors ($M = 12,8$, $ET = 6,0$) et les chercheurs seniors ($M = 13,3$, $ET = 5,3$). Le niveau d'expertise des participants spécialistes en neurosciences ne semble pas influencer sur leur niveau de connaissance de PubMed. Ce sont même les chercheurs en cours de formation, qui utilisent pourtant PubMed depuis moins longtemps que les autres (en moyenne 5,5 années, au lieu de 8 à 11 ans pour les chercheurs en poste), qui obtiennent le meilleur score de connaissance des fonctionnalités de cet outil.

Résultats issus des verbalisations

Dans la formulation de leurs requêtes, 6 spécialistes sur 12 n'ont utilisé que des mots-clés simples juxtaposés (que PubMed considère par défaut comme reliés par le booléen « ET »), ou liés explicitement par « ET ». Les 6 autres ont éventuellement ajouté une limitation du champ de recherche pour obtenir, par exemple, uniquement les revues de question. Parmi ces 6 participants, 2 seulement ont parfois utilisé des modes d'interrogation plus avancés impliquant des fonctions évoluées de l'interface. Aucun cas d'utilisation de booléen autre que « ET » n'a été observé. Notons que si l'on considère l'ensemble des mots-clés utilisés par les 12 spécialistes pour formuler leurs requêtes dans les différentes tâches, on constate que dans 97,5% des cas ces mots-clés étaient soit inclus dans le libellé de la tâche en cours (87%), soit directement dérivés de mots de ce libellé, comme dans les cas d'utilisations de « cell », « visual » ou « brain » au lieu de « cellular », « visually » et « central nervous system ». Ce n'est donc que dans 2,5% des cas que 4 des spécialistes (1 doctorant et 3 chercheurs juniors, mais aucun des chercheurs seniors) ont utilisé des mots-clés « innovants », en faisant appel à leur expertise du domaine. Là encore, il ne semble pas que le niveau d'expertise de ces spécialistes influence la manière dont les spécialistes élaborent leurs requêtes.

Conclusion et Perspectives

Cette étude comportementale, conçue comme une étude exploratoire, révèle que lorsqu'ils doivent rechercher avec PubMed des références bibliographiques dans leur domaine de compétences, les spécialistes en neurosciences sont dans l'ensemble très efficaces. L'efficacité de la recherche et le

nombre de relectures du libellé de la tâche pendant la recherche sont similaires pour les 5 tâches que nous avons proposées, indépendamment du niveau de spécificité du vocabulaire utilisé ou du nombre de mots-clés caractérisant chaque tâche. L'efficacité de la recherche ne dépend par ailleurs apparemment pas ou peu de la connaissance qu'ont les spécialistes de l'outil PubMed, qui est elle très variable d'un spécialiste à l'autre, probablement du fait de leur absence de formation aux outils informatiques de recherche bibliographique (Rouet et al., 2005). Le temps de lecture initiale du libellé et le temps moyen de complétion de chaque tâche diffèrent en fonction de la tâche, mais cette variabilité ne semble pas liée à la spécificité du vocabulaire utilisé, et le nombre de mots-clés du libellé ne semble influencer que le temps de lecture initiale du libellé de la tâche.

L'analyse descriptive des résultats obtenus pour chacun des spécialistes suggère que ni le niveau de connaissance des fonctionnalités de PubMed, ni l'efficacité de la recherche de références ou le mode opératoire mis en œuvre ne dépendent de leur niveau d'expertise du domaine. Le niveau de connaissance des fonctionnalités de PubMed ne semble pas non plus lié au temps depuis lequel chaque spécialiste utilise PubMed dans le cadre professionnel.

Il est par ailleurs frappant de constater que la plupart des participants n'utilisent que les fonctionnalités et les modes d'interrogation les plus simples de PubMed. De plus, lorsque des mots-clés sont fournis par le libellé des tâches, les spécialistes se contentent très largement de reprendre ces mots-clés dans leurs requêtes sans en générer de nouveaux en utilisant leur expertise. Ces données sont probablement à lier au fait que les spécialistes souhaitent que leurs recherches d'information bibliographique et documentaire soient les plus faciles et rapides possibles (Vibert et al., 2006).

L'ensemble de ces données suggère qu'une expertise de haut niveau dans un domaine scientifique (de niveau doctoral ou supérieur) est capable de compenser une connaissance parfois superficielle des outils informatiques de recherche d'information. Cette expérience est en train d'être complétée par une étude contrôle où les mêmes tâches de recherches de références dans le domaine des neurosciences seront proposées à des biologistes non spécialistes en neurosciences intégratives, mais utilisateurs réguliers de PubMed, et de niveau équivalent aux spécialistes en neurosciences. Nous nous attendons à ce que ces non spécialistes en neurosciences soient beaucoup moins efficaces que les spécialistes du domaine dans la recherche de références en neurosciences. Nous espérons ainsi pouvoir séparer les influences respectives de l'expertise du domaine et de l'expertise du système sur l'efficacité et le mode opératoire de la recherche d'informations par les scientifiques de haut niveau. Nous envisageons par ailleurs des expériences ultérieures sur l'emploi des

opérateurs booléens autres que ET dans les requêtes, où nous poserions des problèmes devant théoriquement conduire les participants à utiliser d'autres opérateurs comme SAUF et OU.

Références bibliographiques

- Borgman, C.L. (1986). The user's mental model of an information retrieval system: An experiment on a prototype online catalog. *International Journal of Man-Machine Studies*, 24, 47-64.
- Colombi, T., & Baccino, T. (2003). Le rôle de la syntaxe et de la structure syntaxique dans la sélection des liens hypertextuels. *Le Travail Humain*, 66, 45-64.
- Dalglish, A., & Hall, R. (2000). Uses and perceptions of the World Wide Web in an information-seeking environment. *Journal of Library and Information Science*, 32, 104-116.
- De Groot, S.L., & Dorsch, J.L. (2003). Measuring use patterns of online journal and databases. *Journal of the Medical Library Association*, 91, 231-240.
- Hölscher, C., & Strube, G. (2000). Web search behavior of internet experts and newbies. *Computer Networks*, 33, 337-346.
- Marchionini, G. (1995). *Information seeking in electronic environments*. Cambridge University Press.
- Marchionini, G., Dwiggins, S., Katz, A., & Lin, X. (1993). Information seeking in full-text end-user-oriented search systems: The roles of domain and search expertise. *Library and Information Science Research*, 15, 35-69.
- Ondrusek, A.L. (2004). The attributes of research on end-user online searching behavior: a retrospective review and analysis. *Library and Information Science Research*, 26, 221-265.
- Rouet, J.F., Macedo-Rouet, M., Bouin, E., Deshouillères, B., & Menu, P. (2004, Août). *Effects of discipline expertise in the assessment of information quality and source credibility on the Web*. Paper presented at the annual meeting of the Society for Text and Discourse, Chicago, Illinois, USA.
- Rouet, J.F., Vibert, N., Ros, C., Ramond, M., & Deshouillères, B. (2005). L'usage des systèmes d'informations électroniques en recherche scientifique : le cas de la neurophysiologie. Dans E. Raufaste & A. Tricot (Eds.), *Actes des 3èmes Journées d'Etude en Psychologie Ergonomique EPIQUE* (pp. 71-80). Toulouse, France : Université de Toulouse – Le Mirail.
- Sutcliffe, A.G., Ennis, M., & Watkinson, S.J. (2000). Empirical studies of end-user information searching. *Journal of the American Society for Information Science*, 51, 1211-1231.

Vibert, N., Rouet, J.F., Ros, C., Ramond, M., & Deshoullières, B. (2006). *The use of online electronic information systems in scientific research: the case of neurophysiology*. Manuscrit soumis pour publication.