



HAL
open science

La sélection collaborative de pages Web pertinentes

Jérôme Dinet

► **To cite this version:**

| Jérôme Dinet. La sélection collaborative de pages Web pertinentes. 2005. hal-00005767

HAL Id: hal-00005767

<https://telearn.hal.science/hal-00005767>

Preprint submitted on 1 Jul 2005

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

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

La sélection collaborative de pages Web pertinentes

Jérôme Dinet

Université de Metz

Equipe Transdisciplinaire sur l'Interaction et la Cognition (ETIC)

UFR sciences humaines et art

Ile du Saulcy

57045 Metz Cedex

dinet@sha.univ-metz.fr

RÉSUMÉ. *A l'aide d'une étude expérimentale, nous avons testé les impacts d'un dispositif informatique simple pouvant aider à la recherche collaborative de pages Web pertinente chez les étudiants. Dans l'une des conditions expérimentales, les membres d'un même groupe d'étudiants pouvaient avoir des indications quant aux pages Web déjà visitées par les autres membres de leur groupe. Plusieurs indicateurs comportementaux et de performances des utilisateurs (nombre de pages consultées, temps pour réaliser la tâche, pertinence des pages retenues, etc.) ont été examinés. Trois tâches de recherche de pages Web ont été définies, ces trois tâches étant de difficulté croissante. Nos résultats ont principalement montré que les indications fournies à chaque membre du groupe entraînaient une diminution significative d'une part, du temps mis pour réaliser l'activité et d'autre part, du pourcentage de re-visites de pages Web, mais uniquement lorsque le thème était difficile. Les implications pédagogiques sont discutées à la fin de l'article.*

MOTS-CLÉS: *analyse et évaluation des usages ; travail de groupe ; outil de travail collaboratif ; recherche d'informations.*

1. L'aide à la recherche collaborative d'informations sur le Web

Quels que soient les niveaux et cursus universitaires considérés, la recherche collaborative d'informations sur le Web est de plus en plus fréquente. Nous définissons cette activité en reprenant une définition existante :

“Collaborative Information Retrieval is an information access activity related to a specific problem solving activity that, implicitly or explicitly, involves human beings interacting with other human(s) directly and/or through texts as information sources in a work task related information seeking and retrieval process either in a specific workplace or in a more open community or environment” [HANSEN & JÄRVELIN in press].

Même si quelques études ont montré que la recherche collaborative d'informations sur le Web présentait certains inconvénients tels que l'inégale répartition du travail entre les usagers (par exemple, [LIPPONEN & al. in press] [NURMELA & al. 99]), plusieurs travaux ont montré que la recherche collaborative de pages Web améliore les performances des apprenants [BHARAT 00] [COCKBURN & MCKENZIE 01] [DIAMADIS & POLYZOS 04] [DUMAIS & al. 01] [LIPPONEN & al. 03] [TAUSCHER & GREENBERG 97] :

- le temps total nécessaire à la recherche d'informations sur Internet diminue significativement lorsque plusieurs individus recherchent ensemble des informations sur un même thème ;

- le volume des informations traitées et lues par ces individus augmente significativement ;

- l'organisation des informations trouvées sur Internet semble meilleure dans le sens où la fonctionnalité de « mise en favoris » est d'une part, beaucoup plus utilisée et d'autre part, plus structurante ;

- enfin, le nombre de pages re-visitées diminue significativement lorsque la recherche d'informations est collaborative. En effet, lorsqu'une recherche d'informations sur Internet est menée par un seul individu, le nombre de pages consultées ayant déjà été précédemment vues est considérable : selon les études, le pourcentage de pages re-visitées varient de 30% à 61%, avec des maximums de 92% pour certains individus ! Or, re-visiter des pages engendre forcément une perte de temps, temps que l'individu pourrait mettre à profit pour consulter des pages Web non vues.

Parmi les outils électroniques conçus pour aider les membres d'un groupe à réaliser une recherche d'informations collaborative sur le Web, notre choix s'est porté sur celui développé par [DIAMADIS & POLYZOS 04]. Le lecteur pourra se reporter à leur article pour avoir le descriptif complet de l'architecture logicielle.

Cet outil est censé aider des utilisateurs appartenant à un même groupe de « chercheurs d'informations » à localiser et sélectionner efficacement et rapidement des pages Web pertinentes par rapport à des thèmes prédéfinis. Cet outil, qui prend la forme d'une fenêtre supplémentaire située au-dessus du navigateur, offre un certain nombre d'informations relatives à la consultation antérieure (si consultation il y a eu) de la page Web affichée dans le navigateur située au-dessous. Très concrètement, si deux utilisateurs d'un même groupe travaillent, simultanément ou non, à rechercher des sites et pages Web, une fenêtre située au-dessus du navigateur fournit à chacun d'eux des indications telles que (a) l'adresse IP de sa machine, (b) le nom du groupe auquel appartient l'utilisateur, (c) les noms des autres membres du groupe, (d) le nom du membre de son groupe qui a déjà visité cette page Web, si première visite il y a eu, (e) les dates et heures de cette précédente visite, (f) l'évaluation qu'a faite le dernier visiteur de cette page, cette évaluation ayant été opérée en cochant l'une des deux cases suivantes : pertinente/non pertinente. Depuis le serveur, l'expérimentateur peut alors récupérer toutes les informations relatives aux comportements de navigation des utilisateurs de chaque groupe (adresses des pages Web visitées, durée des consultations, évaluation faite par les utilisateurs, etc.).

2. Méthode

2.1. Participants et conditions

Quarante étudiants de quatrième année de biologie (âge moyen = 23.4 ans) ont été recrutés pour participer à cette expérience. Dix groupes de 4 membres ont été créés, la constitution des groupes ayant été réalisée sur la base des choix personnels des étudiants. Cinq groupes devaient rechercher des pages Web pertinentes par rapport à trois thèmes prédéfinis sans le recours à un outil informatique leur fournissant des indications sur les pages déjà visitées par les autres membres du groupe (condition « sans indication »). Cinq autres groupes devaient rechercher des pages Web sur les mêmes thèmes en ayant la possibilité d'utiliser l'outil informatique leur fournissant des indications relatives aux pages Web affichées dans le navigateur (cf. point 1.).

2.2. Thèmes des recherches d'informations

Les participants devaient rechercher des pages Web pertinentes par rapport à trois thèmes relatifs à des concepts de la biologie, ces thèmes ayant été proposés après avis de cinq experts (enseignants) en biologie qui les ont distingués selon leur niveau de difficulté :

- Thème « facile » (facile) : « Quels sont les risques d'épidémies de grippe mortelle et quelles mesures sont prises pour prévenir ces épidémies ? » ;
- Thème moyennement facile (moyen) : « Toitures végétales : solutions et intérêts » ;

- Thème difficile (difficile) : « Les effets de l'anthropisation sur les herbiers de posidonies et leur écosystème ».

2.3. Procédure

Les dix groupes de quatre étudiants devaient réaliser trois séances de recherches d'informations. Le temps pour mener l'activité était totalement libre. Chaque groupe réalisait ses 3 recherches d'informations dans la même semaine (lors de trois sessions distinctes), dans une salle informatique équipée pour les besoins de l'expérience. En d'autres termes, tous les participants disposaient des mêmes outils informatiques et les recherches d'informations étaient opérées en simultané par les quatre membres de chaque groupe. La seule différence résidait dans le fait que certains groupes (condition « avec indication ») avaient la possibilité d'utiliser un outil informatique leur fournissant des indications relatives aux pages Web affichées dans le navigateur, telles que savoir si ces pages avaient déjà été visitées par un autre membre du groupe, l'évaluation faite sur ces pages, etc. (cf. point 1.). Tous les participants des groupes devant travailler avec l'outil donnant des indications sur les pages Web visitées ont bénéficié d'une séance d'entraînement durant laquelle les explications sur les différentes indications leur ont été données.

2.4. Indicateurs comportementaux

Pour chacun des trois thèmes imposés de recherche d'informations, cinq variables dépendantes ont été mesurées :

- le nombre de pages Web visitées par chaque groupe. Ont été considérées comme visitées et donc comme consultées (lues ?), toutes les pages Web ayant été affichées au moins pendant 5 secondes à l'écran ;
- le nombre de pages Web retenues (et donc jugées pertinentes) par chaque groupe ;
- la pertinence des pages Web retenues (évaluées par cinq experts en biologie, enseignants en biologie), sur une échelle de 0 « pas du tout pertinente » à 10 « très pertinente » ;
- le temps mis par chaque groupe pour effectuer la tâche (en minutes) ;
- le pourcentage de pages Web re-visitées par chaque groupe.

3. Résultats et discussion

La table 1 présente les résultats pour les cinq variables dépendantes mesurées, pour chaque condition (avec vs sans les indications sur les pages Web déjà visitées), pour chacun des trois thèmes de recherches d'informations (facile, moyen, et difficile).

Thème	Condition					
	Avec indication			Sans indication		
	facile	moyen	difficile	facile	moyen	difficile
Nbre pages Web visitées	21.6	33.8	38.6	23	34.4	37.8
Nbre pages Web retenues	12.2	8.6	12.2	10.8	8.8	10
Pertinence des pages retenues	6.2	6.3	6	7.02	6.4	5.6
Temps (en minutes)	33.4	44.8	59.2	28.4	43.2	89.6
Nbre re-visites (en pourcentage)	1	1	7.4	21.8	24.2	28.4

Table 1. Moyenne pour chacune des 5 variables dépendantes, pour les deux conditions et les trois thèmes

Les résultats montrent principalement que les indications sur les pages Web déjà visitées ont eu des impacts significatifs et bénéfiques sur (1) le temps mis pour réaliser l'activité uniquement lorsque la recherche d'informations est difficile, et (2) le pourcentage de re-visites de pages Web, surtout lorsque le thème est difficile.

En revanche, les indications sur les pages Web déjà visitées par les autres membres du groupe n'ont pas eu d'impact (3) sur la pertinence des pages Web trouvées par les groupes, (4) sur le nombre total de pages Web visitées, et (5) sur le nombre de pages Web retenues.

D'un point de vue pédagogique, notre étude a montré que la sélection collaborative de pages Web pertinentes pouvait être améliorée grâce à un dispositif technique relativement simple fournissant des informations relatives aux pages Web que consultait chaque membre du groupe. En effet, en connaissant les pages Web déjà visitées par un membre du groupe et en connaissant l'avis qu'il a émis par rapport à cette page (pertinente vs non pertinente), un étudiant peut consacrer son temps à l'examen d'autres pages Web, réduisant ainsi d'une part, le temps total de réalisation de la tâche et d'autre part, le nombre de re-visites inutiles.

Mais, plusieurs limitations méthodologiques empêchent toute généralisation de nos résultats. D'une part, notre étude était uniquement centrée sur la sélection des pages Web et ne renseigne donc pas sur la qualité d'un possible produit « final » (synthèse, résumé, rapport, etc.), véritable objectif de toute recherche d'informations. D'autre part, les participants dans les deux conditions de réalisation étaient différents. Enfin, l'homogénéité et/ou l'hétérogénéité inter et intra-groupes n'ont pas été mesurées.

4. Bibliographie

- [BHARAT 00] Bharat K., "SearchPad : explicit capture of search context to support web search", *WWW9, Proceedings of the Ninth International World Wide Web Conference*, Amsterdam, The Netherlands, May 15-19, 2000.
- [DIAMADIS & POLYZOS 04] Diamadis E.T., Polyzos G.C., "Efficient cooperative searching on the Web : system design and evaluation", *International Journal of Human-Computer Studies*, vol. 61, p. 699-724.
- [DUMAIS & al. 01] Dumais S., Cutrell E., Chen H., "Optimizing search by showing results in context", *CHI'01, Proceedings of the ACM Conference on Human-Computer Interaction*, Seattle, USA, March 31 – April 5, 2001, ACM Press, New-York, p. 277-284.
- [HANSEN & JÄRVELIN in press] Hansen P., Järvelin K., "Collaborative information retrieval in an information-intensive domain", *Information Processing and Management*.
- [LIPPONEN & al. 03] Lipponen L., Rahikainen M., Lallimo J., Hakkarainen K., "Patterns of participation and discourse in elementary students' computer-supported collaborative learning", *Learning and Instruction*, vol. 13, n°5, 2003, p. 487-509.
- [LIPPONEN & al. in press] Lipponen L., Rahikainen M., Hakkarainen K., Palonen T., "Effective participation and discourse through a computer network : investigating elementary students' computer-supported interaction", *Journal of Educational Computing Research*, in press.
- [NURMELA & al. 99] Nurmela E., Lehtinen E., Palonen T., "Evaluating CSCL log files by social network analysis", *Proceedings of CSCL '99: The third international conference on computer support for collaborative learning*, Erlbaum, Mahwah, NJ, 1999, p. 434-444.
- [TAUSCHER & GREENBERG 97] Tauscher L., Greenberg S., "How people revisit web pages: empirical findings and implications for the design of history systems", *International Journal of Human-Computer Studies*, Vol. 47, n°1, 1997, p. 97-137.