

Evaluation d'un système Web Socio Sémantique en recherche d'information

Jean-Pierre Cahier, Nadia Gauducheau, L'Hédi Zaher, Manuel Zacklad

ISTIT Fre CNRS 2732, Equipe Tech-CICO, Université de Technologie de Troyes, 12 rue Marie Curie, BP 2060, 10010 Troyes cedex {prénom.nom}@utt.fr

Résumé

Un dispositif expérimental a permis de mesurer certains points de comparaison entre les performances de sujets utilisant **i)** une aide à la navigation basé sur les principes du Web Socio Sémantique et recourant à une approche d'IC (outil Agoræ), **ii)** un moteur de recherche et **iii)** la navigation dans une structure simple de mots clés.

Mots clés : évaluation, ontologie sémiotique, recherche d'information, Web socio sémantique.

1 Introduction

La recherche présentée est un travail pluridisciplinaire articulant les préoccupations de chercheurs en IC avec les problématiques en psychologie cognitive sur les aides à la RI. Il s'agit d'évaluer le système Agoræ, dispositif d'aide à la navigation basé sur les ontologies sémiotiques [9] et les principes du Web Socio Sémantique [2], et de comparer différents dispositifs de RI pour suggérer des critères de conception d'aide à la navigation.

Appliquées à la caractérisation de ressources documentaires, les ontologies sémiotiques offrent une modélisation du domaine prenant en compte les besoins en gestion d'information (recherche, classement...) d'une communauté d'utilisateurs. Basées sur une approche d'IC visant à capitaliser l'expérience d'experts du domaine considéré [1] [2], à la différence des thésaurus, elles recourent à des points de vue multiples mettant en relief différentes facettes des ressources considérées et exploitent des relations plus riches que ces derniers entre les expressions caractérisant les ressources (les « thèmes »).

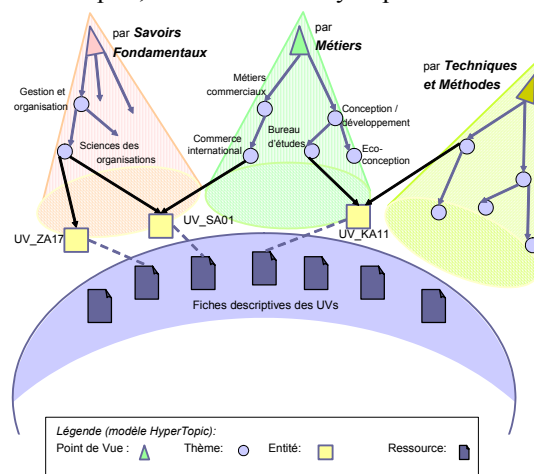
2 Environnements et modèles

La RI s'appuie sur deux classes d'environnements. Dans un *dispositif par requête* (par exemple, moteur de recherche), le sujet interroge sur les « documents » contenus dans sa base. La requête est comparée avec les représentations des documents dans la collection [4]. Pour être utilisés pleinement, ces outils impliquent des connaissances sur la façon dont le système opère. Des travaux expérimentent des systèmes aidant l'utilisateur dans la formulation et reformulation de la requête, ou s'intéressent à l'effet des techniques de visualisation des réponses [3].

Dans un *système par navigation*, la structuration des informations guide l'utilisateur au long de sa recherche. Son utilisation ne nécessite pas de connaissances sur le fonctionnement informatique. En revanche, il conduit souvent à un phénomène de désorientation. Une solution pour palier à ce problème est de fournir à l'utilisateur une carte du site, sous forme de liste (table des matières, index) ou sous forme graphique. L'approche du Web Socio Sémantique implique la conception d'un dispositif de navigation utilisant des ontologies sémiotiques, offrant de multiples points de vue sur les ressources et pouvant évoluer selon les besoins des utilisateurs et les évolutions du domaine (FIG. 1).

FIG. 1 : Exemple simplifié extrait de l'ontologie sémiotique utilisée, concernant les UV d'une Université

D'autre part, nous avons analysé pour cette étude les



différents modèles cognitifs associés à la RI. *Activité exploratoire*, la RI est aussi une *activité de résolution de problème*. Ainsi la RI dans un hypertexte peut être analysée [7] comme un cycle de trois processus, l'évaluation, la sélection et le traitement. Trois types de connaissances sont impliquées [8] dans la navigation : les connaissances du domaine, les connaissances métadocumentaires et les connaissances opératoires sur l'utilisation des outils. Les ontologies sémiotiques apportent notamment des connaissances métadocumentaires. Enfin, la RI relève des *problèmes de « conception »* reposant sur la planification [6], la

RI consistant en un changement de représentation du problème prenant en compte les contraintes.

Finalement pour nous la RI est une activité *d'exploration* où la représentation se construit au fil de la recherche. C'est une activité *cyclique* qui enrichit progressivement la représentation du but de l'utilisateur. Cette activité *implique des changements de systèmes de représentation* incluant des plans déclaratifs du domaine.

3 Expérimentation et résultats

Deux questions ont été dégagées de cette analyse pour guider l'évaluation du dispositif de navigation préconisé dans le cadre d'une approche « Web Socio Sémantique ». D'une part, un système de navigation permet-il d'être plus efficace qu'un système par requête ? D'autre part une carte du domaine s'appuyant sur les ontologies sémiotiques est-elle un système d'aide à la navigation efficace ?

Trois groupes d'étudiants ont réalisé une tâche de RI en temps limité (30 minutes) dans un catalogue informatisé sur Internet., respectivement avec un moteur de recherche, un système par mots-clés et l'outil Agoræ. Nous avons analysé l'effet du système de RI sur leurs performances. Les utilisateurs du moteur ont proposé davantage de réponses ($m=11.3$) que les utilisateurs des sites par navigation ($F(1,30)=13.13$, $p<0.002$), c'est-à-dire avec Agoræ ($m=8.1$) et avec le site par mots-clés ($m=8.2$). Il semble donc que la RI avec aide à la navigation soit plus coûteuse en temps. Le nombre de réponses correctes est similaire dans les trois systèmes ($F(2,30)=1.21$, ns), entre 3.5 et 4.3 sur 12 en moyenne (cf. tableau 2). Le taux de réponses correctes correspond au nombre de réponses correctes divisé par nombre total de réponses fournies par le sujet. Cela permet de déterminer si le sujet propose des réponses adéquates ou au contraire s'il fait souvent des erreurs. Le taux de réponses correctes avec Agoræ est supérieur (53%) par rapport aux utilisateurs du moteur de recherche (35%) ($F(1,30)=4.42$, $p<0.05$).

4 Discussion et perspectives

La comparaison entre systèmes différents ne permet pas toujours d'observer des différences notables et interprétables [5] [6]. Cependant la meilleure fiabilité des réponses obtenues avec Agoræ nous semble d'autant plus convaincante que le moteur de recherche s'appuie ici sur une base de documents de petite taille.

A ce stade, les résultats de cette étude sont encourageants. Ils montrent la supériorité à certains égards des performances obtenues avec l'ontologie sémiotique par rapport au moteur. Néanmoins, l'évaluation d'Agoræ doit être complétée par d'autres expériences. Ainsi, il est envisagé de proposer des tâches permettant de repérer les connaissances que les participants ont construites après utilisation du système. Cela permettrait de montrer que l'efficacité des ontologies sémiotiques est bien liée à un enrichissement de la représentation de la connaissance de l'utilisateur. Cette question est d'autant plus importante

que les enjeux de la conception de systèmes de RI sont non seulement de trouver l'information « recherchée » mais aussi d'apprendre dans un domaine de connaissance [4].

D'autres expérimentations sont en préparation pour compléter l'évaluation du système Agoræ, cette fois pour son apport à la *conception* d'ontologie sémiotique. La qualité d'une ontologie sémiotique repose en amont sur un investissement permanent en IC, l'ontologie devant rester en phase avec les acteurs et leur environnement complexe et changeant. Cette qualité suppose une approche à la fois de travail coopératif et d'IC [1] pour déterminer entre experts du domaine les points de vue et les autres éléments nécessaires pour structurer l'ontologie sémiotique.

Références

- [1] Cahier J.-P. & Zacklad M. (2002), Approche cognitive de la co-construction d'un catalogue multi-points de vue et multi-experts. *Actes de la Conférence Extraction et Gestion des Connaissances (EGC)*, Montpellier 2002, p.221-226, Hermes
- [2] Cahier, J.-P., Zacklad, M., Monceaux, A., Une application du Web socio sémantique à la définition d'un annuaire métier en ingénierie, *Actes des journées Ingénierie des Connaissances*, mai 2004, Lyon.
- [3] Ciaccia, A., Martins, D., Recherche d'informations sur le Web : Etude de l'influence de facteurs liés à l'interface, à l'utilisateur et à la tâche, *Revue d'intelligence artificielle*, 19, 1/2, 2005, p.159-177.
- [4] Dunlop, M. D., Johnson, C.W., Reid, Exploring the layers of information retrieval evaluation, *Interacting with computers*, 10, 1998, p. 225-236.
- [5] Kahn, J. (2000). Dynamique de l'interaction Personne-Système de recherche d'Information, Thèse de doctorat, CNAM, Paris, 2000.
- [6] Kolmayer, E. (1998). Démarche d'interrogation documentaire et navigation. In J.F. Rouet et B. De La Pas-sardièrre (eds.), *Quatrième colloque Hypermédias et apprentissage*, Poitiers, 1998, p.121, 134.
- [7] Tricot, A., Drot-Delange, B., El Boussarghini, R., Foucault, B. Quels savoir-faire les utilisateurs réguliers du web acquièrent-ils? *Revue d'Intelligence Artificielle*, vol 14, n°1-2, 2000, p. 93-112.
- [8] Tricot, A., Rouet, J.-F., Activités de navigation dans les systèmes d'information. In J.M. Hoc & F. Darses (eds.), *Psychologie ergonomique : tendances actuelles*, pp. 71-95, 2004.
- [9] Zacklad, M., Introduction aux ontologies sémiotiques dans le Web Socio Sémantique, *conférence ingénierie des connaissances 2005*, Sophia Antipolis, 31 mai 2005