

Recherche d'information dans le Web socio-sémantique

L'Hédi Zaher, Jean-Pierre Cahier et Manuel Zacklad

Équipe Tech-CICO, ISTIT – CNRS FRE 2732, Université de Technologie de Troyes, 12 rue Marie Curie, BP 2060 - 10010 Troyes cedex, France.

{hedi.zaher, jean-pierre.cahier, manuel.zacklad}@utt.fr

Résumé

Nous avons mesuré expérimentalement les performances de sujets utilisant un dispositif d'aide à la navigation : le système Agoræ basé sur les principes du Web Socio Sémantique et recourant à une approche d'ingénierie des connaissances. Nous avons comparé d'une part ces performances avec celles obtenues en utilisant un moteur de recherche. Nous les avons comparées d'autre part avec celles obtenues par navigation à partir d'une structure simple de mots clés, de type thésaurus, ne faisant pas appel à l'ingénierie des connaissances. Pour ces comparaisons trois groupes d'étudiants débutants se sont vus proposer dans le même temps les mêmes consignes, leur demandant de trouver des unités de valeur (UV) dans le catalogue électronique des enseignements d'une université, équipé par ces trois dispositifs concurrents munis d'un même système de trace. Les résultats nous permettent de conclure à la supériorité à certains égards des performances obtenues avec Agoræ, par rapport au moteur et à la structure simple de mot-clés.

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

1 Introduction

La conception de systèmes de Recherche d'Information (RI) performants est un enjeu majeur des Sciences de l'Information et de l'Ingénierie des Connaissances. Les conséquences sociétales et économiques sont multiples dans les domaines de l'enseignement et de la culture comme dans celui de l'économie. Ces dernières années ont vu l'émergence du courant du Web Sémantique présenté comme un des moyens de rendre plus accessible les contenus disponibles sur le Web à ses utilisateurs par le biais d'une modélisation préalable de la structure sémantique des contenus.

Nous considérons que les orientations actuelles du Web Sémantique le rendent plus adapté à la recherche automatique d'informations stables et peu évolutives par le biais d'agents logiciels, qu'à l'assistance à la RI par des utilisateurs humains au sein de contenus complexes et évolutifs [CCZ02]. Ces analyses nous ont mené à contraster un Web Sémantique Logique ¹ à un Web Socio Sémantique qui exploite des ontologies sémiotiques [ZAC05]. Les ontologies sémiotiques sont des spécifications au sens de Dewey et Bentley [DBE89], c'est-à-dire un processus de dénomination des ingrédients d'une situation problème permettant de progresser vers sa résolution. Appliquées à la caractérisation de ressources documentaires, les ontologies sémiotiques permettent d'offrir une modélisation du domaine prenant en compte les besoins en gestion d'information (recherche, classement...) d'une communauté d'utilisateurs [CAH05]. Basée sur une approche d'ingénierie des connaissances visant à capitaliser l'expérience d'experts du domaine considéré [CZA02], 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 »).

¹ Que nous avons initialement appelé « Web Computationnellement Sémantique » [ZCP03].

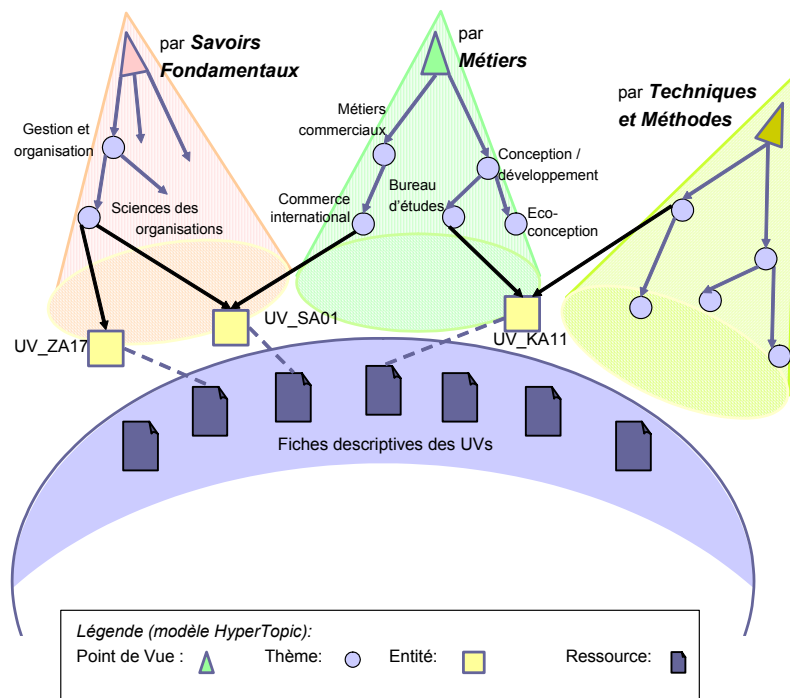


Fig. 1 – Extrait de « l'ontologie sémiotique » utilisée dans cette étude

Au sein de l'équipe Tech-CICO, qui promeut l'approche du Web Socio Sémantique [ZAC05], le formalisme de représentation des connaissances HyperTopic permettant d'exprimer des ontologies sémiotiques [ZAC05] [CAH05] a été défini. Il sert lui-même de base à l'outil Agoræ qui permet de naviguer dans ces ontologies, d'y rattacher des ressources et de les manipuler sous la forme d'une carte de thèmes multi-points de vue.

Nos objectifs ont été de comparer expérimentalement les performances de sujets utilisant un dispositif de RI basé sur les principes du Web Socio Sémantique à celles obtenues d'une part en utilisant un moteur de recherche, et d'autre part, par navigation à partir d'une structure d'aide simple à base d'index de mots-clés.

2 Activité cognitive et recherche d'information

Nous appelons Recherche d'Information (RI) l'ensemble des activités, des processus, des représentations, des inférences et des transformations cognitives qu'élabore et utilise un individu pour trouver des informations utiles et pertinentes par rapport à une tâche qu'il se donne. Nous nous rapprochons ainsi du sens donné par Rouet et Tricot [RTR98] par rapport à l'acceptation du terme qui privilégie l'étude des algorithmes de recherche automatique dans des bases de ressources numériques (bases de données, bases documentaires, etc.).

Dans cette partie nous présenterons une première revue de question des recherches sur les activités de RI conduite d'un point de vue cognitif multidisciplinaire.

- La RI est une activité exploratoire. En « sciences de l'information et de la communication » son étude s'appuie sur l'observation des conduites des sujets dans des situations diverses de RI. Les

auteurs mettent l'accent sur la définition du besoin en information et du problème [MAR95]. Elle joue un rôle fondamental dans toutes les étapes de la RI et dirige la recherche. Le caractère flexible et mouvant de cette définition rend le processus de RI non linéaire. [BAT89] insiste sur la flexibilité du besoin d'information tout au long de la recherche. Elle caractérise la RI comme une activité exploratoire, de type « cueillette de baie », passant par la récolte de « bribes » d'informations récupérées au fur et à mesure des requêtes. Le résultat des interactions avec le système guide l'activité cognitive des sujets.

- La RI est une activité de résolution de problème. [RTR98] proposent un modèle de la RI dans un hypertexte analysée comme un processus de résolution de problème. La RI correspond à un cycle de trois processus, l'évaluation, la sélection et le traitement. Le processus d'évaluation a pour fonction d'identifier les informations qui manquent pour effectuer la tâche. Le processus de sélection correspond à une prise de décision en référence aux critères élaborés dans la phase évaluative. Le processus de traitement met en jeu principalement une activité de lecture-compréhension de texte. Les cycles se succèdent jusqu'à l'atteinte du but. Des mécanismes de gestion cognitive de l'activité accompagnent ces processus (planification, contrôle, régulation). Ce type de modèle permet de rendre compte de l'importance de l'expertise des individus, sur le domaine impliqué dans la RI et sur l'utilisation des outils [TDE00].
- La RI relève des problèmes de « conception » reposant sur la planification. [KOL98] tente d'intégrer les contraintes liées au système dans son analyse de la tâche de recherche d'information dans les bases documentaires. Elle la traite comme un problème de conception. La RI consisterait en un changement de représentation du problème prenant en compte les contraintes. Le changement de représentation se traduit par le passage d'un but flou (besoin d'information) à un but opérationnel (plan d'action sur le système). L'activité du sujet consiste à élaborer des plans déclaratifs (explicitation du but) et procéduraux (mise en œuvre de l'atteinte du but). Les contraintes sont liées au but (quel domaine...) mais également au système d'information (système d'interrogation et fonctionnalités). Un système hypertexte propose une représentation du domaine à l'utilisateur lui permettant d'élaborer des plans déclaratifs contribuant à la focalisation progressive.

De cette première revue nous retenons trois enseignements qui nous semblent essentiels pour la construction des ontologies sémiotiques, la RI étant à la fois :

- une activité d'exploration dans laquelle la représentation du but se construit au fur et à mesure de la recherche ;
- une activité cyclique qui enrichit progressivement la représentation de l'utilisateur ;
- une activité impliquant des changements de systèmes de représentation impliquant notamment des plans déclaratifs du domaine.

3 Recherche d'information et environnement informatiques

Dans la revue proposée au §2, les approches retenues ne tenaient pas systématiquement compte de la spécificité des environnements informatiques de RI. Pour compléter ce point, il faut noter un certain nombre de travaux qui se sont intéressés spécifiquement aux deux grandes classes d'environnements de RI :

- Dans un dispositif par requête, comme un moteur de recherche, la requête est comparée avec les représentations des documents dans la collection [DJR98]. Le système, en réponse à la requête, produit une liste de documents (ou sites...) qui correspond aux documents de la collection compatibles avec la requête (contenant le mot ou des combinaisons de mots, etc.). Pour être utilisés pleinement ces outils impliquent des connaissances sur la façon dont le système opère. Quand une requête n'apporte pas de résultat intéressant, l'utilisateur doit diagnostiquer pourquoi la réponse du système est inattendue et résoudre le problème en proposant une nouvelle requête. Ce système permet de n'avoir accès qu'à une seule partie des ressources (des documents de la collection), ceux correspondant au résultat de la requête. Par conséquent, l'utilisateur doit évaluer l'écart possible entre cette « fenêtre » sur la collection que renvoie le système et son besoin d'information [CHI97]. C'est pourquoi un des enjeux pour ce type d'outils est de

- proposer des feedbacks (retours) pertinents permettant d'apprendre le fonctionnement de l'outil et d'interpréter cette « fenêtre ». Ainsi, certains travaux expérimentent des systèmes aidant l'utilisateur dans la formulation et reformulation de la requête, que ce soit par la sélection de champs préalables [MVR97] ou par la réalisation d'inférences à partir des requêtes précédentes (Beaulieu & Jones, 1998). D'autres travaux s'intéressent à l'effet des techniques de visualisation des réponses [CMA2005].
- Dans les dispositifs par exploration/navigation (ou encore hypertexte) [CHI97], la collection de documents est accessible en navigant de lien en lien. Les liens peuvent se situer entre les documents (système hypertexte classique) et/ou dans une structure décrivant l'organisation des informations (carte conceptuelle, index, etc.). La variété de ces outils est assez large allant d'index peu structurés à des index fortement structurés comme les deux types de systèmes de navigation que nous allons utiliser, respectivement à base de listes de mots-clés et à base d'ontologies sémiotiques. La structuration des informations guide l'utilisateur tout le long de sa recherche (choix d'un lien ou d'un autre). L'utilisation d'un tel outil ne nécessite pas de connaissances sur le fonctionnement informatique. En revanche, la navigation conduit souvent à une désorientation des utilisateurs caractérisée par la consultation de nombreux documents et une baisse des performances dans le temps [MAR95] [MDS98]. L'utilisateur n'obtient pas une vue d'ensemble de la structure des informations et éprouve des difficultés à établir un but et sa planification [KHI95].

Pour être utilisés pleinement les systèmes par requête requièrent de l'utilisateur une bonne connaissance des modalités de fonctionnement du système et en particulier de la manière de formuler les requêtes pour parvenir à des résultats pertinents. De leur côté, les systèmes par navigation sont susceptibles d'entraîner une forme de désorientation des utilisateurs. Pour y pallier, il est bénéfique de proposer des aides à la navigation sous la forme d'une structuration des informations utiles à la navigation.

Cette conclusion provisoire nous semble montrer que dans le cas où les connaissances du domaine sont complexes (ou les utilisateurs novices), cas dans lesquels les systèmes par navigation semblent offrir de meilleures performances, les ontologies sémiotiques sont susceptibles de fournir un guidage palliant aux risques de désorientation cognitive des utilisateurs. Ces conclusions vont nous permettre de formuler notre hypothèse de recherche.

4 Hypothèse expérimentale

Les analyses des stratégies de résolution de problèmes dans les activités de RI montrent que la construction de la représentation du but est au cœur du processus de recherche. Plusieurs facteurs influencent la représentation construite par l'utilisateur, par exemple ses connaissances préalables, le type de tâche et la représentation du domaine fournie par le système [KOL98]. Ce dernier point nous semble important car il différencie un moteur de recherche d'un dispositif de type navigation, et parmi les différents systèmes d'aide à la navigation il avantage ceux qui représentent mieux les connaissances du domaine.

Le moteur de recherche ne fournit pas de représentation du domaine à l'utilisateur et celui-ci doit, par les réponses du système et son expertise (domaine/outil), construire cette représentation. En revanche, avec les dispositifs par navigation qui offrent ce type de représentation, le risque de désorientation lié à ce type de dispositif diminue lorsque la structuration des informations est accessible à l'utilisateur.

Nous faisons donc l'hypothèse selon laquelle un dispositif de recherche par navigation exploitant des ontologies sémiotiques facilitera la construction de la représentation du but à atteindre pour des utilisateurs n'étant pas experts dans le domaine. Il devrait leur permettre d'atteindre de meilleures performances qu'un moteur de recherche et qu'un dispositif d'aide par simple structure de mots-clés.

5 Expérimentation

5.1 Procédure, population et tâche

Trente-trois étudiants de l'Université de Technologie de Troyes ont participé à cette étude. En 1^{ère} ou 3^{ème} année de formation d'ingénieur, tous arrivent pour la première année dans cette Université. De langue maternelle française, ils sont assez familiers d'Internet et de la recherche d'information.

Les participants réalisent une tâche de RI en temps limité (30 minutes) dans un catalogue informatisé sur Internet. Celui-ci rassemble un ensemble de fiches descriptives d'UV en sciences humaines/culture et systèmes d'informations. La tâche consiste à sélectionner un enseignement en référence à une consigne. Avant d'effectuer la tâche, le site est présenté aux étudiants et une consigne d'entraînement est proposée. Douze consignes sont présentées dans un ordre aléatoire alternant les domaines. Trois groupes de 11 étudiants ont réalisé la tâche, respectivement avec un moteur de recherche, un système par mots-clés et le système Agoræ. Les étudiants n'ont encore suivi aucune des UV décrites.

La tâche de RI consiste à partir d'une consigne, à choisir une UV dans le catalogue. La consigne présente un stage pour lequel des compétences ou des connaissances spécifiques sont requises. Chaque UV est décrite par une fiche (réalisée par l'enseignant). Le code des UV a été complètement modifié par rapport aux codes usuels de l'Université, pour garantir des conditions d'égale ignorance chez tous les participants. Une seule UV permet à chaque fois de répondre à la demande. La tâche a été pré-testée auprès de 11 étudiants représentatifs de la population cible afin de déterminer leur compréhension des consignes et le niveau de difficulté.

Exemple de consigne: « Dans le cadre d'un prochain stage d'ingénieur, vous souhaitez renforcer vos compétences en choisissant une UV parmi celles proposées à l'UTT. Dans ce stage, vous devrez organiser un système de télé-enseignement recourant à des didacticiels en ligne. Quelle UV choisir pour vous préparer ? »

Quatre-vingt quatorze UV - parmi les 300 proposées à l'université - ont été sélectionnées pour construire le catalogue, 53 de sciences humaines/culture et 41 de systèmes d'information. Ces domaines ont été sélectionnés en raison de l'expertise nécessaire dont nous disposions pour construire les sites et consignes.

5.2 Dispositifs

« Carte de thèmes » Agoræ

Le système « Agoræ » permet de naviguer dans une carte de thèmes organisés en points de vue, représentant elle-même une ontologie sémiotique suivant le modèle HyperTopic selon lequel d'autres applications de ce type ont été déjà réalisées, [CZM04]. Il ne s'agit pas d'un hypertexte « classique » au sens où l'utilisateur ne navigue pas de document en document (liens entre les documents) mais dans une structure (description du domaine) permettant d'accéder aux documents. Les liens se situent donc au niveau de la structure. Le « Point de Vue » est un descripteur qui correspond à un angle sous lequel est envisagé l'UV (cf. Figure 1). Un expert a réalisé la carte qui comporte environ 400 Thèmes. Il s'agit d'un enseignant chercheur de l'université expérimenté en ingénierie des connaissances.

Chacune des 94 UV est classée selon plusieurs thèmes appartenant à tous les points de vue. Dans l'expérience, cinq points de vue sont proposés aux étudiants participants : par savoirs fondamentaux ; par méthodes, techniques et outils ; par projet ; par métier et fonction ; par secteur d'activités.

Consigne 1/12 : Dans le cadre d'un prochain stage, vous souhaitez renforcer vos compétences en choisissant une UV parmi celles proposées à l'UTT. Il s'agit d'un stage dans lequel vous participeriez à l'élaboration d'une formation au respect de l'environnement à destination d'entreprises. Une partie du stage porterait sur la réalisation de fiches concernant l'évolution des questions du respect de l'environnement ces 50 dernières années. Quelle UV pourrait vous aider dans cette tâche ?

1- SAVOIRS FONDAMENTAUX (6) 1- Gestion, organisation et économie (20) 2- Informatique, réseaux et télécommunications (9) 3- Autres Sciences de l'ingénieur (5) 4- Mathématiques et Logique (5) 5- Autres Sciences sociales et Humanités (16) 6- Sciences de la Nature et de la Vie (3)	2- METHODES, TECHNIQUES et OUTILS (6) 1- Méthodes de l'ingénieur (12) 2- Techniques managériales et commerciales (6) 3- Techniques financières (3) 4- Outils et techniques industrielles (7) 5- Outils et techniques informatiques (4) 6- Méthodes adressant les aspects humains et organisationnels (21)
3-METIER et FONCTION (14) 01- Management général de l'entreprise (1) 02- Métiers commerciaux (3) 03- Conception et développement du produit (10) 04- Logistique, Gestion de Production et Productique (4) 05- Métiers de l'Informatique et de l'Organisation (10) 06- Ressources Humaines (6) 07- Services financiers 08- Communications d'entreprise interne et externe (3) 09- Service juridique 10- Service documentation 11-Fonction d'animation de groupes ou de projets 12- Formation, accompagnement du changement 13- Métiers artistiques	4-PROJET (3) 1- Projet lié à l'entreprise (20) 2-Projet personnel - Savoir-être (11) 3- Projet social (4)

Fig. 2 – Interface de navigation dans la carte de thèmes (« Agoræ »)

Moteur de recherche

Le moteur de recherche utilisé, PHPDig est classique, de type recherche dans tous les mots « pleins » existants dans les documents. Pour accéder au descriptif d'une UV, l'étudiant doit émettre une requête dans un cartouche. Le moteur restitue une liste d'UV. L'étudiant peut alors consulter le contenu des UV (en cliquant sur la réponse). Pour ne pas défavoriser le moteur de recherche, les consignes contiennent au moins un mot qui, pris tel quel pour effectuer une requête, produit un résultat contenant la fiche de l'UV cible. Ce mot doit également aboutir à au moins 3 autres résultats (autres UV « non cibles »).

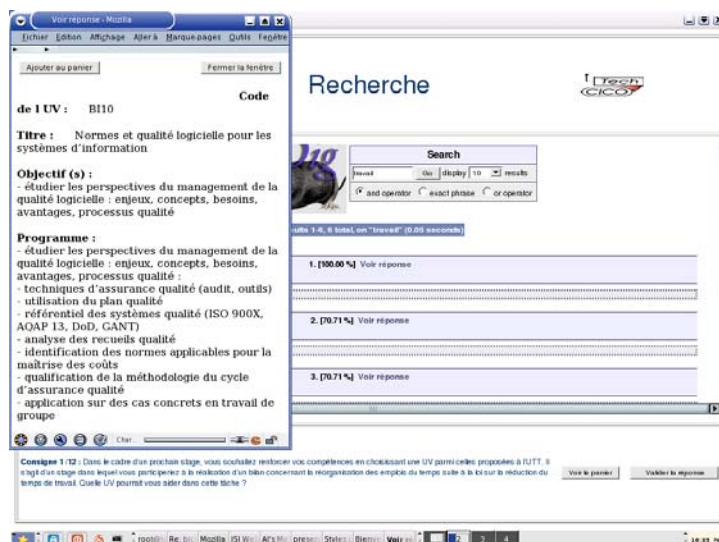


Fig. 3 – Interface recherche par moteur

Structure simple de mots-clés

Dans cette approche, chaque UV a été indexée à partir de sa fiche descriptive selon 7 mots-clés. Compte tenu des recouvrements (un mot-clé indexe en moyenne 1,6 UV), la liste résultant de leur réunion représente 407 mots clés présentés alphabétiquement. Les essais préliminaires de ce site à partir d'une liste alphabétique de 407 mots ayant montré des performances médiocres très défavorables, nous avons amélioré la structure de présentation de ces mots-clés en les divisant en cinq rubriques selon une arborescence simple (« entreprise », « sciences », « techniques », « société », « culture »), de manière à rendre le système plus compétitif. Les mots-clés ont été triés selon ces quatre subdivisions, et ensuite sont présentés par ordre alphabétique. Ce système contraste avec le système Agoræ, qui se base sur une carte de thèmes possédant des subdivisions plus fines et des points de vue multiples considérés par l'expert comme pertinents pour la recherche d'UV. Néanmoins, le taux de rattachement d'une UV est voisin dans les deux systèmes (1,6 UV indexée par un thème ou un mot-clé donné).

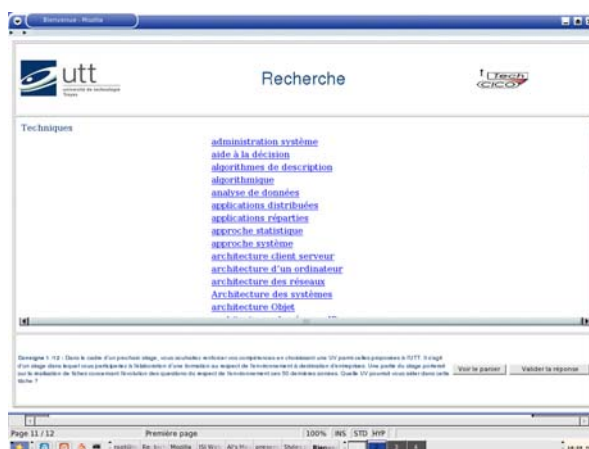


Fig. 4 – Interface de navigation par mots-clés

5 Résultats

Nous avons analysé l'effet du système de RI sur les performances des utilisateurs confrontés à la tâche de RI. Les performances sont exprimées par le nombre de réponses fournies par les utilisateurs dans les 30 minutes qui leur ont été imparties, le nombre de réponses correctes et le taux de réussite (cf. tableau 1). Ces trois mesures ont fait l'objet d'une analyse de variance avec comme facteur inter-sujets le système de RI (Agoræ, moteur, mots-clés).

Il n'y a pas d'effet du système de RI sur le nombre de réponses correctes ($F(2,30)=1.21$, ns). Il y a tout autant de réponses correctes avec les trois systèmes. Néanmoins, les résultats montrent que les utilisateurs du moteur de recherche 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 ($m=8.2$). Le nombre de réponses fournies est similaire pour les deux systèmes par navigation ($F(1,30)=3$, ns). Il semble donc que la RI avec aide à la navigation soit plus coûteuse en temps. Par contre, la fiabilité des réponses avec Agoræ semble supérieure par rapport aux deux autres systèmes. La fiabilité des réponses est évaluée par le taux de réussite à la tâche. Cela 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 est plus élevé pour les utilisateurs de la carte de thèmes « Agoræ » (53%) par

rapport aux utilisateurs du site par mots-clés (44%) et du moteur de recherche (35%). La différence est statistiquement significative pour le moteur de recherche ($F(1,30)=4.42, p<0.05$) mais pas pour le site par mots-clés ($F(1,30)=1, ns$).

	Nombre de réponses fournies		Nombre de réponses correctes		Taux de réussite	
	Moyenne	Ecart type	Moyenne	Ecart type	Moyenne	Ecart type
Agoræ	8,1	2,2	4,3	2	0,53	0,2
Moteur	11,3	1	4	1,3	0,35	0,1
Mots-clés	8,2	3,3	3,5	1,8	0,44	0,24

Tab. 1 – Performances

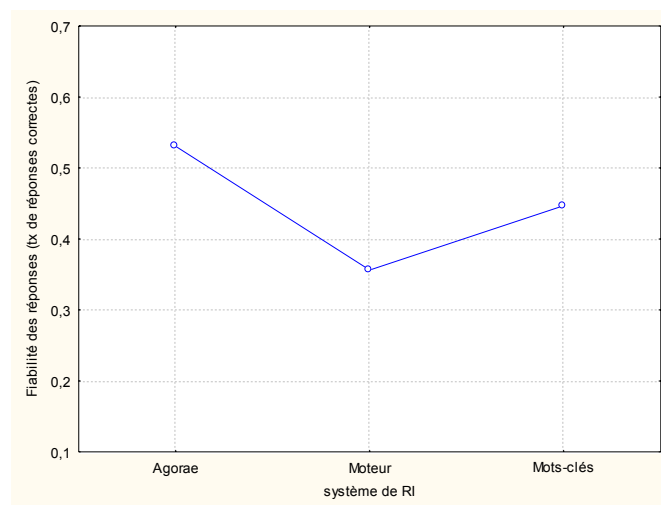


Fig. 5 – Taux de réponses correctes avec les 3 systèmes comparés

Un certain nombre d'autres variables ont été acquises par questionnaire, ou tracées durant l'expérience (notamment les thèmes, mot-clés ou expressions de requête utilisés pour accéder aux UV intermédiaires et finales), afin de suivre l'activité des participants. Par exemple, chaque participant dispose sur les trois systèmes d'un panier (« caddie ») pour stocker plusieurs UV et consulter éventuellement à nouveau le texte de la fiche avant de prendre sa décision finale. Parmi les variables ainsi relevées qui se sont avérées sensibles au type de site, le nombre moyen d'UV ajoutées au panier a été de 1,4 pour Agoræ, de 1,33 pour le moteur, et de 1,7 pour les mots-clés.

6 Discussion et perspectives

Les résultats montrent que du point de vue de la qualité des réponses le dispositif de navigation guidé par des ontologies sémiotiques offre des performances proches, et à certains égards meilleures, que les deux systèmes concurrents étudiés. En particulier, le taux de réponses correctes avec le système « Agoræ » basé sur les principes du Web Socio sémantique est légèrement supérieur au taux mesuré avec le système basé sur le moteur de recherche. Certes, la solution du moteur de recherche montre qu'elle permet de répondre aux consignes plus rapidement, en nombre de réponses fournies, mais nous avons pu établir que cet avantage est seulement apparent, à cause du taux plus élevé de réponses incorrectes avec le moteur.

Il faut souligner que la comparaison entre systèmes différents ne permet pas toujours d'observer des différences notables et interprétables [KAH00] [KOL98]. Cela est lié à la complexité des situations étudiées et à la multiplicité des facteurs pouvant intervenir. La meilleure fiabilité des réponses obtenues avec nous semble d'autant plus convaincante que le moteur de recherche s'appuie sur une base de données de petite taille. La plupart des mots de la consigne ramènent moins de 5 résultats dont la fiche correspondant à l'UV correcte. Dans notre cas, les réponses du moteur sont donc bien moins bruitées qu'elles ne le sont habituellement.

La meilleure fiabilité des réponses obtenues avec Agoræ par rapport au moteur semble confirmer notre hypothèse selon laquelle la cartographie du domaine offerte par l'ontologie aide l'utilisateur à se représenter le but à atteindre. D'autre part, le taux de réponses correctes apparaît légèrement meilleur pour l'approche « multi points de vue » Agoræ que pour l'aide à la navigation par listes de mots-clés, présentant une sémantique « mono point de vue » basé sur un index faiblement structuré.

Néanmoins, outre une reconduction du même protocole (que nous prévoyions d'effectuer avec des effectifs plus nombreux lors de l'arrivée de futures promotions d'étudiants), d'autres expériences sont nécessaires pour confirmer ces premiers résultats et affermir ces facteurs explicatifs. Il serait ainsi intéressant de proposer des tâches permettant de repérer les connaissances que les participants ont construites après utilisation du système. Nous pourrions ainsi montrer que l'efficacité d'un index fortement « structuré », correspondant aux ontologies sémiotiques, est bien liée à un enrichissement de la représentation de la connaissance du domaine. 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 également d'apprendre dans un domaine de connaissance [DJR98].

Si les présentes performances d'une ontologie sémiotique en appui de la RI peuvent sembler encourageantes, elles sont tributaires de la qualité de cette ontologie sémiotique. C'est pourquoi d'autres expérimentations sont également en préparation pour compléter l'évaluation du système Agoræ, mais cette fois du point de vue son apport à la *conception* d'ontologie sémiotique. Nous visons particulièrement l'amélioration de performances que cet outil apporte dans la conception d'ontologie sémiotique au sein de collectifs d'experts distants ou de communautés. La qualité d'une ontologie sémiotique repose en amont sur un investissement permanent en ingénierie de connaissances, l'ontologie devant rester en phase avec ces acteurs et leur environnement complexe et changeant. Cette qualité suppose une approche à la fois de travail coopératif et d'IC [CZA02] pour déterminer entre experts du domaine les points de vue et les autres éléments nécessaires pour structurer la carte de thèmes. C'est pourquoi nous chercherons aussi à évaluer l'apport en performance du système Agoræ dans une tâche de construction collective d'ontologie sémiotique, en le comparant aux moyens traditionnels ou à d'autres outils collecticiels utilisables dans cette tâche.

Références

- [BHL01] Berners-Lee T., Handler J., Lasilla O., "The Semantic Web", *Scientific American*, pages 34-43, mai 2001.
- [BAT89] Bates M., "The design of browsing and berry picking techniques for the on-line interface", *On-line Review*, 13, 5, 1989, p. 407-424.
- [BJO98] Beaulieu M., Jones S., "Interactive searching and interface issues on the Okapi best match probabilistic retrieval system", *Interacting with computers*, 10, 1998, p. 237-248.
- [CAH05] Cahier J.-P., *Ontologies sémiotiques pour le Web socio sémantique: Etude de la gestion coopérative des connaissances avec des cartes hypertopiques*, Thèse en informatique du laboratoire ISTIT/Tech-CICO, sous la direction de M. Zacklad, , soutenue à l'Université de Technologie de Troyes le 14 décembre 2005
- [CZA02] Cahier J.-P. & Zacklad M., « 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 21 - 23 janvier 2002, p.221-226, Hermes

- [CZM04] Cahier J.-P., Zacklad M., Monceaux A., « Une application du Web socio sémantique à la définition d'un annuaire métier en ingénierie », in Actes des journées francophone de l'*Ingénierie des Connaissances*, mai 2004, Lyon. url : <http://www.sociosemanticweb.org>
- [CCZ02] Caussanel J, Cahier J.-P., Zacklad M., Charlet J., « Les Topic Maps sont-ils un bon candidat pour l'ingénierie du Web Sémantique ? », *Conférence Ingénierie des Connaissances IC2002*, Rouen Mai 2002/
- [CHI97] Chiamarella Y., "Browsing and querying: two complementary approaches for multimedia information retrieval", *Hypertext-Information Retrieval- Multimedia*, HIM'97, 1997.
- [CIA05] 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.
- [DBE89] Dewey J. & Bentley A. F., "Knowing and the known". In J. A. Boydston (Ed.), *John Dewey: The later works*, 1925-1953 (Vol. 16, pp. 2-294). Carbondale: Southern Illinois University Press
- [DJR98] Dunlop M. D., Johnson C.W., Reid.. Exploring the layers of information retrieval evaluation, *Interacting with computers*, 10, 1998, p. 225-236.
- [KAH00] Kahn J. *Dynamique de l'interaction Personne-Système de recherche d'Information*, Thèse de doctorat, CNAM, Paris, 2000.
- [KHI95] Kim H., Hirtle S. C., "Spatial metaphors and disorientation in hypertext browsing", *Behaviour and Information technology*, 14, 1995, p. 239-250.
- [KOL98] Kolmayer E., « Démarche d'interrogation documentaire et navigation », In J.F. Rouet et B. De La Passardière (eds.), *Quatrième colloque Hypermédias et apprentissage*, Poitiers, 1998, p.121, 134.
- [MVR97] Magennis M., Van Rijsbergen C. J., "The potential and actual effectiveness of interactive query expansion", in *Proceedings of ACM SIGIR Conference on Research and Development in Information Retrieval*, Philadelphia, 1997, p. 324-332.
- [MAR95] Marchionini, G. (1995). *Information seeking in electronic environment*, Cambridge university Press, 1995.
- [MDS98] McDonald S., Stevenson R., "Navigation in hyperspace: an evaluation of the effects of navigational tools and subject matter expertise on browsing and information retrieval in hypertext", *Interacting with computers*, 10, 1998, p. 129-142.
- [RTR98] Rouet, J.-F., & Tricot, A. (1998). Chercher de l'information dans un hypertexte : vers un modèle des processus cognitifs, *Hypertextes et Hypermédias*, n° hors série, 1998, p. 57-74.
- [TDE00] Tricot, A., Drot-Delange, B., El Boussarghini, R., Foucault, B. (2000). Quels savoir-faire les utilisateurs réguliers du web acquièrent-ils? *Revue d'Intelligence Artificielle*, vol 14, n°1-2, 2000, p. 93-112.
- [ZAC05] Zacklad, M., (2005) Vers le Web Socio Sémantique: introduction aux ontologies sémiotiques, IC'2005 Ingénierie des Connaissances, Nice, Juin 2005
- [ZCP03] Zacklad M., Cahier J.P., Pétard X, « Du Web Cognitivement Sémantique au Web Socio Sémantique », *Journée « Web Sémantique et SHS »* du 7 mai 2003, url : <http://www.lalic.paris4.sorbonne.fr/stic/as5.html>