



**HAL**  
open science

## MAWA, dispositif de “ navigation sociale ”

Christophe Alcantara, Jean-Thierry Julia, Jean-Marie Pecatte, Nicolas Singer,  
Sylvie Trouilhet

► **To cite this version:**

Christophe Alcantara, Jean-Thierry Julia, Jean-Marie Pecatte, Nicolas Singer, Sylvie Trouilhet. MAWA, dispositif de “ navigation sociale ”. Document numérique et société, Sep 2006, France. pp.109-121. sic\_00378771

**HAL Id: sic\_00378771**

**[https://archivesic.ccsd.cnrs.fr/sic\\_00378771](https://archivesic.ccsd.cnrs.fr/sic_00378771)**

Submitted on 26 Apr 2009

**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.

## ***MAWA*, dispositif de « navigation sociale »**

### **Auteurs :**

Christophe ALCANTARA, doctorant en Sciences de l'information et de la communication, LERASS, Univ. Paul Sabatier-Toulouse 3, *christophe.alcantara@wanadoo.fr*

Jean-Thierry JULIA, maître de conférences en Sciences de l'information et de la communication, LERASS, Univ. Paul Sabatier-Toulouse 3, *jean-thierry.julia@iut-tlse3.fr*

Jean-Marie PECATTE, maître de conférences en Informatique, IRIT, Univ. Paul Sabatier-Toulouse 3, *jean-marie.pecatte@iut-tlse3.fr*

Nicolas SINGER, maître de conférences en Informatique, IRIT, Univ. Paul Sabatier-Toulouse 3, *nicolas.singer@iut-tlse3.fr*

Sylvie TROUILHET, maître de conférences en Informatique, IRIT, Univ. Paul Sabatier-Toulouse 3, *sylvie.trouilhet@iut-tlse3.fr*

Avec la collaboration de

Nikos SMYRNAIOS, docteur en Sciences de l'information et de la communication, GRESEC, Université Stendhal–Grenoble 3, *smyrnaios@gmail.com*

### **Résumé**

Le système *MAWA*, objet de cette communication, s'inscrit dans la thématique des systèmes multi-agents pour les TIC. Ensemble d'agents répartis sur l'internet, ceux-ci coopèrent pour enrichir les connaissances relatives à la navigation des utilisateurs. Après avoir replacé le dispositif *MAWA* dans le contexte des outils relatifs au « web collaboratif », le texte retrace les suites d'une première expérimentation du dispositif, dans le cadre d'une recherche commune entre chercheurs en informatique et en communication, dont les objectifs, par une approche plus qualitative des navigations recueillies, étaient notamment de parfaire les algorithmes au cœur du dispositif. Cette coopération précise notamment, dans un registre SIC, et conformément aux spécificités « pragmatiques » de l'outil, quelques pistes de développement pour le dispositif.

### **Mots-clés :**

Navigation. Web collaboratif. Approche pragmatique.

## ***MAWA*, dispositif de « navigation sociale »**

Christophe ALCANTARA, Jean-Thierry JULIA<sup>1</sup>,  
LERASS, Université Paul Sabatier-Toulouse 3  
Jean-Marie PECATTE, Nicolas SINGER, Sylvie TROUILHET  
IRIT, Université Paul Sabatier-Toulouse 3

Chercher et synthétiser de l'information à partir de ressources distribuées sur le web est un problème complexe auquel l'utilisateur est souvent confronté ; toute aussi délicate est en amont, du côté de la production de tels documents numériques, la tâche de celui qui met en ligne de l'information et souhaitera connaître l'usage qui en est fait.

Le système multi-agent *MAWA* traite de l'accès effectif à l'information ainsi que de son organisation, dans un univers où les documents sont distribués et où peu de règles de structuration sont imposées.

Autour de cet outil, la coopération entre deux équipes (l'une en SIC, l'autre en Informatique) retracée ici s'attache à un travail de confrontation de l'outil au terrain : ce sont les résultats de ces investigations que nous développons dans cette communication. Nous abordons dans une première partie, le contexte ayant présidé à la réalisation de l'outil, comparé à des outils similaires ; la deuxième partie concerne les premiers retours de son expérimentation auprès d'utilisateurs.

### **1. Positionnement de l'outil *Mawa***

*MAWA* optimise les conditions d'analyse quantitative de l'usage fait de l'information en ligne. Il s'attache en particulier à la façon dont celle-ci est réellement accédée : l'accès à chacune des pages web est recueilli (de façon transparente, et préservant à l'utilisateur les garanties juridiques de libre consultation) ; au fur et à mesure des navigations, ces informations sont intégrées dans un système qui construit et maintient une « cartographie » de consultation du web. Compte tenu de l'étendue des thèmes abordés sur le web, l'outil ne s'appuie pas sur l'utilisation d'ontologies prédéfinies : les analyses sont décontextualisées, basées sur les structures des pages, notamment méta-balises et taxinomie des liens.

---

<sup>1</sup> Avec la collaboration de Nikos SMYRNAIOS, GRESEC, Université Stendhal-Grenoble 3.

### **Un environnement hypertextuel**

Le web est architecturé autour de la notion d'hypertexte, ce qui place les parcours au centre de la consultation et de la recherche d'information. Ces parcours se font entre entités dont la définition varie en fonction du niveau de détail que l'on veut donner au parcours.

Ainsi du niveau de détail le plus fin au plus général on trouvera des parcours entre :

- URLs (qui sont un repérage technique d'une donnée sur le réseau internet)
- pages (qui sont des regroupements de ressources multimédia)
- sites (qui sont des regroupements de pages, classiquement attribuées à la même entité)
- ressources et/ou services (regroupements plus abstraits de pages ou de sites).

Une fois le type d'entité choisi pour représenter les étapes d'un parcours, le web est alors vu comme un graphe orienté dont les nœuds sont ces entités, et les arcs, les liens qui mènent d'une entité à une autre.

Les parcours étudiés sont un sous-graphe du graphe web défini ci-dessus, construit en observant les navigations d'un ou de plusieurs utilisateurs pendant un intervalle de temps déterminé. Tous les liens suivis ne sont pas d'égale importance et leur est associée une étiquette « force du lien » qui vise à renseigner sur le poids de chaque lien. Pour les parcours qui ne concernent qu'un seul utilisateur, la force du lien est le nombre de fois où ce lien a été suivi dans le parcours en question.

Plus formellement, le passage d'un utilisateur d'une entité web à une autre (un arc du graphe) se représente sous la forme d'un couple  $(v_i, v_j)$  où  $v_i$  et  $v_j$  sont les références des entités reliées ( $v_i$  mène à  $v_j$ ). On associe à ce lien suivi une force  $f$  qui est le nombre de fois où le lien a été emprunté par l'utilisateur. Cette force s'interprète comme la force d'interconnexion des deux entités. Une session de navigation (la représentation formelle d'un parcours pendant un laps de temps) est alors un ensemble de couples  $(v_i, v_j)$  suivis pendant la durée de la session, chacun ayant une force associée.

Nous distinguons les sessions de navigation utilisateur, qui caractérisent la navigation effectuée par un seul usager, et les sessions de navigation consolidées, qui sont obtenues par appariement de sessions utilisateur (PECATTE *et al.*, 2004), et qui visent à représenter une synthèse des navigations effectuées par un collectif d'utilisateurs ; elles permettent de faire émerger un comportement collectif. Le nombre de sessions utilisateurs ayant participé à la réalisation de la session consolidée est appelé « poids de la session ».

Le but de MAWA est donc d'un côté de permettre le recueil de parcours individuels, de l'autre d'assembler ces parcours en comportements collectifs<sup>2</sup>. Son originalité réside dans ce qui peut être évoqué sous la notion de « navigation sociale » : les expériences individuelles en navigation sont capitalisées au profit d'une collectivité d'utilisateurs, permettant de passer d'un mode de navigation individuel à un web plus « coopératif ».

### **MAWA et web coopératif**

Nous considérons le web coopératif comme l'ensemble des systèmes qui utilisent l'activité des utilisateurs pour ajouter de l'information aux pages. Il comporte, en particulier, les systèmes d'annotations (Nestor, Annotea, SyDoM...), les annuaires partagés de favoris (Del.icio.us...), et les outils de filtrage collaboratif (Letizia, WebWatcher, Alexa...).

Les systèmes d'annotations permettent de créer et de partager des commentaires sur le contenu des documents accessibles sur le web. Nestor est un exemple de système gérant (entre autres) les annotations (ZEILIGER *et al.*, 1997) ; il reste local au client et ne permet pas de

<sup>2</sup> Le lecteur trouvera dans (Singer, 2005) une description plus complète du fonctionnement, de la constitution à la gestion des sessions.

coopération entre utilisateurs. Le Web annotation service est un autre exemple, où les annotations peuvent être stockées sur un serveur (VASUDEVAN *et al.*, 1999). Ces systèmes peuvent être vus comme une possibilité pour les utilisateurs de modifier les pages web ou tout au moins d'en donner un complément accessible à ceux qui les suivront. Ils sont particulièrement utilisés dans le cas de l'annotation de documentations techniques, où les utilisateurs rendent compte de leurs expériences dans l'utilisation de telle ou telle fonctionnalité du produit décrit.

Les annotations peuvent utiliser les potentialités du web sémantique. Le système Annotea permet d'annoter des documents sans les modifier (KAHAN *et al.*, 2001) ; les annotations sont stockées en tant que méta-données sur des serveurs d'annotations et présentées aux utilisateurs ayant un client capable de les interpréter. SyDoM est un outil d'annotation pour le web sémantique améliorant la présentation de contenu pour une communauté donnée ; cette amélioration est réalisée en fonction de la modélisation du domaine faite par les membres de la communauté (ROUSSEY *et al.*, 2001).

Si ces systèmes répondent en partie à notre problématique, dans la mesure où les notes créées peuvent référencer d'autres pages, les soumissions de notes restent toutefois manuelles et dépendantes de la « bonne volonté » des acteurs humains.

Les annuaires partagés tels que Del.icio.us<sup>3</sup> permettent de mettre en commun des signets, appréciations, étiquettes... classés par thématiques. Sur le site *technorati*<sup>4</sup> par exemple, la taille d'affichage d'un thème dépend de son niveau d'actualité. Là encore, la quantité des données dépend de l'implication des participants. De plus, les informations n'arrivent pas spontanément à l'utilisateur, mais sont centralisées sur une même URL. Ces annuaires thématiques à base de signets pourraient être associés à un outil de « *profiling* » et entrer dans la famille des outils de filtrage.

Les outils de filtrage de l'information extraient automatiquement parmi des ensembles d'informations celles qui correspondent à un profil défini. On y distingue les systèmes cognitifs et les systèmes collaboratifs.

Le filtrage cognitif se base sur une indexation du contenu des documents pour les rapprocher d'un profil utilisateur. Ce rapprochement est donc la mise en adéquation des thèmes abordés dans les documents avec les thèmes intéressants l'utilisateur. Le filtrage collaboratif (LUMINEAU, 2003) met en commun les évaluations et les parcours des utilisateurs et utilise cette connaissance collective pour prédire un comportement ou faciliter la décision d'un utilisateur. Le filtrage collaboratif est souvent mis en œuvre *via* une architecture multi-agent (Easy-DOR, Broadway, InfoSleuth), approche qui propose des solutions pour des applications complexes, distribuées, pour lesquelles les données sont évolutives et où une supervision centralisée n'existe pas.

Le système à base d'agents Easy-DOR (CHEVALIER *et al.*, 2003) aide l'utilisateur dans sa recherche d'information grâce à une approche coopérative basée sur l'utilisation de la hiérarchie de signets définie par chacun d'eux. Si le document visité correspond à un signet appartenant à la structure multi-arbre regroupant l'ensemble des hiérarchies des divers utilisateurs, le système propose tous les documents pertinents correspondant au même centre d'intérêt ; la pertinence est calculée à partir de l'organisation des hiérarchies.

Letizia est un agent de navigation, lié à un seul utilisateur, qui travaille en tandem avec le navigateur (LIEBERMAN *et al.*, 2001), en se basant sur son comportement. Comme MAWA, il s'agit d'une assistance personnalisée ; toutefois, le système se limite à la seule la vision du client et ne propose pas d'adaptations des documents, comme l'ajout de liens virtuels.

<sup>3</sup> <http://del.icio.us/>

<sup>4</sup> <http://www.technorati.com/>

Broadway est un système très proche du système que nous proposons par ses méthodes et objectifs (TROUSSE, 2000). Il utilise l'historique des navigations pour prédire quelles seront les prochaines pages visitées par un utilisateur, éventuellement pour les lui conseiller : il analyse les *logs* des serveurs web et met en évidence des sessions qui sont utilisées comme « cas type » pour conseiller aux futurs utilisateurs une suite pertinente à leur navigation (techniques de raisonnement à partir de ces cas).

WebWatcher (AMSTRONG *et al.*, 1995) ressemble beaucoup à Broadway par ses objectifs, mais utilise des mots-clés pour conseiller certains liens à l'utilisateur. Il compare ces mots-clés à ceux des utilisateurs précédents, également aux mots contenus dans les balises des documents et des liens. Pour savoir si ces conseils sont pertinents, il demande à l'utilisateur en fin de session si ce dernier a trouvé ou non l'information recherchée. Bien qu'intéressant, ce « feedback » n'est pas actuellement utilisé dans la version actuelle de *MAWA*.

InfoSleuth est une plate-forme multi-agent pour l'interrogation et la gestion des modifications de textes semi-structurés (BAYARDO *et al.*, 1998). Il trouve une application dans le domaine du web où la majorité des ressources est de ce type. Il se compose d'un ensemble d'agents collaborant pour répondre à la requête de l'utilisateur. La pertinence de la recherche d'information dépend alors des compétences et des techniques implantées dans les agents de recherche (fouille de données, recherche basée sur les ontologies).

Pour clore cette classification non exhaustive, citons encore Alexa, outil se rapprochant de notre système. En plus d'être un moteur de recherche classique basé sur une indexation des pages, Alexa met à la disposition des utilisateurs une partie cliente, leur permettant ainsi d'obtenir de l'information sur leurs navigations. Il utilise cette information pour dresser un palmarès des sites les plus visités, mais également pour dresser une carte des sites qui ont un rapport entre eux. Cette carte est obtenue en étudiant le comportement des utilisateurs du web (ce qui le rapproche de *MAWA*), mais aussi en analysant les étiquettes des liens afin de procéder à des associations sémantiques entre sites. Enfin, des suggestions peuvent être formulées par les utilisateurs, de façon à enrichir la carte. Les différences entre Alexa et *MAWA* reposent sur l'architecture du système (centralisée *vs* répartie) et sur la façon dont les parcours sont analysés. Alexa se contente de repérer les sites web qui sont partie prenante d'un parcours et les considère alors comme liés ; *MAWA* s'intéresse autant aux parcours qu'aux sites, et cherche à obtenir des parcours types représentatifs d'un comportement collectif. En cela, il utilise des algorithmes capables de comparer des graphes web, et peut être utilisé pour analyser la façon dont un site ou des sites sont parcourus.

Au terme de ce panorama, si nous retrouvons dans *MAWA* plusieurs caractéristiques précédemment décrites (mise en relation des expériences individuelles, classification de ces expériences...), mentionnons plus précisément que le système :

- cherche à assister de manière transparente l'utilisateur et lui propose spontanément des documents susceptibles de l'intéresser ;
- tente de pallier le manque éventuel d'implication d'un utilisateur par la prise en compte implicite de son comportement lorsqu'il consulte ;
- est totalement distribué et continue de fonctionner même si des agents sont défectueux.

## **2. Retour sur expérimentation**

La collaboration entre équipes menée durant les derniers mois se situait au carrefour de plusieurs objectifs généraux : pour les chercheurs en informatique, contribuer à affiner les stratégies classificatoires du prototype *MAWA* sur la base de résultats autres que calculatoires ; pour ceux en SIC, prolonger d'un point de vue formel le travail autour de parcours et figures

type de navigation (JULIA, SMYRNAIOS, 2006) – le texte présent ne s’attachant qu’au premier pan de recherche évoqué.

L’étude que nous nous proposons relevait d’une « réintroduction » de l’acteur humain au sein du dispositif de modélisation *Mawa*, et s’attachait à approfondir, sous couvert des énoncés tels que les liens empruntés, cet échange entre utilisateurs et machines ; et plus encore entre utilisateurs eux-mêmes, *via* le dispositif *MAWA* d’« aide » à la navigation.

### **Lecture « qualitative »**

Le terrain expérimental investi conjointement – aisément accessible au vu d’un programme de recherche souscrit pour quelques mois<sup>5</sup> – se sera tourné vers les pratiques de navigation étudiantes des deux départements d’IUT. Il s’agissait – pour une durée qui se sera révélée réduite par rapport aux objectifs escomptés – de travailler sur les résultats issus des procédures *MAWA*, recueillis sur la base des navigations (libres et non-encadrées) des étudiants. Pour cette première confrontation, grandeur réelle, de *MAWA* à un ensemble de pratiques de navigation, il aura été entendu que le type de services web consultés importait peu au vu des finalités de l’expérimentation : de façon opérationnelle, celle-ci visait une lecture « informationnelle et communicationnelle », et non plus strictement calculatoire des résultats synthétisés jusque-là par cartographies de parcours consolidés. C’est en effet sur ce point – et la levée d’une hypothèque « qualitative » – qu’était pressenti un développement de *Mawa*, et l’accroissement de pertinence des résultats fournis par le dispositif. Dans l’état jusque-là du projet, seules des méthodes statistiques étaient utilisées durant la consolidation ; or, pour valider la méthode retenue pour le test de ressemblance, il fallait tenter d’analyser la « signification » de la consolidation de 2 parcours, et s’assurer que cette caractéristique supplémentaire était conservée.

Durant cette phase de recueil (janvier à mai 2006), ont été répertoriées, 16 916 sessions, 93 301 liens et 41 438 pages web. Mais la difficulté sera survenue du « décentrage » des motivations assignées dès lors aux productions informationnelles fournies par *MAWA* par rapport aux objectifs initiaux du dispositif : jusque-là imaginées à des fins automatiques de production et de présentation de sessions consolidées, elles n’en restent que difficilement exploitables au bénéfice d’une approche qualitative (notamment en l’absence de modalités de traitement d’informations qui restent éminemment brutes et statistiques).

Autour d’un tel terrain expérimental, les résultats – informationnels – auxquels l’exploitation humaine de ces données peut amener ne s’avèrent que de piètre pertinence : ils mettent en évidence des phénomènes assez attendus autour de seuls résultats prégnants qui viennent souligner des parcours autour de sites très ordinaires (MSN, Google, Yahoo !, Hotmail, etc.) ; ils sont le fait trop évident des navigations recueillies, et tel un « bruit », rendent difficile quelque attention à des phénomènes plus particuliers, non pas en termes de contenus de sites, mais de fonctionnement des algorithmes *MAWA*.

Ce constat est bien évidemment le fait du terrain choisi, qui, s’il pouvait augurer d’une telle cartographie du web – et il y aboutit comme d’autres dispositifs mentionnés plus haut y aboutissent – ne permet plus d’analyses supplémentaires (apport potentiel de *MAWA*, du fait de ses particularités, notamment centrées sur les sessions et non pas seulement sur les sites). Aux prises avec l’hétérogénéité des pratiques ici de mise – qui ne sont pourtant pas contradictoires mais bien conformes à des finalités en matière de cartographie – la « plus-value » de *MAWA* en termes de *sessions* se dissout au fil du recueil et devient inexploitable, ou plus exactement exploitable pour de seules conclusions qui rejoignent d’autres, en termes de *sites*, comme avec

---

<sup>5</sup> IRIT/ LERASS, 2005-2006, *Pratiques des internautes en situation d’autonomie*, Action scientifique de l’Université Paul Sabatier.

Alexa. Sur ce point d'une information qui sera au final cumulée, fait défaut une restitution diachronique des résultats, qui aurait pu être pertinente conjointement avec l'approche « sessions » de *MAWA*. Elle autoriserait en effet différentes vues sur une construction cartographique dynamique, si ce n'est en termes de sites (mêmes sites au palmarès), tout du moins et plus finement en termes de parcours différenciés au fil du temps vers les sites en question. Une telle « cartographie dynamique », sur la base des modalités d'accès effectives et non plus sur de seuls résultats en fréquence de consultation, pourra être un premier résultat – indirect – découlant de l'expérimentation, quand elle nécessiterait de modifier les algorithmes *MAWA* : pour l'instant, au terme des phases de consolidation et de fusion, et dans ce sens de façon dommageable, *MAWA* ne conserve pas la « mémoire » de telles modifications de la base d'informations relative au graphe des sessions.

### **Quelques pistes de recherche**

Si le point jusque-là évoqué n'était pas quant à lui contradictoire avec l'hétérogénéité des pratiques de navigation relevées, les difficultés d'un approfondissement plus « qualitatif » se heurtent dès lors à cette caractéristique du terrain choisi. Nous évoquions le nombre des informations restituées, difficilement exploitable par l'observation humaine. Or, pour parfaire les algorithmes *MAWA*, cette première étape s'avère nécessaire, et reste donc pour l'instant très partielle ; si elle ne préjuge pas à terme d'une impossibilité calculatoire d'œuvrer sur de tels algorithmes affinés, elle obère pour l'instant cette démarche heuristique d'une mise au jour de tels algorithmes. Ce constat, pour nous dans l'ordre méthodologique, rejoint assurément d'autres conclusions largement partagées dans le champ des SIC :

*« Construire un support de connaissances communes suppose donc que les individus appartenant à un même groupe [...] disposent de caractères stables, voire invariants dans leur réseau de connaissances, et que l'on puisse connaître (saisir, distinguer, décrire...) ces invariants. Une connaissance "commune" ne sort pas toute armée ni d'un individu, ni d'un groupe, ni d'un document »* (RABOUTET, CAHUZAC, 2006, p. 67).

La particularité de *MAWA*, rappelons-le, réside bien dans cette approche de la navigation sur le web par l'entremise des *sessions* effectives des usagers : elle n'est certes pas une approche « sémantique » de la chose (le dispositif ne procède à aucun recueil ou aucune exploration des contenus de la page web), mais se situe néanmoins loin d'une stricte approche en termes de « comptage ». Le recueil du « parcours » emprunté, à l'entre-deux de ces positions, pourrait être à rapprocher d'une approche « pragmatique » – une consultation en train de se faire –, tout du moins de la part automatique et observable d'une « pragmatique de navigation ». C'est là ce qui fonde le propre du dispositif, et son perfectionnement doit tenir compte de cette dimension inconnue d'autres dispositifs. Aussi, l'approche de la sorte automatique et informatique (et non sémantique) pourra être pertinente et conforme au regard du constat selon lequel « *ce mode collaboratif servait davantage à échanger des informations sur le projet et à générer une sociabilité de groupe [...] que véritablement mettre en synergie des compétences et des connaissances expertes* ». (GAGLIO, ZACKLAD, 2006, p. 97)

La spécificité de *MAWA* consiste notamment en cette part d'échange (automatisé) d'informations autour d'un projet ; et c'est bien dans le sens d'une telle sociabilité, prise en charge pour partie par la machine, que nous évoquions effectivement l'expression de « navigation sociale », qui va dans le sens de la remarque de E. Katz, comme quoi « *les technologies de communication nous mettent en rapport les uns avec les autres selon des modalités qui sont largement indépendantes des messages transmis* » (KATZ, 1989, p. 83).



Cette « navigation sociale » est encore conforme aux conclusions de toutes premières investigations en matière de multimédia, dans les années 80-90 (DILLON, 1996), qui préconisaient, en place des structures formelles, compositions et organisations des différents contenus (et à l'instar d'une option pragmatique dans le registre de la langue comme quoi le « sens » est *sous-déterminé* par le matériel linguistique), de ne s'attacher qu'à la *structure rhétorique* des documents numériques, de leur *structure argumentative* (ici dans les transitions « argumentatives » que sont les liens), ou encore des « scénarios d'interaction » (TRICOT, ROUET, 1998) : ce que tente de modéliser MAWA à l'occasion de cette « navigation sociale ».

Force est pourtant de constater à ce jour, que dans le but d'en mettre au jour ses modalités, un passage plus « informationnel » paraît obligé avant que de pouvoir modéliser une telle « sociabilité » électronique. Plusieurs solutions permettraient toutefois de franchir une telle étape opératoire et de proposer des voies pour la modélisation et la consolidation des navigations recueillies. Afin de « rétrécir » momentanément la cartographie, et rendre exploitable les données fournies par MAWA, des aménagements d'expérimentation sont à envisager :

- délimiter « dans le temps » le recueil des navigations et circonscrire par exemple l'analyse autour de phénomènes restreints, liés à un événement ou bien encore artefact suscité aux fins de l'expérimentation ;
- délimiter « dans l'espace », c'est-à-dire recentrer autour d'un public d'internautes qui pourrait plus aisément être caractérisé autour d'intérêts partagés ; si une « communauté de pratique » (LEFEBRE *et al.*, 2004) était dans notre terrain d'expérimentation évidente – mais par trop disparate –, sans pour autant viser une communauté autour d'une tâche (qui risque de ne renvoyer que similitude des sessions), une communauté d'« intérêt »<sup>6</sup> semble une condition nécessaire pour pouvoir faire émerger un perfectionnement des algorithmes. Dans ce sens, de nouvelles expérimentations MAWA sont à envisager, par exemple en milieu professionnel, autour d'une même appartenance organisationnelle, et par là autour d'intérêts contingents à un même secteur ;
- délimiter « l'objet » : une voie pourrait enfin être exploitée dans le fait de ne recueillir des sessions qu'autour d'un unique site (à l'intérieur et/ou connexes à ce site), site identifié dans sa thématique, qui servirait de « pivot » pour reconnaître une telle communauté d'intérêt. Ce pourrait être le cas autour d'un public étudiant, mais cette fois autour d'un site d'enseignement, ou encore dans un autre type d'organisation, autour du site interne d'une entreprise.

Enfin, de façon plus « sémantique », une dernière voie s'ouvrirait, qui complexifierait d'autant le dispositif. Il s'agirait, à l'image des systèmes qui se retrouvent autour de la procédure de « marquage », de faire que l'internaute puisse lui-même caractériser, thématiquement, le champ d'intérêt à l'instant de sa session (de façon libre, contributive, ou bien de façon strictement réactive à l'intérieur d'un thésaurus, ou bien encore d'une liste de mots-clés jusque-là recueillie). Une telle optique, jusque-là cartographique et dès lors « qualitative », pourrait être un substitut à l'impossibilité d'une autre caractérisation « sémantique » de la session (éventuellement durant le seul temps de constitution de la base des sessions MAWA). Si tel devait être le cas – qui marquerait certes un changement dans la stratégie de développement de MAWA –, reste que la spécificité du dispositif pourrait être de

---

<sup>6</sup> « Groups of people who share a concern, a set of problems or a passion about topic, and who deepen their knowledge and expertise in this area by interacting on an ongoing basis » (Soulcier, 2004), d'après (Wenger, 1998).

co-construire, simultanément et conjointement, graphes des sessions et graphes des mots-clés, selon des algorithmes se correspondant et se complétant.

## Conclusion

A ce stade de la collaboration entre les équipes, il n'aura certes pas été possible de dégager quelques résultats prégnants, autres que ceux auxquels aboutissent d'autres dispositifs autour d'un web collaboratif. Mais rappelons que les buts de l'expérimentation n'étaient pas ici, et que cette même expérimentation était notamment guidée par l'intuition de prolonger le dispositif *MAWA* dans le sens d'une lecture plus « informationnelle » si ce n'est sémantique. Les résultats obtenus dans ce sens auront permis de discerner la stricte spécificité de *MAWA* – approche pour l'instant plus « pragmatique » que « sémantique » – et d'envisager de possibles prolongements afin d'affiner les algorithmes mis en œuvre.

Ainsi, le programme d'étude autour du dispositif se poursuit (sur la base des observations mentionnées plus haut : cartographie dynamique ; stratégies de délimitation du recueil, éventuelle adjonction d'un module « sémantique » par mots-clés de l'internaute), qui pourra dès lors articuler plus profondément les deux regards que le présent travail aura tenté de relier.

## Références bibliographiques

- AMSTRONG, R., FREITAG, D., JOACHIMS, T., MITCHELL, T., (1995). WebWatcher: A learning Apprentice for the World Wide web. *AAAI Spring Symposium on Information Gathering from Heterogeneous, Distributed Environments*. Stanford (CA) : p. 6-12.
- BAYARDO, R., BOHRER, W., BRICE, R., CICHOCKI, A., FOWLER, G., HELAL, A., KASHYAP, V., KSIEZYK, T., MARTIN, G., NODINE, M., RASHID, M., RUSINKIEWICZ, M., SHEA, R., UNNIKRIISHNAN, C., UNRUH, A., WOELK, D., (1998). Infosleuth: agent-based semantic integration in open and dynamic environments. In *Reading in Agents*, HUHNS, M. N., SINGH, M. P., eds. San Francisco, (CA) : Morgan Kaufmann, p. 205-216.
- CHEVALIER, M., JULIEN, C., (2003). Interface coopérative et adaptative pour la recherche d'informations sur le web. *19<sup>èmes</sup> journées des bases de données avancées*. Lyon.
- DILLON, A. P., (1996). Myths, Misconceptions, and an Alternative Perspective on Information Usage and the Electronic Medium, in *Hypertext and Cognition*. ROUET, J.-F., LEVONEN, J. J., DILLON, A. P., SPIRO, R. J., eds. Mahwah (NJ) : Lawrence Erlbaum Associates, p. 25-42.
- GAGLIO, G., ZACKLAD, M., (2006). La circulation documentaire en entreprise comme analyseur de pratiques professionnelles. Une étude de cas. *Sciences de la Société*, n° 68, *Dimensions sociales du document*, p. 92-109.
- JULIA, J.-T., SMYRNAIOS, N., (2006). Figures d'un cheminement, ou "Ceci n'est pas un parcours multimédia". *Les Cahiers de Champs Visuels*, n° 3-4, *Le parcours multimédia* (à paraître).
- KAHAN, J., KOIVUNEN, M.-R., PRUD'HOMMEAUX, E., SWICK, R., (2001). Annotea: An Open RDF Infrastructure for Shared Web Annotations. In *Proceedings of the WWW10 International Conference*. Hong Kong.
- KATZ, E., (1989). La recherche en communication après Lazarsfeld. *Hermès*, n° 4, p. 72-92.
- LEFEBRE, P., ROOS, P., SARDAS, J.-C., (2004). Les théories des communautés de pratique à l'épreuve : conditions d'émergence et organisation des communautés. *Systèmes d'Information et Management*, vol. 9, n° 1, p. 25-48.
- LIEBERMAN, H., FRY, C., WEITZMAN, L., (2001). Why Surf Alone?: Exploring the web with Reconnaissance Agents. *Communications of the ACM*, p. 69-75.
- LUMINEAU, N., (2003). Un tour d'horizon du filtrage collaboratif. *Rapport de l'action spécifique CNRS sur la personnalisation de l'information (RTP 9)*.

- PÉCATTE, J.-M., SINGER, N., TROUILHET, S., (2004). Mawa : a multi-agent tool for cooperative information retrieval. Nice : *CNRIUT*, p. 97-104.
- RABOUTET, C., CAHUZAC, H., (2006). Le document comme espace de connaissance(s). Entre fonction sociale et fonction cognitive. *Sciences de la Société*, n° 68, *Dimensions sociales du document*, p. 60-79.
- ROUSSEY, C., CALABRETTO, S., PINON, J.-M., (2001). SyDoM: A Multilingual Information Retrieval System for Digital. In *Proc. International Conference ICCP/IFIP On Electronic Publishing (ELPUB'2001)*. Kenterbury (UK) : p. 150-164.
- SINGER, N., PÉCATTE, J.-M., TROUILHET, S., (2005). The Multi-agent Cooperative Navigation System Mawa: A Model of Dynamic Knowledge Specialization for a User-centric Analyse of the web. In *International Conference on Intelligent Agents, web Technologies and Internet Commerce (IAWTIC'2005)*. Vienna (Austria).
- SOULIER, E., (2004). Les communautés de pratique au cœur de l'organisation réelle des entreprises. *Systèmes d'Information et Management*, vol. 9, n° 1, p. 3-24.
- TRICOT, A., ROUET, J.-F., dir., (1998). Les hypermédias : approches cognitives et ergonomiques. *Hypertextes et Hypermédias*, n° hors série.
- TROUSSE, B., (2000). Evaluation of the Prediction Capability of a User Behaviour Mining Approach for Adaptive web Sites. In *Proceedings of the 6th RIAO Conference Content-Based Multimedia Information Access*. Paris.
- VASUDEVAN, V., PALMER, M., (1999). On web Annotation: Promises and Pitfalls of Current web Infrastructure. *HICSS'99*.
- WENGER, E., (1998). *Communities of practice*. New-York : Cambridge University Press (réed. 2004).
- ZEILIGER, R., REGGERS, T., BALDEWYNS, L., JANS, V., (1997). Facilitating web Navigation: Integrated Tools for Active and Cooperative Learners. *International Conference on Computers in Education'97*. Charlottesville (VA).