HAL
open science

Recherche d'informations personnalisées

Christelle Petit-Rozé, Emmanuelle Grislin-Le Strugeon, Mourad Abed, Guillaume Uster, Christophe Kolski

To cite this version:

Christelle Petit-Rozé, Emmanuelle Grislin-Le Strugeon, Mourad Abed, Guillaume Uster, Christophe Kolski. Recherche d'informations personnalisées. Conference Internationale NîmesTIC 2000, Sep 2000, Nîmes, France. hal-03396521

HAL Id: hal-03396521

https://uphf.hal.science/hal-03396521v1

Submitted on 14 Feb 2024

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.

Conference Internationale NîmesTIC 2000, Ingénierie des Systèmes et NTIC, Nîmes, France, september, 2000.

Recherche d'informations personnalisées

Christelle Rozé (*,**), Emmanuelle Grislin-Le Strugeon (*), Mourad Abed (*),
Guillaume Uster (**), C. Kolski(*)

(*) LAMIH – RAIHM, Université de Valenciennes et du Hainaut-Cambrésis, Le Mont Houy, 59313 Valenciennes cedex 9, France.
(**)INRETS – ESTAS, 20, rue Elisée Reclus, 59650 Villeneuve d'Ascq, France.

{christelle.roze, strugeon, abed, kolski}@univ-valenciennes.fr ; guillaume.uster@inrets.fr

Résumé : Cet article décrit une méthode de recherche d'informations personnalisée. Ceci consiste à présenter toute l'information nécessaire et uniquement l'information nécessaire en fonction de l'utilisateur concerné. Ce système a la particularité d'être géré par des agents intelligents. Cette " intelligence " sera composée de leur capacité à cerner les besoins spécifiques d'un utilisateur donné et à rechercher les informations pertinentes pour répondre à ces besoins.
Cette méthode sera appliquée sur un système d'information voyageur destiné, d'une part, à faciliter l'usage des transports collectifs, en présentant l'offre et les conditions d'accès, et d'autre part, améliorer le confort des usagers en les accompagnant dans l'exécution de leur déplacement.

Mots-clés : systèmes d'information, transport, agents intelligents, recherche d'informations.

Abstract : *This article describes a method for the personalization of the information. This consists in presenting all, and only, the information that is necessary according to the concerned user. This system has the characteristic to be managed by intelligent agents. This " intelligence " will be made up of their capacity to determine the specific needs for a given user and to seek relevant information to meet these needs.*
This method will be applied to a traveller information system intended to facilitate the use of collective transport, by presenting the offer and the conditions of access, and to improve the comfort of the users by accompanying them in their travel.

Keywords : *information system, transport, intelligent agents*

1 Introduction

Un "système d'information", comme son nom l'indique, est destiné à fournir de l'information à un utilisateur. En fait, il devrait, idéalement, permettre à l'utilisateur de récupérer de *l'information* à partir des *données* auxquelles a accès le système. Or, cette transformation des données en information, à savoir cette plus-value apportée aux données qui sont triées, classées, validées, est bien souvent négligeable : le système laisse à l'utilisateur la charge de retrouver l'information dans la masse de données qui lui est fournie.

C'est pourquoi notre objectif est d'aider l'utilisateur dans sa démarche de recherche d'informations en lui fournissant un résultat personnalisé. En effet, dans le contexte actuel de mise à disposition de grandes masses de données, l'heure est à la *personnalisation* de l'information nécessaire et suffisante en réponse à une requête particulière, et non, une large diffusion de toutes les informations disponibles.

Nous allons débuter cet article en décrivant le niveau de personnalisation que nous pouvons obtenir actuellement dans les moteurs de recherche, puis nous présenterons notre méthode de personnalisation.

2 Les systèmes existants

2-1 Personnaliser l'information

Personnaliser l'information, cela signifie s'adapter aux buts, préférences et capacités de l'utilisateur :

- l'adaptation aux buts consiste à prendre en considération le but que cherche à atteindre l'utilisateur et donc à se focaliser sur ses centres d'intérêts. Par exemple, lors d'une recherche par internet, le but de l'utilisateur peut être d'obtenir un rapport sous le format Postscript sur le thème "agent" ; ce qui diffère notablement du but "trouver un moyen (téléphone, fax ou e-mail) de contacter M. Dupont de la société X. Dans le domaine de la supervision de procédé, un but de l'utilisateur peut être de trouver le composant à l'origine d'un dysfonctionnement. Cette prise en compte du ou des buts de l'utilisateur (et par conséquent, l'adaptation à la tâche de l'utilisateur) est une des préoccupations fondamentales des concepteurs d'interfaces adaptatives.
- l'adaptation aux préférences de l'utilisateur existe de façon répandue dans les interfaces adaptables, à savoir paramétrables par l'utilisateur, grâce à un ensemble d'options et de menus de type "personnaliser". Un exemple d'une forme un peu plus "automatisée" d'adaptation est donnée par la configuration d'un environnement graphique particulier et l'exécution de certaines applications utiles à l'utilisateur lors de sa connexion à un système en s'identifiant.
- l'adaptation aux capacités de l'utilisateur consiste à lui délivrer de l'information selon une forme et dans des délais acceptables (utilisables) par lui. Par exemple, cela peut consister à se restreindre à la limite de temps (attente) dont dispose l'utilisateur avant récupération d'une information. Cela peut se concrétiser également par l'adaptation aux moyens matériels et/ou logiciels dont dispose l'utilisateur pour récupérer ou visualiser l'information fournie (ex. : taille d'écran, débit du réseau, etc.)

La personnalisation de l'information conduit ainsi nécessairement à modéliser l'utilisateur du système d'information. Nous nous sommes intéressés à la façon dont cela a été réalisé à l'aide d'agents logiciels intelligents et dans le contexte de la personnalisation de l'information disponible sur Internet. En effet, l'Internet constitue l'exemple typique d'une source de données difficile à exploiter.

2-2 Agents personnels de recherche d'information sur internet

Ils existent des systèmes capables d'aider les utilisateurs dans leur recherche d'informations sur Internet. Ces systèmes offrent de plus en plus de services. J. Delgado [4] référence trois types différents de systèmes d'informations :

- Les moteurs de recherche qui proposent une liste de réponses suite à une demande de l'utilisateur.
- Les modèles " push " d'informations capables de proposer des informations sans demande explicite de l'utilisateur. Dans cette catégorie, nous pouvons ajouter le système LETIZIA, créé par H. Lieberman [8]. C'est un assistant personnel pour le parcours du Web. Il enregistre les URLs choisies par l'utilisateur, lit les pages et dessine un profil de l'utilisateur au fur et à mesure que celui-ci visite des pages. À partir de là, il recherche d'autres pages susceptibles d'intéresser l'utilisateur et présente ses résultats sur une fenêtre indépendante.
- Les agents intelligents.

Nous avons affiné cette classification. Et plus précisément, nous avons sous-classé les systèmes à base d'agents intelligents. La classification qui suit est ordonnée suivant le degré de personnalisation de l'information fournie à l'utilisateur.
Au premier niveau, nous classons les systèmes se comportant comme de simples moteurs de recherche basés sur une architecture d'agents intelligents. Ils fournissent une information suite à une demande de l'utilisateur exprimée sous forme de mots-clés. Souvent, ces systèmes contiennent une arborescence de catégories qui facilite leurs recherches. Par exemple : Voila[1], Lokace[2], Altavista[3] …
Au niveau suivant, nous classons les systèmes basés sur une architecture " multi-agents ", capable de lancer une recherche sur plusieurs moteurs en même temps. Par exemple : Bulls Eye[4], Copernic[5], Webfer- retPro[6], Mata Hari[7] …
Sur le niveau supérieur, nous classons les agents capables de fournir une information en fonction d'une liste de préférences établie au préalable. C'est le cas de Cognisoft[8] qui diffuse des informations aux utilisateurs en fonction de leurs besoins et préférences, ou encore d'Autonomy Knowledge Server[9] capable de comprendre les centres d'intérêts des utilisateurs. Il peut créer des agents capables de rechercher des informations pertinentes suivant les intérêts personnels de leur utilisateur [8, chap 4].
Enfin au dernier niveau, nous classerons notre système. Nous concevons un système capable de rechercher des informations pour un utilisateur mais également de lui fournir un résultat *personnalisé,* dans le but de faciliter et d'accélérer la recherche.

A partir du niveau 3, une certaine modélisation de l'utilisateur est perceptible. Ces modèles peuvent être envisagés selon deux perspectives :
- la flexibilité et la richesse du modèle : dans quelle mesure de nouvelles caractéristiques peuvent être ajoutées au modèle et quelle peut être la nature de ces caractéristiques ;
- la capacité d'apprentissage du logiciel : quel est le moyen — manuel, automatique, semi-automatique — utilisé pour intégrer de nouvelles caractéristiques au modèle.

2-3 Richesse du modèle de l'utilisateur

Le modèle de l'utilisateur est principalement basé sur un ensemble de mots clés qui caractérisent ses centres d'intérêts. Ces mots-clés sont soit simplement listés (reliés par des booléens), soit organisés plus finement, par exemple à l'aide de réseaux sémantiques [9].
La personnalisation concernant les domaines d'intérêt peut être évidemment enrichie par différentes autres caractéristiques de l'utilisateur. Par exemple, il est possible d'ajouter au modèle les connaissances ou besoins de l'utilisateur concernant certains de ces domaines. La manière la plus simple de réaliser ceci semble de définir un ensemble de profils caractéristiques d'utilisateurs. Le système SAIRE [9], par exemple, distingue 3 "stéréotypes" (*science*, *general* et *non science*) selon le niveau de connaissances scientifiques de l'utilisateur. Le système ConCall [15] classe les utilisateurs selon leur profil "métier" (chercheur, éditeur, ...) et cible ainsi leurs besoins spécifiques.
Néanmoins, les modèles existants apparaissent relativement "pauvres" au regard des connaissances utilisées par certains systèmes d'assistance (en particulier appliqués à d'autres domaines) et qui intègrent des éléments tels que les croyances de l'utilisateur sur le système (comme dans RESCUER [14] ou BGP-MS [10]), ou une représentation de sa tâche …

2-4 Capacité d'apprentissage de l'agent intelligent

La capacité d'apprentissage est également un aspect qui permet de distinguer les différents systèmes.
L'acquisition de nouvelles caractéristiques de l'utilisateur peut se faire :
- manuellement : en laissant le soin à l'utilisateur de juger de ce qui l'intéresse ou non. Le système propose par exemple une liste de thèmes ou de sites web à l'utilisateur qui valide ceux qui correspondent à ses besoins [2] [12] ;
- semi-automatiquement : l'agent logiciel propose des informations qui lui semblent répondre aux besoins de l'utilisateur, lequel valide ensuite explicitement ce qui lui convient effectivement ;

[1] http://www.voila.com
[2] http://www.lokace.com
[3] http://www.altavista.com
[4] http://www.intelliseek.com
[5] http://www.copernic.com
[6] http://www.ferretsoft.com
[7] http://www.thewebtools.com
[8] http://www.verity.com
[9] http://www.autonomy.com

- automatiquement : le logiciel utilise l'historique des actions de l'utilisateur (sites fréquemment visités, liens suivis, par exemple) pour enrichir le modèle par inférence [3].

Ces différentes méthodes d'apprentissage de connaissances sur l'utilisateur peuvent être combinées au sein d'un même système : la validation "manuelle" a lieu lors des premières utilisations, puis le système enrichit par lui-même le modèle progressivement. Par exemple, le système Ringo [12] crée les profils par classification à partir des centres d'intérêts exhibés par les utilisateurs précédents. UM-tool [1] utilise des stéréotypes d'utilisateurs et un système de re-classification dynamique.

Dans notre système, nous utilisons une méthode d'apprentissage des connaissances hybrides : dans un premier temps une acquisition manuelle, puis une acquisition automatique. Ceci nous permet d'enrichir le modèle utilisateur par rapport aux modèles classiques et donc de personnaliser le plus possible l'information.

Nous allons donc présenter la méthode de personnalisation que nous avons retenue, puis dans une seconde partie nous expliquerons le système sur lequel sera appliquée cette méthode.

3 Méthode de personnalisation

Ce paragraphe est destiné à présenter la démarche suivie pour personnaliser une information ainsi que l'architecture adoptée.

3-1 Démarche proposée

Nous avons choisi de personnaliser l'information suivant deux phases : une étape de personnalisation partielle puis de personnalisation globale. Ces deux phases consistent à récupérer toutes les données possibles sur l'utilisateur et de les ré-utiliser pour ses futures autres recherches.

3-1-1 Initialisation/Paramétrage de la méthode

La première phase consiste à classer tout nouvel utilisateur en fonction de son profil dans une (ou plusieurs) catégorie(s).
Dans notre système, le profil est déterminé à l'aide d'un ensemble restreint de données obtenues par des questions directes à l'utilisateur lors de sa première connexion au système. Par ce questionnaire, l'utilisateur donne explicitement des exemples de ces centres d'intérêt (les exemples positifs) et des exemples de ce qu'il n'aime pas (les exemples négatifs). L'utilisateur est donc représenté dans le système par un ensemble de mots-clés organisés à l'aide de réseaux sémantiques [9].
Le système peut ensuite déduire un certain nombre de comportements ou informatiquement parlant, de valeurs d'attributs, probables pour cet utilisateur lambda.

Lors de l'inscription du premier utilisateur du système, deux solutions se présentent :
- Soit nous disposons d'une base d'exemples avant l'utilisation de cette méthode. Dans ce cas, nous pouvons définir des catégories par rapport à cet échantillon d'utilisateur et classer le nouvel utilisateur par rapport aux catégories créées.
- Soit nous ne disposons d'aucune information. Dans ce cas, nous ne pouvons pas déterminer les catégories d'utilisateurs. Nous appliquerons l'algorithme de classification au fur et à mesure des utilisations de cette méthode par les utilisateurs.

Les catégories sont définies à partir de données sur lesquelles un algorithme issu d'une méthode de data-mining est appliqué.

3-1-2 Recherche d'information

La seconde phase de cette méthode complète la semi-personnalisation. Le système possède des renseignements sur chaque utilisateur, à savoir des informations sur son état civil, sur ses préférences ainsi que sur les recherches déjà effectuées. La personnalisation concernant les domaines d'intérêt peut être évidemment enrichie par différentes autres caractéristiques de l'utilisateur.
Le modèle de l'utilisateur est complété au fur et à mesure des utilisations du système.

Lorsqu'un utilisateur désire faire une recherche particulière, il se connecte à notre système et pose sa requête. Le système va tout d'abord vérifier si une recherche similaire a déjà été demandée. Si c'est le cas, il pourra utiliser la réponse choisie par l'utilisateur, l'adapter à la nouvelle demande et la proposer. Sinon, à partir des résultats

antérieurs, le système apprendra les centres d'intérêts de l'utilisateur afin de constituer une nouvelle requête suffisamment précise pour obtenir des résultats ciblés sur les attentes de l'utilisateur. Pour cela,

- soit le système crée directement une requête très précise en fonction du profil de l'utilisateur, lance une recherche sur cette requête, puis propose à l'utilisateur les résultats obtenus.
- soit le système crée une requête simple, lance la recherche, puis filtre les informations en fonction du profil de l'utilisateur avant de lui présenter les résultats.

Toutes les recherches demandées par l'utilisateur seront ajoutées à la base de données ainsi que le résultat choisi.

La figure 1 résume les deux premières étapes de la personnalisation.

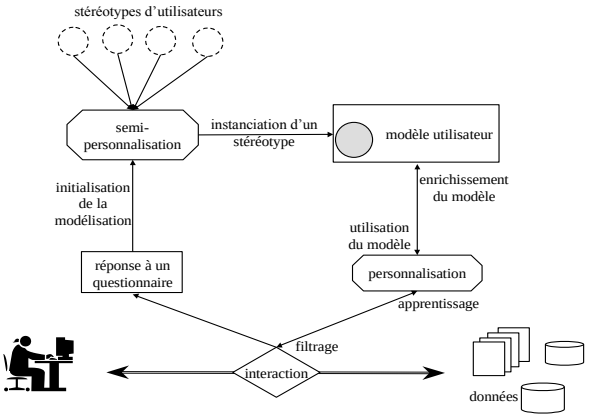

Figure 1 : Cycle de personnalisation de l'information

3-1-3 Envoi d'informations à l'utilisateur

Le système doit être capable de proposer des informations à l'utilisateur sans que celui-ci ne les ait demandées (c'est le principe du push-up). Il existe deux types d'informations possibles.

Soit le système envoie des informations de type " semi-personnalisé ". Dans ce cas, le système se base sur les différentes catégories définies et envoie les informations à tous les utilisateurs d'un même profil. Il s'agit en quelque sorte d'une " publicité " ciblée en fonction d'un profil d'utilisateurs.
Soit le système envoie des informations de type " personnalisé ". Dans ce cas, le système n'envoie cette information qu'à un seul utilisateur. Il peut s'agir d'une précision à une requête antérieure ou une autre proposition de réponse mieux adaptée à la demande de l'utilisateur.

3-2 Architecture proposée

Pour concevoir la méthode de personnalisation, nous avons besoin d'agents logiciels. Ces agents auront plusieurs rôles. Tout d'abord, lors de l'initialisation du système, ils devront définir des catégories. Pour cela, ils devront utiliser les données [13] représentatives des utilisateurs et à l'aide de techniques issues du data-mining [7], ils devront élaborer des catégories [9][15].
Ensuite, ils devront maintenir le système. Pour cela, ils seront chargés de déterminer la (ou les) catégorie(s) dans laquelle (ou lesquelles) ils devront introduire ce nouvel utilisateur. Puis au fur et à mesure de ses utilisations, les agents logiciels vont emmagasiner des informations qui serviront à mettre à jour continuellement leurs connaissances sur cet utilisateur.
Enfin, d'autres agents auront pour mission de renseigner l'utilisateur convenablement (c'est-à-dire conformément à son profil).
Pour résumer le rôle d'un agent logiciel, nous pouvons donner ces 4 fonctionnalités :

- Tout d'abord, être capable de rechercher une information demandée.
- Etre capable d'apprendre les préférences de l'utilisateur.
- Etre capable de choisir l'information par rapport aux préférences apprises.
- Et enfin, être capable de réduire le volume d'informations destinées à l'utilisateur.

D'un point de vue technique, l'agent sera composé des caractéristiques suivantes :

- un module de communication [5] pour coopérer avec les autres agents, se partager les tâches de la recherche, comparer les résultats obtenus.
- un module d'apprentissage pour apprendre les préférences des utilisateurs, apprendre à s'adapter à l'environnement.
- une base de connaissances des moteurs de recherche ou des sites concernant un thème particulier, également de la connaissance sur les autres agents, sur ce qu'ils connaissent, ce qu'ils savent faire.
- un module de contrôle.
- et des capacités de mobilité pour parcourir le réseau à la recherche d'information.

4 Application

Notre objectif applicatif concerne l'information liée aux transports terrestres de personnes, dans le but, d'une part, de faciliter l'usage et l'accès aux modes de transport collectifs et, d'autre part, d'améliorer le confort des usagers en les accompagnant dans leurs déplacements [16].
L'information à délivrer doit permettre la préparation d'un déplacement et son suivi. La préparation de déplacement répond aux questions de choix des modes de transports à utiliser et de leur chaînage : "quel(s) mode(s) de transport utiliser pour aller d'un point A à un point B ? A quels horaires ? Avec quel(s) changement(s) ? A quel coût ?" etc. Le suivi de déplacement doit permettre d'informer l'utilisateur des modifications éventuelles à apporter en temps réel à son déplacement en fonction des événements extérieurs tels qu'une perturbation du trafic, un incident, …
Actuellement, chaque exploitant détient et diffuse les données concernant son réseau et son mode de transport. La difficulté réside en la conjugaison des différents réseaux de transports et des données associées.
L'information doit être mise à la disposition du public par l'intermédiaire d'un système d'information qui intégrerait ces différentes données [10] pour constituer une information dite "multimodale" (le système est dénommé SIM).
Différents champs d'étude sont donc ouverts dans ce contexte applicatif : recueil de données en réseau ; intégration de données hétérogènes ; planification ; personnalisation de l'information ; temps réel.
Notre intervention se situe actuellement au niveau de la conception du SIM et concerne plus particulièrement la faisabilité de l'utilisation d'agents intelligents dans ce contexte. La méthode proposée dans le paragraphe précédent intervient dans la présentation personnalisée de l'information multimodale.

L'architecture générale proposée pour le SIM est basée sur des agents logiciels (Cf. figure 2) : des agents d'assistance intelligents au sein du système lui-même, capables de déléguer des tâches de recueil d'information à des agents mobiles et des agents réactifs situés sur les sites des entreprises exploitantes des moyens de transport.

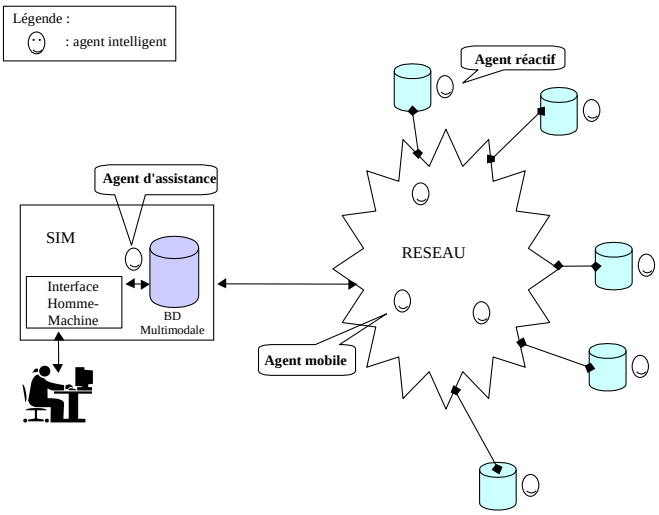

Figure 2 ***: Fonctionnement général du SIM***

[10] Une autre étude, menée par l'INRETS, est actuellement en cours à propos des difficultés juridiques et institutionnelles engendrées par la mise en place d'un tel système.

5 Conclusion

Nous avons présenté dans cet article nos premières idées concernant une méthode de recherche d'informations personnalisées.
Une fois finie, cette méthode sera d'une grande aide pour tous les moteurs de recherche. Dans ce monde d'information, la personnalisation est aussi un moyen de trouver plus rapidement un document, ou encore de reprendre une recherche là où elle avait été arrêtée.

6 Références

[1] G. Brajnik, G. Guida, C. Tasso. *User Modeling in Expert Man-Machine Interfaces: a Case Study in Intelligent Information Retrieval*. IEEE Transactions on SMC, vol. 20, n°1, Jan-Feb. 1990, pp. 166-185.

[2] L. Chen, K. Sycara. *WebMate: A Personal Agent for Browsing and Searching*. In Proceedings of the 2nd International Conference on Autonomous Agents and Multi Agent Systems, Minneapolis, MN, May 10-13, 1998.

[3] D. D'Aloisi, V. Giannini. *The Info Agent: an Interface for Supporting Users in Intelligent Retrieval*. Proceedings. of NIR-IT'95, 21-23 Nov. 1995.

[4] J. A. Delgado R. *Agent-based information filtering and recommender systems*. Thesis of doctor. Institute of Technology, Nagoya (Japan), March 2000.

[5] T. Finin, Y. Labrou et J. Mayfield. KQML as an agent communication language. In *Software agents*, MIT Press, 1997.

[6] P. Gendre. Systèmes d'information multimodale : une bibliographie commentée. Rapport technique, CERTU, Lyon, mars 1999.

[7] R. Lefébure et G. Venturi. *Data Mining*. Eds Eyrolles, 1998.

[8] H. Lieberman. Autonomous interface agents. In *ACM Conference on Human-Computer Interface*, pages 67-74. ACM Press, March 1997.

[9] J. Odubiyi, D. Kocur, S. Weinstein, N. Wakim, C. Gokey, J. Graham. *SAIRE - A scalable agent-based information retrieval engine*. Proceedings of Agents"97, ACM Press, 1997.

[10] W. Pohl, J. Höle. *Mechanisms for flexible representation and use of knowledge in user modeling shell systems*. In A. Jameson, C. Paris & C. Tasso (eds.), "User modeling: Proceedings of the 6th INt. Conf.", Wien, New-York, Springer Verlag, 1997.

[11] H. Samier and V. Sandoval. *La recherche intelligente sur l'internet et l'intranet*. Hermès, septembre 1999.

[12] U. Shardanand, P. Maes Social information filtering: algorithms for automating "word of mouth". Proceedings of CHI'95, ACM Press, 1995.

[13] SITURV. Enquête " ménages-déplacements " de l'agglomération de valenciennoise. Rapport de synthèse, 1997/1998.

[14] M. Virvou, B. Du Boulay. *Human plausible reasoning for intelligent help*. User modeling and user-adapted interaction, vol 9, pp. 321-375, 1999.

[15] A. Waern, C. Averman, M. Tierney, A. Rudström, J. Laaksolahti. *ConCall: an information service for researchers based on EdInfo*, , Report T98:04, oct. 1998.

[16] PREDIT – groupe thématique « Nouveaux services aux usagers. Note stratégique sur l'information multimodale, novembre 1999.