

Action AID

*Gestion des connaissances pour l'aide à la conception
coopérative*

Sophia Antipolis

THÈME 3A



*Rapport
d'Activité*

2001

Table des matières

| | | |
|----------|---|-----------|
| 1 | Composition de l'équipe | 3 |
| 2 | Présentation et objectifs généraux | 3 |
| 3 | Fondements scientifiques | 5 |
| 3.1 | Panorama | 5 |
| 3.2 | Raisonnement à partir de cas | 5 |
| 3.3 | Systèmes de recommandations | 6 |
| 3.4 | Sémantique et conception de systèmes d'informations hypertextes | 8 |
| 4 | Domaines d'applications | 9 |
| 4.1 | Panorama | 9 |
| 4.2 | Recherche collaborative d'informations sur le Web | 9 |
| 4.3 | Vérification statique d'un système d'informations. | 10 |
| 4.4 | Analyse des usages sur internet/intranet | 10 |
| 4.5 | Reconception de sites Web basée sur une analyse de l'usage | 11 |
| 5 | Logiciels | 11 |
| 5.1 | CLF - « Computer Language Factory » | 12 |
| 5.2 | ROL - « Reactive Object Language » | 12 |
| 5.3 | CBR*Tools - Plate-forme objet en raisonnement à partir de cas | 12 |
| 5.4 | Broadway*Tools - Génération de systèmes de recommandations adaptatifs | 13 |
| 5.5 | Broadway-Web - Aide personnalisée à la navigation sur le Web | 14 |
| 5.6 | Hermès - Aide à la confrontation de points de vues sur le Web | 16 |
| 6 | Résultats nouveaux | 17 |
| 6.1 | Panorama | 17 |
| 6.2 | Extraction de patrons de navigation web basée sur une indexation par arbre des suffixes généralisé | 17 |
| 6.3 | Validation de nos outils d'analyse et d'utilisation de l'usage | 18 |
| 6.3.1 | Analyse des logs HTTP | 18 |
| 6.3.2 | Validation de nos plates-formes logicielles | 19 |
| 6.4 | Evaluation du langage ROL pour nos applications Web | 19 |
| 6.5 | Plate-forme pour l'analyse et l'utilisation de comportements visuels et non visuels | 19 |
| 6.6 | Aide à la vérification sémantique de sites Web | 21 |
| 7 | Contrats industriels (nationaux, européens et internationaux) | 21 |
| 7.1 | Contacts industriels | 21 |
| 8 | Actions régionales, nationales et internationales | 22 |
| 8.1 | Actions régionales | 22 |
| 8.2 | Actions nationales | 22 |
| 8.3 | Actions européennes | 22 |

| | | |
|-----------|---|-----------|
| 8.4 | Actions internationales | 23 |
| 9 | Diffusion de résultats | 23 |
| 9.1 | Animation de la communauté scientifique | 23 |
| 9.1.1 | Visites | 23 |
| 9.1.2 | Serveur interne Web | 24 |
| 9.1.3 | Divers | 24 |
| 9.2 | Formation | 24 |
| 9.2.1 | Enseignement universitaire | 24 |
| 9.2.2 | Thèses | 24 |
| 9.2.3 | Stages | 25 |
| 9.3 | Participation à des colloques, séminaires | 25 |
| 10 | Bibliographie | 25 |

1 Composition de l'équipe

Responsable scientifique

Brigitte Trousse [Chargée de recherche, Inria]

Assistante de projet

Sophie Honnorat [AI, Inria, à 1/4-Temps]

Personnel Inria

Thierry Despeyroux [Chargé de recherche, Inria]

Poste d'accueil

Konstantinos Konstantinidis [jusqu'au 15/10/01]

Chercheur doctorant

Mohamed Semi Gaieb [UNSA, boursier 1/2 ATER, jusqu'au 31/08/01]

Doru Tanasa [UNSA, à compter du 01/10/01]

Chercheur post-doctorant

Mohamed Semi Gaieb [UNSA, boursier 1/2 ATER, à compter du 01/09/01]

Chercheur invité

Mireille Arnoux [Université de Bretagne Occidentale, Maître de conférences, à compter du 01/10/01]

Stagiaires

Patrice Carrique [ESSI, UNSA, du 01/07/01 au 30/09/01]

Gilles Dodinet [ESSI, UNSA, jusqu'au 13/04/01]

Ferdinand Dosser [ESSI, UNSA, jusqu'au 13/04/01]

Laurent Perrimond [ESSI, UNSA, jusqu'au 13/04/01]

Cédric Sempe [ESSI, UNSA, jusqu'au 13/04/01]

Doru Tanasa [Ecole Doctorale STIC, UNSA, jusqu'au 30/06/01]

2 Présentation et objectifs généraux

Mots clés : génie logiciel, intelligence artificielle, conception, spécification, hypermédia, sites/services Web, sémantique formelle, adaptation à l'utilisateur, gestion de points de vue, analyse des usages, Web mining, extraction des connaissances à partir de données (ECD), fouille de données, raisonnement à partir de cas, recherche d'informations, navigation, comportements utilisateur, calcul de recommandation, prédiction.

L'action AID a donné naissance le 6 novembre à une nouvelle action AxIS (<http://www.inria.fr/recherche/equipes/axis.fr.html>) : celle-ci poursuivra les recherches menées dans AID depuis 2000 en se focalisant sur un domaine de conception, celui des systèmes d'informations et se verra augmenter de chercheurs issus du thème 4 (ex action ADOPT) sur les aspects "fouille de données, extraction de connaissances à partir de données et classification". Ce rapport d'activité est donc le dernier pour l'action AID.

L'action de recherche AID est une équipe pluri-disciplinaire (Intelligence artificielle et Génie logiciel) qui mène des recherches dans le domaine du Web. Elle vise des méthodes et des outils pour l'aide à l'activité de conception, de réalisation et de maintenance de sites/services Web

évolutifs et adaptatifs. Nos recherches visent à apporter une aide aux deux groupes d'acteurs que sont les *créateurs* et les *utilisateurs* de sites et à permettre d'anticiper dès la conception d'un site les problèmes d'usage et de maintenance (due à des futures évolutions).

Nous avons commencé à appliquer nos recherches à des sites d'une certaine taille (comme des annuaires thématiques/portails, sites institutionnels), où l'intégration de services d'aide à la recherche d'information est pertinente et où des problèmes de maintenance peuvent se poser assez rapidement.

Axes de recherche

Nous menons des recherches sur trois thèmes :

1. Aide à la conception et la réalisation de systèmes de recommandations personnalisées, plus précisément sur les aspects *adaptation à l'utilisateur* et *apprentissage*. Pour cela, nous utilisons des techniques de raisonnement à partir de cas (noté RàPC¹) et plus généralement d'ECD (Extraction de connaissances à partir de données). Ces recherches visent à aider à la recherche collaborative d'informations dans un système d'informations hypermédia (un site Web ou sur le Web par exemple) et/ou à la réutilisation d'expériences passées en conception.
2. Aide à la spécification et à la vérification sémantique de sites Web basées sur des approches formelles en Génie Logiciel (cf. la sémantique naturelle utilisée pour spécifier la sémantique de langages de programmation) et/ou en IA.
3. Aide méthodologique et logicielle pour la gestion de points de vues dans le cadre de projets de conception de sites/services Web (vus comme des systèmes complexes) et la re-conception de tels sites à partir d'une analyse de l'usage du site et/ou du service intégré au site. Ce troisième thème s'appuie sur les résultats des recherches des deux premiers thèmes.

Nos principaux résultats concernent le thème 1 « systèmes de recommandations personnalisées » démarré en 96 pour l'aide à la recherche d'information sur le Web (ou dans un site Web) avec des techniques de raisonnement à partir de cas (cf. notre plate-forme orientée objet en RàPC appelé CBR*Tools pour la réutilisation d'expériences passées, l'approche Broadway avec notre boîte à outils Broadway*Tools).

Coopérations

L'action AID a collaboré avec plusieurs industriels comme le CNES, France Télécom et le centre de recherche européen XRCE (France). De plus, elle participe au niveau national ou international à diverses communautés scientifiques.

¹en anglais *CBR* pour « *Case-Based Reasoning* »

3 Fondements scientifiques

3.1 Panorama

Résumé : *L'action AID mène des recherches sur l'application des techniques d'intelligence artificielle pour l'aide à l'utilisation et à une conception continue de sites Web. Pour cela, nous étudions les apports possibles du raisonnement à partir de cas et de la fouille de données et plus généralement de l'ECD ("Extraction de Connaissances à partir de Données") pour l'élaboration d'une approche de systèmes de recommandations sur le Web. D'autre part nous étudions l'application de méthodes formelles en sémantique des langages de programmation et celles de représentation des connaissances en IA appliquées à la conception et la vérification sémantique de sites Web.*

3.2 Raisonnement à partir de cas

Mots clés : gestion de l'expérience, réutilisation d'expériences passées, indexation, raisonnement à partir de cas case-based reasoning, séquences temporelles.

Glossaire :

Raisonnement à partir de Cas (RàPC). Se dit d'une approche de résolution de problèmes basée sur la réutilisation par analogie d'expériences passées appelées "cas". Un cas est généralement indexé pour permettre de le retrouver suivant certaines caractéristiques pertinentes et discriminantes, appelées "indices" ; ces indices déterminent dans quelle situation (ou contexte) un cas peut être de nouveau réutilisé.

Le raisonnement à partir de cas ^[NMJ⁺97] se décompose habituellement en quatre phases principales ^[AP94,Kol93] :

1. une phase de recherche de cas ayant des similarités (i.e. des indices similaires) avec le problème courant,
2. une phase de réutilisation, permettant de construire une solution au problème courant en se basant sur les cas identifiés dans la phase précédente,
3. une phase de révision de la solution qui permet d'affiner cette dernière grâce à un processus d'évaluation,
4. une phase d'apprentissage, dont le but est de mettre à jour les éléments du raisonnement en prenant en compte l'expérience qui vient d'être réalisée et qui pourra ainsi être utilisée pour les raisonnements futurs.

[NMJ⁺97] A. NAPOLI, A. MILLE, M. JACZYNSKI, B. TROUSSE, ALII, « Aspects du raisonnement à partir de cas », in : *Actes des 6 èmes journées nationales PRC-GDR Intelligence Artificielle*, S. Pesty, P. Siegel (éditeurs), hermes, Paris, p. 261-288, mars 1997.

[AP94] A. AAMODT, E. PLAZA, « Case-Based Reasoning : Foundational Issues, Methodological Variations, and System Approaches », *The European Journal of Artificial Intelligence* 7, 1, 1994, p. 39-59.

[Kol93] J. KOLODNER, *Case-Based Reasoning*, Morgan Kaufmann Publishers, 1993.

Les problèmes difficiles en RàPC sont très fréquemment liés : à la définition et la représentation d'un cas, l'organisation de la base de cas, les diverses indexations utilisées et la définition de « bonnes » mesures de similarités pour la recherche de cas, le lien recherche-adaptation de cas (le meilleur cas étant le cas le plus facilement adaptable), la définition d'une stratégie d'adaptation à partir du (ou des) cas retrouvé(s), l'apprentissage de nouveaux indices, etc.

Nous poursuivons l'évaluation de nos résultats en RàPC en particulier notre modèle d'indexation par situation comportementale, notre plate-forme orientée objet CBR*Tools et boîte à outils Broadway*Tools. De plus, nous étudions plus particulièrement des techniques d'indexation des sessions et algorithmes de recherche de sous-séquences fréquentes d'items pour l'analyse en ligne et hors ligne de l'usage des internautes.

3.3 Systèmes de recommandations

Mots clés : Web, hypermedia, fouille de données, extraction de connaissances à partir de données, KDD, raisonnement à partir de cas, CBR filtrage collaboratif, calcul de recommandations.

L'objectif d'un système de recommandations est d'aider les utilisateurs à faire leurs choix dans un domaine où ils disposent de peu d'informations pour trier et évaluer les alternatives possibles [SM95,RV97,KMM⁺97].

Un système de recommandations peut être décomposé en trois entités de base (cf figure 1) : le groupe d'agents *producteurs* de recommandations, le module de *calcul de recommandations* et le groupe de *consommateurs* des recommandations.

Un défi majeur dans le domaine de la conception de systèmes de recommandations est le suivant :

Comment produire des recommandations *personnalisées* et de haute *qualité* tout en *minimisant l'effort* requis de la part des producteurs et des consommateurs.

Deux grandes approches complémentaires sont proposées dans la littérature : 1) l'approche basée sur le contenu et fondée sur l'apprentissage automatique de profils utilisateurs et 2) l'approche dite de filtrage collaboratif fondée sur des techniques de fouille de données. Le profil utilisateur est une structure de données qui décrit les centres d'intérêts d'un utilisateur dans l'espace des objets à recommander. Celui-ci est une structure construite dans la première approche ou donnée dans la seconde par l'utilisateur . ce profil est utilisée soit pour filtrer les objets disponibles (on parle alors de filtrage basé sur le contenu), soit pour recommander à l'utilisateur ce qui a satisfait d'autres utilisateurs ayant un profil similaire (on parle alors de filtrage collaboratif) [RV97].

-
- [SM95] U. SHARDANAND, P. MAES, « Social Information Filtering: Algorithms for Automating Word of mouth », *in: CHI'95: Mosaic of creativity*, ACM, p. 210-217, Denver, Colorado, mai 1995.
- [RV97] P. RESNICK, H. R. VARIAN, « Recommender systems », *Communications of the ACM* 40, 3, 1997, p. 56-58.
- [KMM⁺97] J. A. KONSTANT, B. N. MILLER, D. MALTZ, J. L. HERLOCKER, L. R. GORDON, J. RIEDL, « GroupLens: Applying collaborative filtering to usenet news », *Communications of the ACM* 40, 3, 1997, p. 77-87.

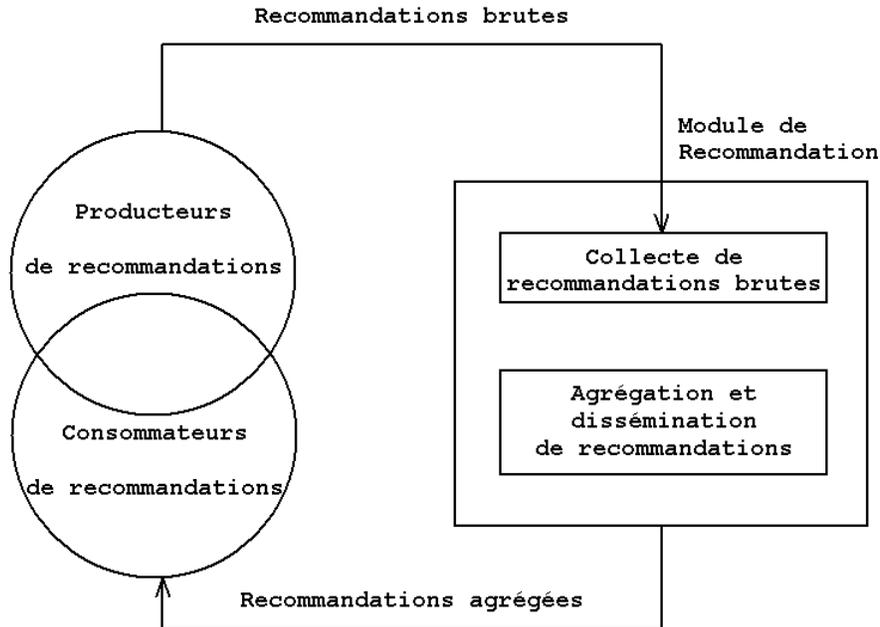


FIG. 1 – Architecture générale d'un système de recommandation

Dans l'action AID, nous poursuivons le développement d'une approche hybride de calcul de recommandations basée sur l'analyse du contenu visité et centrée fouille de données où les comportements passés d'un **groupe** d'utilisateurs sont utilisés pour calculer les recommandations (cf. filtrage collaboratif). La plupart des autres approches fondées sur la fouille de données sont principalement des approches statistiques où l'ordre d'occurrence d'événements dans l'historique n'est pas pris en compte lors du calcul de recommandations. Citons comme exemple, dans le domaine d'aide à la navigation sur le Web, le système FootPrints ^[WM97] et le système de Yan et al ^[YJGMD96].

Les problèmes difficiles pour la mise en œuvre de notre approche concernent les aspects suivants :

1. fournir des techniques d'identification et d'extraction de comportements pertinents (i.e. des enseignements ou des cas) à partir des données brutes des historiques,
2. définir des méthodes et des techniques de mesure de similarités entre comportements,
3. définir des techniques d'inférence de recommandations personnalisées à partir des comportements pertinents passés identifiés (ou à partir des cas remémorés).

[WM97] A. WEXELBLAT, P. MAES, « Footprints: Visualizing histories for web browsing », *in : Proceedings of RIAO'97, Computer Assisted Information Retrieval on the Internet*, Montreal, 1997.

[YJGMD96] T. YAN, M. JACOBSEN, H. GARCIA-MOLINA, U. DAYAL, « From user access patterns to dynamic hypertext linking », *Computer Network and ISDN systems 28*, mai 1996, p. 1007-1014, (proceedings of the 5th international WWW conference).

Nous étudions l'ensemble des trois problèmes ci-dessus en explorant la possibilité d'application des techniques RàPC et plus généralement de l'ECD.

Nous étudions la classe de systèmes de recommandations s'appuyant sur une **réutilisation d'expériences passées** issues d'un groupe d'utilisateurs utilisant des techniques de raisonnement à partir de cas (RàPC). Dans cette classe de systèmes, nous privilégions les deux types de systèmes suivants :

- ceux dont le calcul de recommandations se base sur une réutilisation d'expériences d'un groupe d'utilisateurs **recherchant de l'information** dans un système d'informations hypertexte comme le Web ou dans un site internet/intranet. Ces systèmes visent une **aide personnalisée** à l'activité de recherche d'informations ;
- ceux dont le calcul de recommandations s'appuie sur une **réutilisation d'expériences passées issues d'experts** en vue de fournir un aide à des activités de **conception**.

3.4 Sémantique et conception de systèmes d'informations hypertextes

Mots clés : Web sémantique, typage, documents semi-structurés, spécification, maintenance, évolution, sites Web, sémantique formelle, génie logiciel.

Concevoir et maintenir un système d'informations hypertexte comme un site Web est une tâche difficile. Il est beaucoup plus facile de trouver des informations inconsistantes qu'un site bien maintenu sur Internet. Notre but est d'étudier et de construire les outils nécessaires à la conception, à la production et à la maintenance de sites complexes et cohérents avec une approche pluri-diciplinaire (GL et IA).

Il existe un très fort parallèle entre un document structuré (tel qu'un site Web) et un programme, et le monde du Web est un très bon utilisateur d'idées développées il y a plusieurs années dans le monde du génie logiciel : la notion de syntaxe abstraite se retrouve dans un langage tel que XML et les DTD de même que l'idée de séparation entre structure et présentation concrète.

Jusqu'à présent, le monde du Web s'est principalement intéressé à la présentation des pages (HTML, CSS, XSL) et à la structure syntaxique du contenu des pages (XML), mais très peu à la sémantique des sites. Notons cependant les efforts du consortium W3C autour du "Web sémantique" (XML, RDF et "RDF schema") ainsi que certains travaux de recherche issus de l'IA basés sur une approche ontologique comme WebMaster^[HM99]. Notre approche diffère dans le fait que nous voulons exploiter plus loin le parallèle entre programmes et sites Web pour mieux aborder la sémantique formelle des sites.

Il existe (au moins) deux sens au terme "sémantique" : ce peut être l'étude scientifique du sens des unités linguistiques, mais aussi l'étude de propositions d'une théorie déductive du point de vue de leur vérité ou de leur fausseté. C'est à cette dernière définition que nous voulons nous référer en formalisant une partie du contenu des systèmes d'information.

Nous pouvons d'ores et déjà distinguer les "aspects statiques" d'un site qui peut être vu

[HM99] F. V. HARMELEN, J. V. D. MEER, « WebMaster: Knowledge-based Verification of Web-pages », in: *Twelfth International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems IEA/AIE'99*, 1999.

comme un ensemble de contraintes globales (pas seulement syntaxiques, mais aussi sémantiques et dépendantes du contexte) qui doivent être vérifiées et la “aspects dynamiques” qui prend en compte la navigation d’un utilisateur dans un site et rejoint donc l’analyse des usages. Pour les aspects dynamiques, nous pensons à moyen terme, formaliser les notions de qualité et de fiabilité d’un site en faisant un parallèle avec la notion de preuve de programme.

4 Domaines d’applications

4.1 Panorama

Mots clés : Web, multimédia, commerce électronique, e-marketing, e-CRM, santé, ingénierie, éducation, environnement, télécommunications, conception de sites Web, conception de services adaptatifs, recherche d’information, analyse des usages, analyse des logs, personnalisation.

Résumé :

Les domaines d’application visés par le projet concernent les systèmes d’informations hypermédia soit sur le Web [Tro99] (sites de commerce électronique, sites institutionnels, sites éducatifs, sites pour la santé ou l’environnement par exemple) soit en intranet (santé, etc.). Nous visons particulièrement des services d’aide à la recherche collaborative d’informations sur le Web, à la fois du point de vue utilisateurs (cf. section 4.2), et du point de vue concepteurs de sites Web (cf. section 4.3, 4.4, 4.5).

Cependant nos recherches en RàPC peuvent être appliquées dans de nombreux domaines pour des problèmes d’aide à la recherche et à la réutilisation d’expériences passées, par exemple en conception, diagnostic, etc. ; ces recherches visent particulièrement des domaines d’applications ayant des problèmes nécessitant une prise en compte d’historiques et donc une indexation de cas par des situations comportementales, comme en supervision et contrôle de systèmes dynamiques (cf environnement, santé) ou en conception de systèmes d’assistance.

4.2 Recherche collaborative d’informations sur le Web

Mots clés : réutilisation, expérience, groupe d’utilisateurs, filtrage collaboratif, analyse de comportements utilisateurs.

Le Web est un hypermédia de documents hétérogènes et dynamiques. Cet espace virtuel, qui est en expansion croissante, offre une énorme quantité de données. Deux types de méthodes permettent de retrouver des documents pertinents dans cet espace : l’interrogation et la navigation. L’interrogation est appropriée lorsque l’utilisateur a une idée précise de ce qu’il cherche, ou bien peut être utilisée comme point de départ pour une session de navigation.

[Tro99] B. TROUSSE, « Présentation des travaux de recherche relatifs à l’IA et Internet de l’Action AID », juillet 1999, In *Dossier IA et Internet - S. Stinckich - Bulletin de l’AFIA - numero 38.*

Cependant, la complexité et l'étendue du Web rendent, d'une part, difficile l'indexation des documents nécessaire pour une recherche par interrogation, et peuvent, d'autre part, entraîner la désorientation de l'utilisateur lors d'une navigation. L'assistance des utilisateurs dans leur recherche d'information sur le Web est donc un problème important : notre approche vise une recherche d'informations collaborative (basée sur des techniques RàPC), où l'expérience d'un groupe de personnes est mise à profit.

Ce domaine d'application a constitué encore notre investissement majeur avec nos travaux autour de Broadway. Notre partenaire principal sur ce sujet est France Télécom (R&D).

4.3 Vérification statique d'un système d'informations.

Mots clés : sémantique formelle, sémantique naturelle, contraintes globales, spécifications, vérification statique.

Avec le développement fulgurant de l'Internet et l'utilisation de ses standards comme mode de diffusion et de représentation de données, la conception des sites Web devient une activité de plus en plus importante. Or, si les outils qui existent actuellement permettent de structurer syntaxiquement l'information, il n'y a pas à l'heure actuelle d'outils de vérification de la cohérence sémantique de cette information.

La notion de vérification sémantique est fondamentale, car c'est elle qui peut assurer la qualité et la cohérence d'un site à travers le temps en particulier lors des mises à jour et des refontes des sites. Cette vérification passe par une phase déclarative dans laquelle des règles globales sont données. Ces règles permettent par la suite de s'assurer que, quand un site est modifié, il vérifie toujours les règles imposées.

Le besoin de cohérence ne concerne pas uniquement le contenu des systèmes d'informations mais aussi les diverses méta-données qui peuvent lui être attachées. On peut considérer que le fait de présenter une même information sous différentes formes (à destination humaine pour le contenu et à destination machine pour certaines annotations) ne fait qu'augmenter le risque d'incohérence, et que donc des méthodes de contrôle de la cohérence doivent être développées.

4.4 Analyse des usages sur internet/intranet

Mots clés : Web mining, analyse des usages, comportements utilisateur, prédiction, e-CRM, e-marketing, commerce électronique, réutilisation des usages.

La motivation de nos recherches sur l'analyse des usages sur internet/intranet s'inscrit dans une approche globale de la conception visant à anticiper les usages. En effet, peu d'aide en termes d'analyse de l'usage est offert aux concepteurs de sites pour anticiper les usages.

Les méthodes (off-line) d'analyse des fichiers logs HTTP proposées actuellement sur le marché n'adoptent pas une approche cognitive de la notion de *session utilisateur* ni ne prennent en compte le *degré de satisfaction des utilisateurs*. Notre méthode d'analyse, que nous avons appliquée à plusieurs reprises, dépasse ces limites et propose de caractériser les sessions utilisateurs en fonction de critères objectifs à acquérir auprès des concepteurs relativement à une notion de réussite ou d'échec. Nous avons actuellement un savoir-faire dans l'application de

notre méthode (off-line ou on-line) pour des sessions de type navigation et/ou formulation de requêtes.

D'autre part, notre modèle d'indexation de situations comportementales utilisé dans nos systèmes de recommandations nous permet, dans un site donné, d'extraire - soit on-line soit off-line à partir des logs - des comportements jugés intéressants par l'analyste via l'utilisation d'un patron puis de les indexer.

4.5 Reconception de sites Web basée sur une analyse de l'usage

Mots clés : conception continue, analyse des usages, re-conception, ingénierie concourante, points de vues, négociation.

L'étude des sessions utilisateurs à partir des enregistrements fournis par les fichiers de logs des serveurs ainsi que l'analyse de l'usage d'un système de recommandations à la Broadway sont de bonnes sources de réflexion quant à la refonte d'un site Web. Elles permettent, en effet, dans une certaine mesure, de quantifier les difficultés rencontrées par l'utilisateur ou l'évolution éventuelle des usages et peuvent donc suggérer des changements dans la structure des sites, changements qui peuvent ou non modifier la sémantique (comme l'ajout d'une nouvelle contrainte globale). Analyse de l'usage et vérification sémantique sont donc deux ingrédients nécessaires de la reconception d'un site Web.

Enfin, au niveau méthodologie de conception, nous souhaitons appliquer nos travaux théoriques sur les points de vues (cf contrat CNES 95-97) au domaine de la conception de sites Web² et utiliser éventuellement notre système Hermès pour supporter la confrontation de points de vues propres à cette conception (cf. contrats CNES et CNRS-GIS). Il s'agit pour nous d'aider à la structuration des principaux critères d'évaluation utilisés : facilité d'utilisation, aspect économique, facilité d'évolution du site, etc.

5 Logiciels

Mots clés : informatique décisionnelle, service Web, personnalisation, gestion de l'expérience, raisonnement à partir de cas, plate-forme orientée objet, réutilisation, UML, patron de conception, système de recommandations, filtrage collaboratif, coopération, argumentation, Web, langages de spécification, sémantique formelle, CLF, CBR*Tools, Broadway*Tools, ROL, programmation parallèle asynchrone, objets réactifs, bus logiciel, Java, Prolog.

Les logiciels de l'équipe sont conçus pour la plupart avec l'atelier de conception objet *Rational Rose* et réalisés avec le langage de programmation Java. Ces logiciels sont décrits à l'adresse suivante : <http://www-sop.inria.fr/aid/software.html>

²Nous nous intéressons actuellement principalement aux sites Web de type répertoires thématiques, portails (cf. Voilà) ou catalogues (commerce électronique) et aux sites institutionnels.

5.1 CLF - « Computer Language Factory »

Mots clés : langages de spécification, sémantique, Centaur, Prolog.

Participant : Thierry Despeyroux [correspondant].

CLF, développé par Thierry Despeyroux (dans l'ex-projet CROAP) est un ensemble d'outils et de formalismes de spécification de la syntaxe et la sémantique de langages informatisés. CLF propose actuellement les langages AS ^[Des96] (Abstract Syntax) et CS (Concrete Syntax).

Une partie de CLF a été directement adaptée au monde Prolog comme une extension des DCG ("Definite Clause Grammars"). Par rapport aux DCG, notons une plus grande facilité d'expression due à la possibilité de récursions gauches et à la compilation de certaines règles de grammaire permettant d'utiliser certaines optimisations très importantes des compilateurs Prolog (indexation de clauses). De plus l'analyseur ainsi réalisé contient une méthode générique permettant de faire le lien entre les occurrences du terme construit et les positions textuelles dans le texte source, une caractéristique qui manque à la plupart des générateurs d'analyseurs lexicaux.

Cette extension des DCG a été ensuite utilisée pour construire un analyseur XML utilisé pour nos travaux sur la vérification sémantique de systèmes d'informations écrits en XML (cf. section 6.6).

5.2 ROL - « Reactive Object Language »

Mots clés : programmation parallèle asynchrone, langage de spécification, objets réactifs, bus logiciel.

Participants : Semi Gaieb [correspondant], Thierry Despeyroux.

ROL [12], développé par Sémi Gaieb dans le cadre de son travail de thèse, est un langage de spécifications d'objets communiquant par l'intermédiaire d'un bus logiciel. ROL permet de décrire clairement les interactions entre des objets, indépendamment du langage dans lequel ils sont écrits, et indépendamment de la configuration de l'application.

ROL a été utilisé pour la spécification du logiciel Hermes ainsi que pour un système d'aide à la navigation dans un répertoire thématique (cf educaid FT & INRIA).

5.3 CBR*Tools - Plate-forme objet en raisonnement à partir de cas

Mots clés : raisonnement à partir de cas, plate-forme objet, composants logiciels, réutilisation, UML, patron de conception.

Participant : Brigitte Trousse [correspondante].

Résumé : *CBR*Tools est une plate-forme à objets développée dans l'équipe depuis 97 pour faciliter le développement d'applications nécessitant des techniques de raisonnement à partir de cas.*

[Des96] T. DESPEYROUX, « AS, for Abstract Syntax - Manual - V1.0 », *rapport de recherche n°197*, Inria, septembre 1996.

CBR*Tools [Jac98] [JT99] est une plate-forme à objets (ou « object-oriented framework » [JF88,FS97]) en RàPC, qui offre un ensemble de classes abstraites modélisant les principaux concepts nécessaires pour développer une application intégrant des techniques de raisonnement à partir de cas : cas, bases de cas, index, mesures de similarité, contrôle du raisonnement. Elle offre également un ensemble de classes concrètes qui implantent de nombreuses méthodes classiques (indexation par plus proches voisins, indexation par Kd-tree [WAD94], indexation par prototypes [JT94], indexation basée sur une approche neuronale, mesures de similarités standards). CBR*Tools comporte actuellement plus de 200 classes avec notamment deux grands groupes : le package *core* pour le fonctionnement de base et le package *time* pour la gestion spécifique des situations comportementales. La programmation d'une nouvelle application se fait par spécialisation de classes existantes, par composition d'objets ou en utilisant les paramètres des classes existantes.

CBR*Tools vise tout particulièrement des domaines d'application nécessitant une réutilisation de cas devant être indexés par des situations comportementales.

CBR*tools a été installée à France télécom (R&D) à Lannion en 1998 et 2000 dans le cadre de Broadway-Web et educaid (FT-CTI) et a été utilisée dans le cadre d'un contrat XRCE-INRIA (98). Une documentation sur le Web est accessible à l'adresse suivante : <http://www-sop.inria.fr/aid/cbrtools/manual/>.

La plate-forme CBR*Tools a été évaluée via la conception et la réalisation de quatre applications (Broadway-Web, educaid, BeCKB, Broadway-Predict). Nous avons montré que, pour chaque application, l'expertise approfondie nécessaire pour utiliser CBR*Tools ne concerne que 20% à 40% des points d'ouverture validant ainsi l'aide apportée par notre plate-forme tant sur la modélisation que sur l'implantation, grâce à la réutilisation de son architecture abstraite et de ses composants (index, similarité).

5.4 Broadway*Tools - Génération de systèmes de recommandations adaptatifs

Mots clés : agent personnalisé, boîte à outils, système distribué, programmation

-
- [Jac98] M. JACZYNSKI, *Modèle et plate-forme à objets pour l'indexation des cas par situation comportementale: application à l'assistance à la navigation sur le Web*, thèse de doctorat, université de Nice-Sophia Antipolis, Sophia-Antipolis, décembre 1998.
- [JT99] M. JACZYNSKI, B. TROUSSE, « Patrons de conception dans la modélisation d'une plate-forme pour le raisonnement à partir de cas », *Revue l'Objet* 5, 2, 1999, Numéro Spécial sur les patterns orientés objets, D. Rieu et J-P. Giraudon (guest editors).
- [JF88] R. JOHNSON, B. FOOTE, « Designing Reusable Classes », *Journal of Object-oriented programming* 1, 2, 1988, p. 22-35.
- [FS97] M. E. FAYAD, D. C. SCHMIDT, « Object-Oriented Application Frameworks », *Communication of the ACM* 40, 10, 1997, p. 32-38.
- [WAD94] S. WESS, K. ALTHOFF, G. DERWAND, « Using K-d Trees to Improve the Retrieval Step in Case-Based Reasoning », in: *Lecture Notes in Artificial Intelligence, Topics in Case-Based Reasoning*, S. Wess, K. Althoff, M. M. Richter (éditeurs), Springer-Verlag, p. 167-181, 1994.
- [JT94] M. JACZYNSKI, B. TROUSSE, « Fuzzy Logic for the Retrieval Step of a Case-Based Reasoner », in: *Second European Workshop on Case-Based Reasoning (EWCBR'94)*, p. 313-320, Chantilly, 1994.

asynchrone, composants logiciels, réutilisation, services adaptatif, personnalisation.

Participants : Brigitte Trousse [correspondante], Semi Gaieb.

Résumé : *Broadway*Tools est une boîte à outils pour faciliter la réalisation de systèmes de recommandations Web adaptatifs pour l'aide à la recherche d'informations sur le Web ou dans un site internet/intranet. Cette boîte à outils offre actuellement différents serveurs dont un serveur de calcul de recommandations basé sur notre modèle d'indexation comportementale pour l'observation des sessions utilisateurs et sur la réutilisation de sessions passées d'un groupe d'utilisateur. Un système de recommandations réalisé avec Broadway*tools observe les navigations de différents utilisateurs et récolte les évaluations et les annotations de ces utilisateurs pour établir une liste de recommandations pertinentes (documents Web, mots clés, etc.).*

Cette boîte à outils a été utilisée pour deux systèmes d'aide à la navigation : Broadway-Web (ex Broadway-V1) pour l'aide à la navigation sur le Web et Broadway-educaid pour l'aide à la navigation dans un clone d'un site de France Télécom.

5.5 Broadway-Web - Aide personnalisée à la navigation sur le Web

Mots clés : filtrage collaboratif, système de recommandations, aide à la navigation, Web, analyse réutilisation, comportements utilisateurs, profils utilisateurs.

Participants : Brigitte Trousse [correspondante], Ferdinand Dosser, Sémi Gaieb, Laurend Perrimond.

Résumé : *Broadway-Web³ [MT99] est un assistant pour la navigation sur le Web réutilisant les navigations passées d'un groupe d'utilisateurs. Broadway-Web observe les navigations de différents utilisateurs et récolte les évaluations et les annotations de ces utilisateurs pour établir une liste de documents pertinents.*

Broadway-Web (<http://www-sop.inria.fr/aid/broadway/>), issu de la thèse de M. Jaczynski (1998), est un serveur HTTP utilisé comme proxy : il est inséré entre le navigateur et le reste du web et il intercepte ainsi toutes les demandes de documents pour le protocole HTTP. Broadway est alors capable d'observer les différentes navigations des utilisateurs en enregistrant notamment : les adresses des documents visités, un ensemble de mots clefs issus de l'analyse automatique des pages HTML et les évaluations des documents. Durant une navigation, Broadway (cf. Fig 2) peut afficher un ensemble de documents qu'il conseille suivant l'état courant de la navigation, et permet aux utilisateurs d'évaluer ou d'annoter les documents traversés grâce à une barre d'outils insérée dynamiquement dans les pages HTML visualisées.

³Brodway – « BROwsing ADvisor reusing pathWAYS »

[MT99] M.JACZYNSKI, B. TROUSSE, « Broadway : a Case-based System for Cooperative Information Browsing on the World-Wide-Web », in : *Collaboration between Human and artificial Societies. Coordination and Agent-based distributed Computing*, J. A. Padget (éditeur), LNAI Series, p. 264–283, 1999.

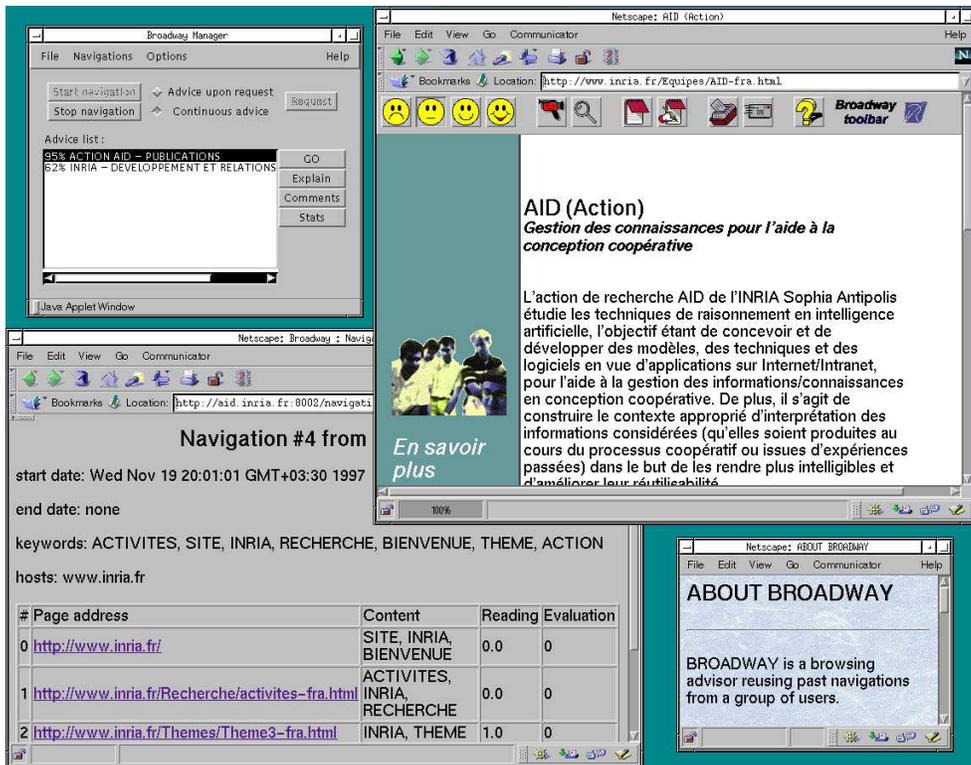


FIG. 2 – Utilisation de Broadway-Web

Broadway intègre le serveur HTTP Jigsaw du W3C programmé en Java et utilise notre plateforme CBR*Tools pour implanter le système de raisonnement à partir de cas permettant la réutilisation des navigations passées.

Broadway-Web (ex Broadway-V1) a été installée à France Télécom (R&D) à Lannion en 98 et a fait l'objet d'une démonstration aux huitièmes rencontres INRIA-Industrie *Informatique pour les Télécommunications et le Multimédia* le 26 novembre 1998. ainsi qu'aux rencontres Inria/Industrie sur la recherche d'informations organisées en décembre 2001.

5.6 Hermès - Aide à la confrontation de points de vues sur le Web

Mots clés : discussion, Web, négociation, coopération, argumentation.

Participants : Semi Gaieb, Brigitte Trousse [correspondante].

Résumé : *Hermès est un système d'aide à la structuration d'une discussion, pouvant être utilisé à la prise de décision collective sur le Web.*

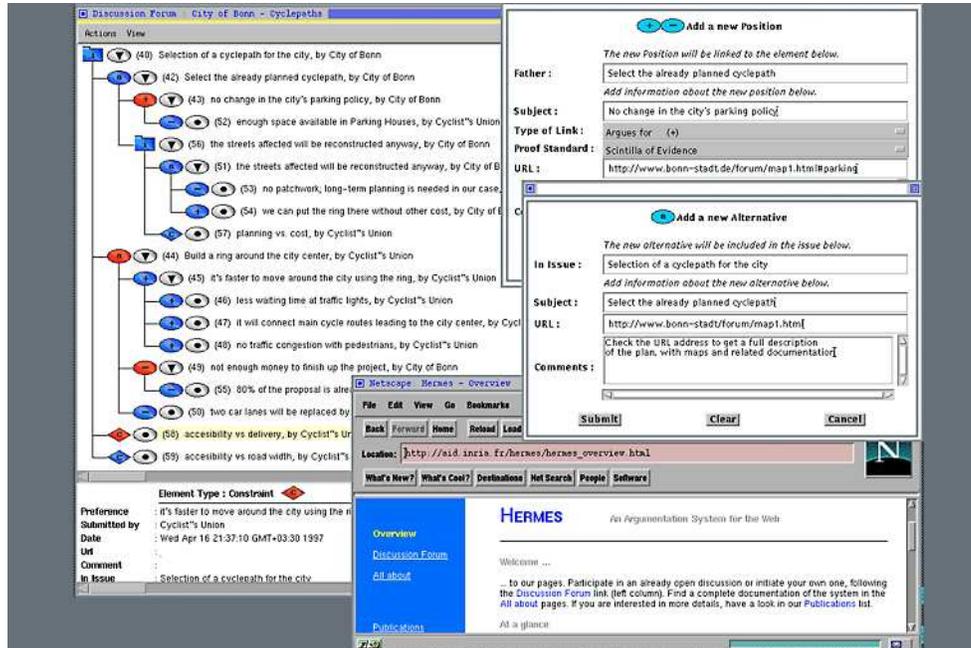


FIG. 3 – Utilisation d'Hermès

L'objectif d'Hermès ^[KTP97,Kar98,KP98] (<http://www-sop.inria.fr/aid/hermes/>) est d'apporter une aide plus puissante que celle des systèmes existants qui proposent uniquement des forums de discussions hiérarchiques.

Pour cela, le système Hermès (cf. Fig 3) offre des mécanismes de raisonnement qui contrôlent la discussion. De plus, il permet d'exprimer différents arguments individuels (ou préférences) et de les noter dans le but d'aider à la sélection d'un certain choix. Hermès vise des coopérations sur le Web entre agents humains travaillant ensemble dans des lieux différents et à un même instant.

Hermès a fait l'objet de recherche à l'EPFL ainsi qu'à University of Science and Technology - Hong Kong.

6 Résultats nouveaux

6.1 Panorama

Résumé : *Nos principaux résultats sont 1) un algorithme de recherche de sous-séquences fréquentes basé sur une indexation de ces séquences par un arbre des suffixes généralisé 2) une validation de nos outils d'analyse (cf. logaudience) sur des logs HTTP et de Broadway*Tools avec la réécriture de Broadway-Web, 3) les travaux de thèse de S. Gaieb autour du langage ROL, 4) la spécification d'une plateforme d'expérimentation d'analyse et d'utilisation de comportements visuels et non visuels des internautes et enfin 5) la conception et la réalisation d'outils basés sur Prolog (parseurs XML, etc.) pour l'aide à la vérification sémantique de sites Web.*

6.2 Extraction de patrons de navigation web basée sur une indexation par arbre des suffixes généralisé

Mots clés : comportement utilisateur, analyse des usages, sous-séquences fréquentes d'items, navigation.

Participants : Doru Tanasa, Brigitte Trousse.

Dans le cadre d'un stage de DEA [20], nous avons effectué des recherches sur l'indexation de séquences temporelles en utilisant un arbre des suffixes généralisé. Les séquences d'événements ont de nombreuses applications. Citons les alarmes dans un réseau de télécommunication, les données cliniques, les valeurs des actions sur le marché boursier, les sessions des utilisateurs

-
- [KTP97] N. KARACAPILIDIS, B. TROUSSE, D. PAPADIAS, « Using Case-Based Reasoning for Argumentation with Multiple Viewpoints », *in: Case-Based Reasoning Research and Development, Proceedings of the 2nd Int. Conference on Case-Based Reasoning (ICCBR-97), Providence, Rhode Island, July 25-27, 1997*, D. Leake, E. Plaza (éditeurs), Lecture Notes in AI 1266, Springer-Verlag, Berlin, p. 541-552, 1997, <http://www-sop.inria.fr/aid/papers/97iccb.html>.
- [Kar98] N. KARACAPILIDIS, « Hermes - Supporting Argumentative Discourse in Multi-Agent Decision Making », *ERCIM News*, 32, 1998.
- [KP98] N. KARACAPILIDIS, D. PAPADIAS, « Hermes: Supporting Argumentation Discourse in Multi-Agent Decision », *in: Proceedings of the AAAI-98 Conference*, Madison WI, AAAI/MIT Press, p. 827-832, juillet 1998.

d'un logiciel. Généralement on obtient ces séquences par l'observation des valeurs de paramètres d'un processus pendant une période de temps donnée.

Nos recherches sur l'indexation de séquences temporelles s'intègrent dans nos celles sur les systèmes du type Broadway. Pour une session de navigation d'un utilisateur avec le système Broadway-Web par exemple, nous avons quatre séquences temporelles (URL, contenu, évaluations explicite et implicite). L'objectif de notre travail est de permettre une recherche rapide de sessions similaires en termes de sous-séquences fréquentes. Nous avons choisi d'indexer ces sessions avec un arbre des suffixes généralisé qui permet la recherche d'un patron en $O(n)$ où n est la longueur du patron recherché, structure principalement utilisée pour l'indexation des chaînes de textes et des séquences génétiques [Wei73,Gus97].

Nous avons implémenté un algorithme (appelé APRIORI-GST) basé sur notre méthode d'indexation de sessions, n pour l'extraction des patrons fréquents (sous-séquences fréquentes d'items). La structure de base de l'algorithme est similaire à celui d'Apriori [AS95]. L'originalité de notre approche consiste dans l'utilisation de cet index (GST) pour l'indexation des sessions de navigation Web. L'avantage de cet index GST dans ce cas est double : 1) l'incrémentalité qu'il apporte (on ne doit pas refaire l'index chaque fois que la base de données des sessions change) et 2) lors de l'extraction des patrons (par l'algorithme APRIORI-GST) on utilise l'index pour calculer le support d'une séquence. Nous avons présenté nos principaux résultats en [15, 19].

6.3 Validation de nos outils d'analyse et d'utilisation de l'usage

Participants : Mireille Arnoux, Patrice Carrique, Gilles Dodinet, Ferdinand Dossier, Nathalie Evan, Sémi Gaieb, Laurend Perrimond, Cédric Sempe, Doru Tanasa, Brigitte Trousse.

6.3.1 Analyse des logs HTTP

Nous avons poursuivi le développement de notre boîte à outils **logaudience** par capitalisation des diverses modélisations des sessions des internautes naviguant sur le Web ou dans un site Web réalisées ces deux dernières années (contrat FT CTI, prédiction du comportement utilisateur, etc.) lors de la conception de systèmes de recommandations. Nous avons mené une expérimentation avec **logaudience** et notre algorithme APRIORI-GST sur quatre jours de logs HTTP du site inria-sop. Cette expérimentation a permis de valider nos outils et algorithme [15] et aussi de montrer l'importance de la modélisation des informations du site corrélées avec les logs.

-
- [Wei73] P. WEINER, « Linear pattern matching algorithms », *in: 14th IEEE Symp. on Switching and Automata Theory*, p. 1-11, 1973.
- [Gus97] D. GUSFIELD, *Algorithms on Strings, Trees, and Sequences*, Cambridge University Press, 1997.
- [AS95] R. AGRAWAL, R. SRIKANT, « Mining Sequential Patterns », *in: Proc. 11th Int. Conf. Data Engineering, ICDE*, P. S. Yu, A. L. P. Chen (éditeurs), IEEE Press, p. 3-14, 6-10 1995, <http://citeseer.nj.nec.com/agrawal95mining.html>.

6.3.2 Validation de nos plates-formes logicielles

Cette année, nous avons initié la reconception de Broadway*Tools par capitalisation issue de nos applications Web (nouvelles interfaces et classes abstraites en Java) et réécrit Broadway-Web dans cette nouvelle version [17].

6.4 Evaluation du langage ROL pour nos applications Web

Mots clés : annuaire thématique, raisonnement à partir de cas, programmation asynchrone, langage de spécifications, objets réactifs, bus logiciel, Web, décision, point de vue, confrontation, négociation, forum de discussion, argumentation, conception.

Participants : Thierry Despeyroux, Semi Gaieb, Brigitte Trousse.

La validation de ROL sur *educaid* s'est terminée en début d'année 2001 et les travaux de thèse de S. Gaieb ont été soutenus en juillet 2001 dont voici le résumé :

Les logiciels font de plus en plus souvent appel au parallélisme et aux composants distribués. Ces notions ont été petit à petit introduites dans les langages de programmation (ADA, Java) ou donné lieu au développement de langages spécialisés et de *middlewares*. Cependant, il existe peu d'outils de conception de haut niveau pour de tels programmes. Dans cette thèse, nous proposons un modèle de communication basée sur les bus logiciels à messages et un langage de spécification des interactions entre objets de type réactifs asynchrones distribués nommé ROL (*Reactive Object Language*). Ce langage permet la spécification du réseau que forment ces objets et leurs réactions aux messages qu'ils reçoivent. Les spécifications ROL sont totalement indépendantes de l'implémentation. Pour leur exécution, nous proposons une traduction vers le langage objet Java. Actuellement, la partie configuration de l'application est laissée à la charge du programmeur. En spécifiant deux applications Web parallèles asynchrones et distribuées existantes (Hermès et Educaid) avec notre langage ROL, nous montrons que sa syntaxe déclarative permet de simplifier la programmation et la maintenance de systèmes composés d'objets réactifs par rapport à une approche basée sur un langage à objets traditionnel.

6.5 Plate-forme pour l'analyse et l'utilisation de comportements visuels et non visuels

Mots clés : eye-tracking, analyse, comportements, navigation.

Participants : Mireille Arnoux, Patrice Carrique, Nathalie Evan, Sémi Gaieb, Brigitte Trousse.

Le but du projet e-Behaviour est de proposer une plate-forme pour l'analyse et l'utilisation des comportements visuels et non visuels des internautes dans un système d'aide à la navigation. L'originalité de ce projet est la prise en compte par le système de recommandations, du parcours visuel des utilisateurs sur l'écran. La « saisie » par un système d'« Eye-Tracking » et l'analyse des données relatives à ce parcours se font en collaboration avec le LPEQ/UNSA.

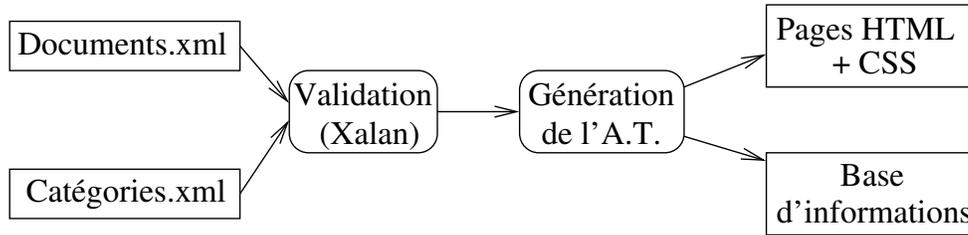


FIG. 4 – Génération d'un annuaire thématique intégrant un système de recommandations

La plate-forme se compose de deux parties : (1) une partie formée de pages HTML (ou XML) et (2) une partie formée de serveurs et de servlets.

1. Après validation (avec xalan) de deux fichiers XML contenant les informations sur la structure de l'annuaire thématique (catégories) et sur les documents qu'il recense (URL, titre, auteurs, description, ...), on génère automatiquement deux composantes dépendantes de l'annuaire thématique choisi :

- (a) les pages HTML de l'annuaire thématique. Pour garder une marge de manœuvre dans la configuration de l'affichage de ces pages, elle utiliseront deux fichiers CSS (un pour les catégories nœuds et un pour les catégories feuilles). Les pages HTML contiendront des liens permettant de naviguer dans l'annuaire thématique tout en transmettant au reste du système des informations relatives à ces navigations grâce à des servlets.
- (b) un serveur d'information contenant toutes les informations concernant les documents. L'utilité de ce serveur est justifiée par la simplification des paramètres des servlets et des serveurs (cf partie suivante).

2. La deuxième partie est composée de serveurs et de Servlets communiquant via un serveur d'événements auxquels ils s'abonnent. Chaque clic sur un lien de l'annuaire thématique permet de naviguer dans l'annuaire thématique tout en transmettant « l'information du clic » (i.e. page pointée par le lien) au recommandeur. Celui-ci met à jour sa « mémoire » et calcule une liste de recommandations. Leur calcul fait en partie appel au serveur d'informations qui utilise la base d'informations générée (Cf. partie précédente). Le serveur de logs enregistre tous les événements (début et fin de navigation, pages visitées et recommandations).

Les serveurs (d'événements, de recommandations, de logs et d'information) sont indépendants de l'annuaire thématique et utilisent dans Broadway*Tools les composants objet relatifs à la partie spécialisation d'un service de recommandations dans un annuaire thématique. Les servlets correspondent, elles, à l'intégration de ces serveurs dans ces annuaires.

Le serveur de recommandations utilisé dans cette plate-forme est en cours de réalisation et s'inspire de celui de Broadway-Web. Les principales modifications par rapport au recommandeur de Broadway-Web concernent la prise en compte des comportements visuels (en termes de zones fixées et d'un score de lecture).

6.6 Aide à la vérification sémantique de sites Web

Mots clés : sémantique, sites Web, services adaptatifs, Web Semantics, approches formelles, sémantique naturelle, typage, vérification, Centaur, CLF, adaptation à l'utilisateur, personnalisation.

Participants : Thierry Despeyroux, Brigitte Trousse.

Nous avons poursuivi notre évaluation de l'apport d'une sémantique de style sémantique naturelle [Kah87] pour la spécification et la vérification d'une page Web, voire d'un site Web [14, 13].

Notre motivation au niveau du "Web dsSémantique" concerne principalement l'aide à la spécification et à la vérification de sites Web. Très peu de travaux abordent la vérification sémantique de sites, que ce soit avec des techniques issues de l'IA ou du génie logiciel : citons l'un d'eux (en IA) WebMaster [HM99].

Notre approche est inspirée de travaux précédents en sémantique des langages de programmation, traçant un parallèle entre la syntaxe des langages de programmation et la structure des sites Web (ou de documents semi-structurés) et entre la sémantique des programmes et la sémantique des sites Web, appliquant des notions de types et de règles sémantiques aux documents présents sur le Web.

Cette année, nous nous sommes d'une part affranchis du système Centaur utilisé précédemment, d'autre part nous nous sommes intéressés à des données réelles (comme des pages Web du site de l'Inria et aussi des données d'analyse générées en XML à partir d'une base de données).

Si on rapproche la vérification de systèmes d'information de la notion de compilation de programmes, apparaît rapidement la nécessité d'exprimer des règles de dépendance comme celles qui apparaissent dans un « makefile ». Cependant, la complexité de la structure d'un système d'information nous a fait toucher les limites d'un programme tel que « make » : manipulation d'un grand nombre de fichiers et de répertoires, délocalisation, utilisation de données accessibles par leur URL etc. .

Nos travaux ont fait l'objet d'une présentation à deux conférences internationales (HYPERTEXT'01 et Webnet'01).

7 Contrats industriels (nationaux, européens et internationaux)

7.1 Contacts industriels

Nous avons présenté les recherches de l'équipe à plusieurs reprises :

-
- [Kah87] G. KAHN, « Natural Semantics », *in: Proceedings of STACS'87*, Lecture Notes in Computer Science n 247, Springer-Verlag, Berlin, février 1987. aussi Rapport de Recherche de l'INRIA Sophia Antipolis N 601, <http://www.inria.fr/rrrt/rr-0601.html>.
- [HM99] F. V. HARMELEN, J. V. D. MEER, « WebMaster: Knowledge-based Verification of Web-pages », *in: Twelfth International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems IEA/AIE'99*, 1999.

- le 8 février à une délégation d'Hitachi pour discuter en particulier de collaborations possibles en relation avec l'accord cadre que nous avons avec eux ;
 - le 5 octobre, à trois membres d'une délégation du laboratoire de recherche de DENSO (MM Namba, Ueno et Taguchi) ;
 - le 31 octobre, à D. Galarreta (CNES, Toulouse).
 - le 30 novembre à quatre membres de SAP AG (Corporate Research) des sites de Sophia (France) et de Karlsruhe (Allemagne) : L. Gomez, C. Hebert, J. Posegga, J. Schaper
- Nous avons poursuivi nos contacts avec la société Scorex SAM (Monaco) en vue de préciser l'application de nos recherches dans le domaine du commerce électronique et du "scoring".

8 Actions régionales, nationales et internationales

8.1 Actions régionales

- Action Colors (2001-2002) : le projet **e-Behaviour** (cf. section 6.5) est une coopération principalement entre notre équipe et le LPEQ (UNSA). <http://www-sop.inria.fr/COLOR/2001/e-BEHAVIOUR.html>

8.2 Actions nationales

- INRIA Rocquencourt : collaboration sur l'analyse de l'usage avec Yves Lechevallier, Aicha Oubaiche et sur la vérification sémantique de documents XML avec Marc Csernel dans le cadre des recherches proposées dans AxIS (créée le 6 novembre 2001). <http://www-sop.inria.fr/axis/>.
- Site stic : dans le but de montrer l'intérêt de nos travaux dès les premières phases de la conception du futur site **stic** dont l'Inria est à l'origine, nous avons étudié l'intégration d'un serveur de traces de sessions et d'un service de recommandations personnalisées dans un annuaire thématique généré à partir des ressources en XML déjà identifiées (cf. en intranet à l'inria sophia antipolis le site de nom **didacticiel**). Une démonstration de notre travail a eu lieu à Sophia Antipolis en octobre auprès de D. Chouchan et Th. Vieville.
- Laboratoire des Usages, CNRT, Sophia Antipolis (http://www-sop.inria.fr/aid/Labo_des_usages.pdf), laboratoire qui vise sous la forme d'un GIS à éclairer les principales pistes d'une recherche sur les usages des TIC et d'une recherche technologique inspirée par ces usages (B. Trousse).
- GDR-PRC I3 - Groupement de Recherches « Information - Interaction - Intelligence » du CNRS <http://www-timc.imag.fr/I3/>. Groupe national de raisonnement à partir de cas qui organise chaque année un séminaire de recherche (B. Trousse).

8.3 Actions européennes

- IST Ontoweb : l'action AID participe au réseau OntoWeb (Ontology-based Information Exchange for Multilingual Electronic Commerce and Information Integration) qui a été proposé en 2000 à l'initiative de Dieter Fensel (Division of Mathematics & Computer

Science, Vrije Universiteit Amsterdam). De l'INRIA, interviennent également les équipes ACACIA (Sophia), Orpailleur (Nancy) et Exmo (Grenoble).

- COST Action 282 (2001-2005) : “Knowledge Exploration in Science and Technology”. Y participent de l'INRIA, ORPAILLEUR, VERSO et AID.

Nous avons eu diverses actions au niveau européen :

- ERCIM : B. Trousse a participé en juin à l'évaluation des candidats au prix européen Cor Baayen (ERCIM) de la meilleure thèse.
- Maccinet : nous avons réalisé la première version (installée sur une machine Inria) du site du réseau européen Maccinet en collaboration avec J-A. Désidéri et T. Nguyen (SINUS).

8.4 Actions internationales

B. Trousse a donné une conférence invitée au 3ème séminaire international SYNASC'01 *Symbolic and Numeric Algorithms for Scientific Computing* (organisé par RISK-Linz et l'université West de Timisoara) le 3 octobre à Timisoara, Roumanie. Cette conférence était intitulée : “Design of web sites and adaptive recommender systems based on usage analysis”.

9 Diffusion de résultats

9.1 Animation de la communauté scientifique

B. Trousse est membre du comité de rédaction de deux revues :

- Revue d'Intelligence Artificielle RIA publiée par Hermès qui diffuse les résultats de la communauté francophone d'IA (rédacteur en chef : M. Pomerol)
- Revue électronique I3 du GDR-I3 (rédacteurs en chef : C. Garbay et H. Prade) : <http://www.Revue-I3.org/>

9.1.1 Visites

L'action AID a reçu cette année plusieurs visites :

- 15 juin : une délégation finlandaise à l'initiative de Minna-Kristiina Ojavainio (chef de projet) de l'“Institute for Extension Studies” de l'Université of Tampere ;
- 5 octobre : Jérôme Chailloux (Institut Pasteur) ;
- à plusieurs reprises : Thierry Baccino, Sylvain Denis et Thérèse Colombi de l'UNSA (LPEQ de l'UNSA) dans le cadre du projet e-behaviour (Colors) ;
- 3 décembre : Dr. Joseph A. Konstan, Professeur Associé au département “Computer Science & Engineering” de l'Université du Minnesota (USA), et responsable du projet de recherche GroupLens (systèmes de recommandations de type filtrage collaboratif) supporté par la NSF et la société Net Perceptions..

B. Trousse a rendu visite au Prof Viorel Negru à l'université de West Timosoara (Roumanie) en octobre.

9.1.2 Serveur interne Web

L'action AID a poursuivi la mise au point de son site Web qui permet d'accéder à un certain nombre d'informations relatives à nos recherches et en particulier à notre plate-forme objet CBR*Tools de partage et de réutilisation d'expériences et aux deux systèmes Broadway et Hermès : <http://www-sop.inria.fr/aid/>.

Dans le cadre du travail du poste d'accueil (Kostas Kontantinidis), une nouvelle version en XHTML basé sur les outils du W3C (avec et sans frame) a été réalisée. Des outils de génération de la page *publications* ont été réalisés à partir d'un fichier .bib Une machine pour nos expérimentations sur le World Wide Web (Hermès et Broadway) est installée à l'Inria .

9.1.3 Divers

T. Despeyroux est président de l'AGOS, membre titulaire de la commission technique paritaire (CTP) et membre titulaire du conseil d'administration de l'INRIA en tant que représentant du personnel.

9.2 Formation

9.2.1 Enseignement universitaire

Nous avons participé cette année aux enseignements suivants :

- DEA Informatique (resp. E. Kounalis) à l'UNSA Sophia Antipolis : cours sur les *frameworks objet* (4h) dans le cours du tronc commun (TC4) "Langages et Modèles à objets" (resp. M. Blay) : B. Trousse
- DESS "Ergonomie et NTIC" (resps. Th. Baccino et J. Araszkievies) à l'UNSA : cours (18h) sur la *conception de sites Web évolutifs* et sur la *conception de sites/services adaptatifs d'aide à la recherche d'informations* : B. Trousse.
- ESINSA à l'UNSA, Sophia Antipolis : TP (92h) d'*algorithmique & programmation en java* : S. Gaieb.
- ESSI à l'UNSA, Sophia Antipolis : TP (84h) de *Java* : D. Tanasa..

9.2.2 Thèses

L'action AID est équipée d'accueil de doctorants et stagiaires de DEA de la formation doctorale des sciences pour l'ingénieur de Nice-Sophia Antipolis (DEA Informatique).

Thèse soutenue :

1. **Semi Gaieb**, Outil d'aide à la programmation parallèle asynchrone, université de Nice-Sophia Antipolis, juillet 2001 (directeur de thèse Th. Despeyroux).

Thèse en cours :

1. **Doru Tanasa**, Trace et analyse de l'usage pour l'aide à la reconception d'un site Web, Université de Nice-Sophia Antipolis (directeur de thèse Brigitte Trousse).

9.2.3 Stages

Nous avons accueilli 7 étudiants qui ont travaillé sur les projets et stages suivants (ordre chronologique) :

- **Cédric Sempe** et **Gilles Dodinet**, étudiants de l’ESSI 3ème année (2000-2001, UNSA), ont effectué leur projet jusqu’à mi-avril 2001 sur **logaudience**, un ensemble d’outils génériques d’analyse de logs HTTP en termes de sessions. Ces outils visent à répondre à certains besoins identifiés au niveau du site Web INRIA national [18].
- **Ferdinand Dosser** et **Laurent Perrimond**, étudiants de l’ESSI 3 ème année (2000-2001, UNSA), ont effectué leur projet jusqu’à mi-avril 2001 ; [17].
- **Doru Tanasa**, étudiant du DEA d’Informatique (2000-2001, UNSA) dans le cadre du programme SOCRATES, a effectué son stage de 6 mois à compter du 15 janvier sur l’analyse et la réutilisations de séquences temporelles [20]. Nous avons accueilli cet étudiant à nouveau jusqu’à fin septembre pour poursuivre le travail commencé [15].
- **Patrice Carrique**, étudiant de maîtrise d’informatique (2000-2001) de l’UNSA a effectué un stage de trois mois (été 2001) à mi-temps sur **logaudience** et sur le projet e-behaviour [16].
- **Nathalie Evan**, étudiante de l’ESSI (UNSA, 2001-2002), effectue son projet à compter de mi-octobre sur les aspects trace des comportements visuels et non visuels dans une base de données relationnelle dans le cadre du projet e-behaviour.

9.3 Participation à des colloques, séminaires

Nous avons présenté nos travaux de recherche dans des séminaires et conférences internationales : HYPERTEXT’01, Webnet’O1 et SYNASC’01. Voir la bibliographie de cette année pour plus de détails.

Outre ces conférences ou séminaires, nous avons participé

- à la plate-forme AFIA (25-27 juin) à Grenoble : B. Trousse.
- au 2 ème meeting du réseau européen Ontoweb le 7-8 décembre à Amsterdam : Th. Despeyroux et B. Trousse.

et présenté nos travaux de recherche lors de diverses manifestations nationales INRIA :

- Journées scientifiques sur le thème “Web sémantique” à Nancy les 25 et 26 juin : B. Trousse et T. Despeyroux.
- Journées scientifiques sur le thème “Statistiques” à Rennes les 15 et 16 novembre. : B. Trousse (en collaboration avec Y. Lechevallier).
- Journée INRIA-Industrie sur le thème à Rocquencourt le 6 décembre. : B. Trousse et Sémi Gaieb.

10 Bibliographie

Ouvrages et articles de référence de l’équipe

- [1] C. BRANKI, N. KARACAPILIDIS, B. TROUSSE, « Computer-Supported Cooperative Work in Design », *Journal of Design Sciences and Technology* 6, 1, 1998, 115 pages.

- [2] T. DESPEYROUX, « AS, for Abstract Syntax - Manual - V1.0 », *rapport de recherche*, Inria, septembre 1996, Technical Report no. 197.
- [3] M. JACZYNSKI, B. TROUSSE, « WWW Assisted Browsing by Reusing Past Navigations of a Group of Users », in : *Proceedings of the European Workshop on Case-based Reasoning, LNCS/AI*, Springer-Verlag, Dublin, Ireland, septembre 1998.
- [4] M. JACZYNSKI, B. TROUSSE, « Patrons de conception dans la modélisation d'une plate-forme pour le raisonnement à partir de cas », *Revue l'Objet* 5, 2, 1999, Numéro Spécial sur les patterns orientés objets, D. Rieu et J-P. Giraudon (guest editors).
- [5] M. JACZYNSKI, *Modèle et plate-forme à objets pour l'indexation des cas par situation comportementale : application à l'assistance à la navigation sur le Web*, thèse de doctorat, Université de Nice Sophia-Antipolis, Sophia-Antipolis, décembre 1998, (In french).
- [6] N. KARACAPILIDIS, D. PAPADIAS, « Hermes : Supporting Argumentation Discourse in Multi-Agent Decision », in : *Proceedings of the AAI-98 Conference*, AAAI/MIT Press, p. 827–832, 1998.
- [7] N. KARACAPILIDIS, B. TROUSSE, D. PAPADIAS, « Using Case-Based Reasoning for Argumentation with Multiple Viewpoints », in : *Case-Based Reasoning Research and Development, Proceedings of the 2nd Int. Conference on Case-Based Reasoning (ICCB-97)*, Providence, Rhode Island, July 25-27, 1997, D. Leake, E. Plaza (éditeurs), Lecture Notes in AI 1266, Springer-Verlag, Berlin, p. 541–552, 1997, <http://www-sop.inria.fr/aid/papers/97iccb.html>.
- [8] M. OUSSALAH, B. TROUSSE, « Numéro Spécial sur les représentations par objets en conception », *Revue l'Objet* 4, 2, juin 1998, 226 pages.
- [9] B. TROUSSE, W. VISSER, « Use of Case-Based Reasoning Techniques for Intelligent Computer-Aided-Design Systems », in : *IEEE International Conference on Systems, Man and Cybernetics - Systems Engineering in the Service of Humans*, Le Touquet, France, October 1993.
- [10] B. TROUSSE, « Vers des outils d'aide à la conception coopérative : "Design Groupware" », in : *Connaissances et savoir-faire en entreprise - Intégration et capitalisation*, J.-M. Fouet (éditeur), Hermes, Paris, 1997, ch. 17, p. 317–341.
- [11] B. TROUSSE, « Viewpoint Management for Cooperative Design », in : *Proceedings of the IEEE Computational Engineering in Systems Applications (CESA '98)*, M. K. P. Borne, A. E. Kamel (éditeurs), UCIS - Ecole Centrale de Lille - CD-Rom, avril 1998.

Thèses et habilitations à diriger des recherches

- [12] M. GAÏEB, *Modèle et langage pour la spécification d'architectures basées sur les bus logiciels à messages*, thèse de doctorat, Université de Nice Sophia-Antipolis, juillet 2001.

Communications à des congrès, colloques, etc.

- [13] T. DESPEYROUX, B. TROUSSE, « Maintaining Web sites », in : *Webnet 2001, Orlando, USA*, octobre 2001.
- [14] T. DESPEYROUX, B. TROUSSE, « Web sites and Semantics », in : *HYPERTEXT'01, the twelfth ACM Conference on Hypertext and Hypermedia, Aarhus, Danemark*, p. 239–240, août 2001.
- [15] D. TANASA, B. TROUSSE, « Web Access pattern Discovery and Analysis based on Page Classification and on Indexing Sessions with a Generalised Index Tree », in : *SYNASC 2001, Timisoara, Roumanie*, p. 62–72, octobre 2001.

Divers

- [16] P. CARRIQUE, S. GAIEB, B. TROUSSE, « Projet e-Behaviour : spécifications de la plateforme d'expérimentation », Rapport technique Colors/e-Behaviour version 1, décembre 2001.
- [17] F. DOSSER, L. PERRIMOND, « Broadway V1.1 ou reconception d'un système d'aide à la navigation sur le web : utilisation d'un nouveau framework et prise en compte de bases composites », Mémoire de projet ESSI, Université de Nice Sophia Antipolis (UNSA), avril 2001.
- [18] C. SEMPÉ, G. DODINET, « Analyse de l'usage d'un site Web », Mémoire de projet ESSI, Université de Nice Sophia Antipolis(UNSA), avril 2001.
- [19] D. TANASA, B. TROUSSE, « Découverte et analyse de patrons d'accès Web », Poster paper in "Journées francophones d'Extraction et de Gestion des Connaissances" (EGC'02), janvier 2002, à paraître.
- [20] D. TANASA, « Analyse et réutilisations de séquences temporelles. Application aux comportements d'internautes », Mémoire de DEA d'Informatique, Université de Nice Sophia Antipolis(UNSA), Juillet 2001, 24 pages.