

Action AID

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

Sophia Antipolis

THÈME 3A



*R*apport
*d'**A*ctivité

2000

Table des matières

1	Composition de l'équipe	3
2	Présentation et objectifs généraux	4
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 en 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 de sites Web	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 »	11
5.2	ROL - « Reactive Object Language »	12
5.3	CBR*Tools - Plateforme objet en raisonnement à partir de cas	12
5.4	Broadway*Tools - Pour la génération de systèmes de recommandations adaptatifs	13
5.5	Broadway-1 - 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	16
6.1	Panorama	16
6.2	Raisonnement à partir de cas	16
6.2.1	Approche neuronale pour l'indexation de cas	17
6.2.2	CBR*Tools : ajout de nouveaux composants d'indexation de cas	18
6.3	Conception de systèmes de recommandations personnalisées	18
6.3.1	Réalisation et évaluation du système Broadway-2 (FT)	19
6.3.2	Capitalisation d'objets métiers dans Broadway*Tools	19
6.4	Évaluation du langage ROL pour nos applications Web	19
6.5	Sémantique de sites/services Web	20
6.5.1	Analyse comparative des travaux sur la sémantique de sites Web	20
6.5.2	Aide à la vérification sémantique de sites Web	20
6.5.3	Aide à la vérification sémantique de services Web adaptatifs	21
6.6	Analyse des usages	22

7 Contrats industriels (nationaux, européens et internationaux)	22
7.1 Contrat CEA Cadarache	22
7.2 Contrat France Télécom R&D (CTI)	22
7.3 Contacts industriels	23
8 Actions régionales, nationales et internationales	23
8.1 Actions régionales	23
8.2 Actions nationales	23
8.2.1 GDR-PRC I3	23
8.2.2 Laboratoire des Usages, CNRT, Sophia Antipolis	24
8.2.3 Site Web INRIA	24
8.2.4 Collaborations	24
8.3 Actions internationales	24
8.3.1 Europe	24
9 Diffusion de résultats	25
9.1 Animation de la communauté scientifique	25
9.1.1 Comités de programme	25
9.1.2 Visites	25
9.1.3 Serveur WWW	25
9.1.4 Divers	25
9.2 Formation	26
9.2.1 Enseignement universitaire	26
9.2.2 Thèses	26
9.2.3 Stages	26
9.3 Participation à des colloques, séminaires	27
10 Bibliographie	27

1 Composition de l'équipe

Responsable scientifique

Brigitte Trousse [Chargée de recherche, Inria]

Assistante de projet

Hortense Hammel [Inria, à mi-temps, 1 janvier au 31 octobre]

Sophie Honnorat [Inria, à 1/4-temps, à partir du 1 novembre]

Personnel Inria

Thierry Despeyroux [Chargé de recherche, Inria]

Poste d'accueil

Konstantinos Konstantinidis [à partir du 16 octobre]

Chercheur doctorant

Mohamed Semi Gaieb [UNSA, boursier MESR puis boursier ATER]

Collaborateur extérieur

Michel Jaczynski [ingénieur ILOG, France, 1 avril au 30 septembre]

Bernard Senach [consultant ergonome, Ergomatics, France, 1 avril au 30 septembre]

Stagiaires

Attila Benedek [ESSI, UNSA & West Univ of Timisoara, Roumanie, 1 avril au 20 décembre]

Gilles Dodinet [ESSI, UNSA, 16 octobre au 31 décembre]

Ferdinand Dossier [ESSI, UNSA, 16 octobre au 31 décembre]

Hakim Maadadi [ESSI, UNSA, 1 janvier au 30 avril]

Maamar Maameri [ESSI, UNSA, 1 janvier au 30 avril]

Laurent Perrimond [ESSI, UNSA, 16 octobre au 31 décembre]

Anshuman Rai [ITT Kanpur, Inde, 31 mai au 30 septembre]

Laurent Rineau [ENS Ulm, Paris, 19 juin au 16 septembre]

Cedric Sempe [ESSI, UNSA, 16 octobre au 31 décembre]

Adrien Zamani [ESSI, UNSA, 1 janvier au 30 avril]

2 Présentation et objectifs généraux

Mots clés : génie logiciel, intelligence artificielle, STIC, activité de conception, sites/services Web, sémantique, 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, document semi-structuré.

L'action de recherche AID est une équipe pluridisciplinaire (Intelligence artificielle et Génie logiciel) depuis 1999. Elle mène des recherches dans le domaine du Web, domaine qui fût un choix en 96 [Tro99]. L'action AID 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).

Plus précisément, nous menons des recherches selon 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 la réutilisation d'expériences passées en conception.
2. Aide à la spécification et à la vérification sémantique de sites Web basée 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 thème s'appuie sur les résultats des deux premiers.

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..

Nos principaux résultats concernent le thème « 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 orienté objet

1. ou encore *CBR* pour « *Case-Based Reasoning* »

[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.

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 collabore avec plusieurs industriels comme France Télécom et le centre de recherche européen XRCE (France). De plus, elle participe au niveau national ou international à diverses communautés scientifiques.

3 Fondements scientifiques

3.1 Panorama

Résumé : *L'action AID mène des recherches sur l'utilisation des techniques d'intelligence artificielle pour l'aide à l'utilisation et à la 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 appliquée au Web. D'autre part nous étudions l'utilisation de méthodes formelles des langages en sémantique 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 : réutilisation d'expériences passées, indexation, case-based reasoning, chroniques, temps.

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, dont le but est de rechercher des 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,

[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 on Artificial Intelligence* 7, 1, 1994, p. 39–59.

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

3. une phase de révision de la solution qui permet de l'affiner grâce à un processus d'évaluation,
4. une phase d'apprentissage, chargée 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.

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 obtenus en 99, en particulier le modèle d'indexation par situation comportementale, la plate-forme orientée objet CBR*Tools et la boîte à outils Broadway*Tools. De plus, nous avons étudié plus particulièrement des techniques d'indexation pour une recherche efficace de cas issues d'approches neuronales.

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 peu d'informations leur sont disponibles, afin de trier et d'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 recommandation est le suivant :

Comment produire de 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 fondée sur l'apprentissage automatique de profils utilisateurs et 2) l'approche fondée sur des techniques de fouille de données. Le profil utilisateur est une structure de données qui décrit les

-
- [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.

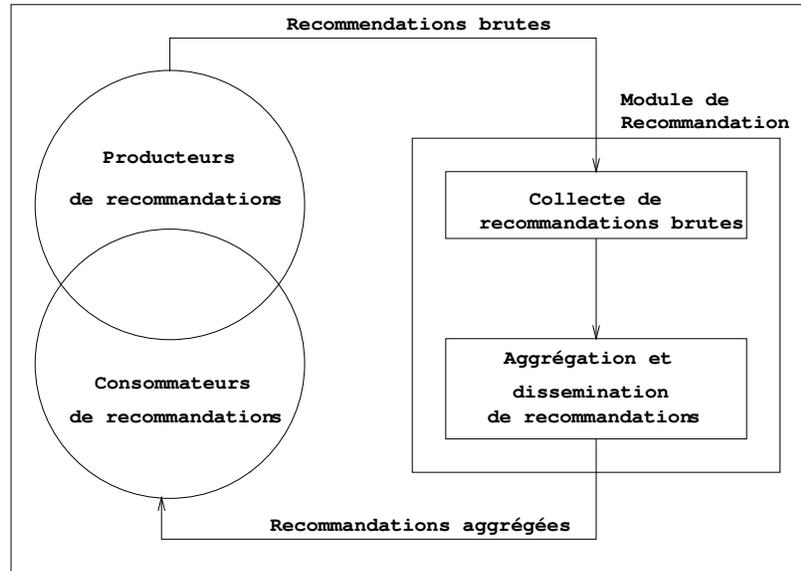


FIG. 1 – Architecture générale d'un système de recommandation

centres d'intérêts d'un utilisateur dans l'espace des objets à recommander. Une fois une telle structure construite, on peut l'utiliser 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].

Dans l'action AID, nous avons poursuivi le développement d'une approche de calcul de recommandations centrée fouille de donnée 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,

[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).

3. définir des techniques d'inférence de recommandations personnalisées à partir des comportements pertinents passés identifiés (ou à partir des cas remémorés).

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 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 en conception de systèmes d'informations hypertextes

Mots clés : Web, sémantique, typage, documents semi-structurés, conception, maintenance, sites Web, 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 avec RDF et "RDF schema" ainsi que certains travaux de recherche issus de l'IA basés sur une approche ontologique ^[HM99] comme On2broker. Notre approche diffère dans le fait que nous voulons exploiter plus loin le parallèle entre programmes et sites Web pour mieux aborder la sémantique des sites.

Nous pouvons d'ores et déjà distinguer la sémantique statique d'un site qui peut être vu comme un ensemble de contraintes globales (pas seulement syntaxiques, mais aussi sémantiques

[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.

et dépendant du contexte) qui doivent être vérifiées et la sémantique dynamique qui prend en compte la navigation d'un utilisateur dans un site et rejoint donc l'analyse des usages. Dans un avenir plus lointain, nous pensons 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 : conception de sites Web, navigation, recherche d'information, annuaires thématiques, portails, sites institutionnels, commerce électronique, sites éducatifs, analyse des usages, analyse des logs, système de recommandations.

Résumé :

Les domaines d'application visés par le projet concernent le Web [Tro99], plus particulièrement 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 4.5).

Cependant nos recherches en RàPC peuvent être appliquées dans de nombreux domaines pour des problèmes d'aide à la recherche et la réutilisation d'expériences passées, par exemple en planification, conception, diagnostic, etc. ; elles 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, médecine) ou en conception de systèmes d'assistance.

De même, nos travaux sur la corrélation de points de vues peuvent également avoir des retombées dans de très nombreux domaines en ingénierie concourante, dans le cadre de la conception de collecticiels ou de systèmes avancés d'informations nécessitant une indexation d'expériences passées en termes de points de vues.

4.2 Recherche collaborative d'informations sur le Web

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

Le World Wide Web (WWW) 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

[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.*

navigation. Cependant, la complexité et l'étendue du WWW 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 WWW 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 de sites Web

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

Avec le développement fulgurant d'Internet et l'utilisation de ses standards comme mode de diffusion et de représentation de données mais aussi en intranet, 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.

4.4 Analyse des usages sur internet/intranet

Mots clés : Web mining, analyse des usages, comportements utilisateur, prediction, 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 à ces concepteurs pour anticiper les usages.

Les méthodes actuelles (off-line) d'analyse des fichiers logs HTTP proposées actuellement sur le marché n'abordent pas le niveau *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 concurrente, 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 réflexions quand 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 propre 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 : personnalisation, expérience passée, mémoire de cas raisonnement à partir de cas, plate-forme orientée objet, réutilisation, UML, système de recommandations, navigation, filtrage collaboratif, décision, négociation, coopération, argumentation, Web, langages de spécification, sémantique, Centaur, programmation parallèle asynchrone, objets réactifs, bus logiciel.

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>

5.1 CLF - « Computer Language Factory »

Mots clés : langages de spécification, sémantique, Centaur.

Participant : Thierry Despeyroux [correspondant].

CLF, développé par Thierry Despeyroux (lorsque celui-ci était membre du projet CROAP) est un ensemble d'outils et de formalismes de spécification de la syntaxe et la sémantique de langages informatisées. CLF propose actuellement les langages AS^[Des96] (Abstract Syntax) et

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

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

CS (Concrete Syntax).

CLF, utilisé dans le système Centaur <http://www-sop.inria.fr/croap/centaur/centaur.html> et basé sur la sémantique naturelle ^[Kah87], est actuellement appliqué à la spécification de XML et utilisé pour la génération d'un environnement d'aide à la vérification sémantique de sites Web (cf. section 6.5.1).

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 [13], 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 et en cours de validation pour des systèmes basés sur l'approche Broadway, en particulier pour l'aide à la navigation dans un répertoire thématique (cf Broadway-AT) (cf section 9.2.3).

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

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

Participante : 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.*

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

-
- [Kah87] G. KAHN, « Natural Semantics », *in: Proceedings of STACS'87*, Lecture Notes in Computer Science n 247, Springer-Verlag, Berlin, février 1987. also Rapport de Recherche de l'INRIA Sophia Antipolis N 601, <http://www.inria.fr/rrrt/rr-0601.html>.
 - [Jac98] M. JACZYNSKI, *Modèle et plate-forme à objets pour l'indexation des cas par situation comportementales: 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.

concepts nécessaires pour développer une application intégrant des techniques de raisonnement à partir de cas : cas, bases de cas, index, similarité, contrôle du raisonnement. Il offre également un ensemble de classes concrètes qui implantent un certain nombre de méthodes classiques (indexation par plus proches voisins, indexation par Kd-tree ^[WAD94], indexation par prototypes ^[JT94], indexation basée sur une approche neuronale [15], 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'applications 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 98 dans le cadre de Broadway 1 et a été utilisée dans le cadre d'un contrat XRCE-INRIA (98) ainsi que dans un FT-CTI (1998-2000). Une première version d'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-1, FT CTI, XRCE, Broadway-Predict). En effet pour évaluer l'utilisation de CBR*Tools, nous avons montré que, pour chaque application, l'expertise approfondie nécessaire varie de 20% à 40% des points d'ouverture validant ainsi l'aide apportée par CBR*Tools 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 - Pour la génération de systèmes de recommandations adaptatifs

Mots clés : agent personnalisé, boîte à outils, système distribué, programmation asynchrone, composants logiciels, réutilisation, services adaptatif, personnalisation.

Participants : Brigitte Trousse [correspondante], Attila Benedek, 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*

[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.

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-1 pour l'aide à la navigation sur le Web et Broadway-2 pour l'aide à la navigation dans un clone d'un site de France Télécom.

5.5 Broadway-1 - 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, Konstantinos Konstandinidis, Laurend Perrimond.

Résumé : *Broadway-1*³ [MT99] est un assistant pour la navigation sur le Web réutilisant les navigations passées d'un groupe d'utilisateurs. Broadway observe les navigations de différents utilisateurs et récolte les évaluations et les annotations de ces utilisateurs pour établir une liste de documents pertinents.

Broadway (V1) (<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 WWW 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 5.5) 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 page HTML visualisées. Broadway intègre le serveur HTTP Jigsaw du W3C (World Wide Web Consortium) programmé en Java et utilise notre plate-forme CBR*Tools pour implanter le système de raisonnement à partir de cas permettant la réutilisation des navigations passées.

Broadway(V1) a été installée à France Télécom (R&D) à Lannion et fut l'objet d'une démonstration aux huitèmes rencontres INRIA-Industrie *Informatique pour les Télécommunications et le Multimédia* le 26 novembre 1998.

3. Brodway – «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.

The image shows a Netscape browser window displaying the AID (Action) website. The browser title is "Netscape: AID (Action)". The address bar shows "http://www.inria.fr/Equipes/AID-fra.html". The page content includes the title "AID (Action)" and the subtitle "Gestion des connaissances pour l'aide à la conception coopérative". Below this, there is a paragraph in French describing the research action at INRIA Sophia Antipolis. To the left of the main content, there is a "Broadway Manager" window with a "Navigation #4 from" summary. The summary includes start and end dates, keywords, and a table of navigation history.

Navigation #4 from
 start date: Wed Nov 19 20:01:01 GMT+03:30 1997
 end date: none
 keywords: ACTIVITES, SITE, INRIA, RECHERCHE, BIENVENUE, THEME, ACTION
 hosts: www.inria.fr

#	Page address	Content	Reading	Evaluation
0	http://www.inria.fr/	SITE, INRIA, BIENVENUE	0.0	0
1	http://www.inria.fr/Recherche/activites-fra.html	ACTIVITES, INRIA, RECHERCHE	0.0	0
2	http://www.inria.fr/Themes/Theme3-fra.html	INRIA, THEME	1.0	0

ABOUT BROADWAY
 BROADWAY is a browsing advisor reusing past navigations from a group of users.

FIG. 2 – Utilisation de Broadway 1

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

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

Participants : Semi Gaieb, Maamar Maameri, Brigitte Trousse [correspondante], Adrien Zamani.

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

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 actuels 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 travaillant ensemble dans des lieux différents et à un même instant.

Hermès fût l'objet de recherche à l'EPFL (C. Vanoirbeek) ainsi qu'à University of Science and Technology - Hong Kong (D. Papadias).

6 Résultats nouveaux

6.1 Panorama

Résumé : *Nos principaux résultats concernent 1) l'adaptation de modèles hybrides neuronaux pour l'indexation de cas 2) l'ajout de composants d'indexation dans notre plate-forme de raisonnement à partir de cas CBR*Tools, 3) l'évaluation de notre approche Broadway de calcul de recommandations pour l'aide à la navigation dans un site Web 4) l'évaluation du langage Rol sur un système de recommandations dans un annuaire thématique 5) l'évaluation de la sémantique naturelle pour l'aide à la vérification sémantique de sites Web.*

6.2 Raisonnement à partir de cas

Mots clés : ràpc, indexation de cas, approche neuronale, efficacité, recherche de cas,

-
- [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.

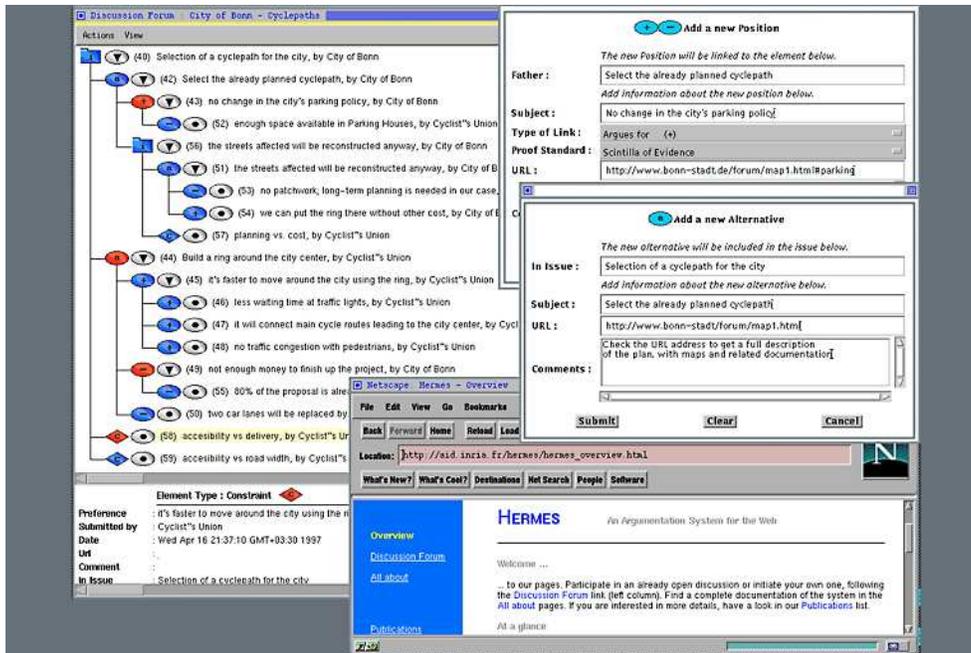


FIG. 3 – Utilisation d'Hermès

apprentissage, composants objet.

Participants : Attila Benedek, Brigitte Trousse.

6.2.1 Approche neuronale pour l'indexation de cas

L'objectif de cette recherche vise à améliorer la performance d'un système de recommandations basé sur du RàPC dans le cas d'une montée en charge du site offrant ce service. L'efficacité du raisonnement d'un système RàPC en termes de mémorisation de nouveaux cas (d'apprentissage) et de remémoration (de rappel) de cas précédents, dépend, d'un côté, de la manière dont les cas sont représentés, indexés, sauves dans le mémoire et, d'un autre côté, des algorithmes utilisés pendant la phase de recherche ou de remémoration.

Parmi les trois différentes formes d'organisation de la mémoire : mémoire plate, mémoire hiérarchique (réseaux à traits partagés et arbres de discrimination [Qui86]) et mémoire connexionniste, nous avons étudié principalement la mémoire connexionniste qui offre une recherche efficace, avec un prix de mémorisation et de sauvegarde optimal, mais aussi avec une précision moindre par rapport à une mémoire plate.

Nous avons fait deux propositions hybrides d'index neuronaux inspirés des modèles Probis

[Qui86] QUINLAN(J.R.), « Induction of Decision Trees », *Machine learning*, no.1, 1986, pp.81-106.

[Mal96] et GCS [Fri94].

Le premier index s'appuie sur le modèle ARN2 [Azc91], qui est un réseau de neurones à base de prototypes avec une structure évolutive, utilisé dans le modèle Probis comme une solution hybride d'indexation RàPC entre la mémoire plate et la mémoire connexioniste. Inspiré par ce modèle, nous proposons un modèle offrant un nouvel algorithme de mémorisation et de remémoration plus efficace (par exemple le traitement des cas frontières) et adapté à nos besoins dans CBR*Tools (traitement des prototypes atypiques)[15].

Le deuxième index est un modèle hybride d'indexation de la mémoire à trois couches: 1) une mémoire plate pour garder les cas qui sont groupés dans le même "cluster", 2) le réseau de neurones GCS pour établir les différents clusters et pour offrir un premier niveau d'indexation, 3) une structure hiérarchique, les arbres k-d [WAD94] pour mettre à jour un deuxième niveau d'indexation des clusters qui se trouvent sur le premier niveau.

6.2.2 CBR*Tools : ajout de nouveaux composants d'indexation de cas

Nous avons conçu et intégré une modélisation objet du premier index dans CBR*Tools en utilisant trois points d'ouverture principaux pour traiter les différents, nouveaux concepts introduits de ce modèle, notamment la distance ou degré de similarité entre deux cas, la représentation numérique d'un cas (par exemple pour une session de navigation sur Internet), et le noyau de l'index qui comprend les algorithmes de mémorisation et de remémoration. L'implantation du deuxième index est en cours de réalisation dans CBR*Tools.

6.3 Conception de systèmes de recommandations personnalisés

Mots clés : comportement utilisateur, filtrage collaboratif, analyse des usages, navigation, 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 : Attila Benedek, Semi Gaieb, Maamar Maameri, Hakim Maamedi, Bernard Senach, Brigitte Trousse, Adrien Zamani.

-
- [Mal96] M. MALEK, *Un modèle hybride de mémoire pour le raisonnement à partir de cas*, thèse de doctorat, Université Joseph Fourier, Grenoble, octobre 1996.
- [Fri94] B. FRITZKE, «Growing Cell Structures- A Self-organizing Network for Unsupervised and Supervised Learning», *rapport de recherche*, ICSI, Berkeley, mai 1994.
- [Azc91] A. AZCARRAGA, A. ET GIACOMETTI, «A Prototype Based Incremental Network Model for Classification Task», *in : In Fourth International Conference on Neural Networks and Their Applications, Nimes, France*, 1991.
- [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.

6.3.1 Réalisation et évaluation du système Broadway-2 (FT)

La réalisation du système de recommandations d'aide à la navigation (Broadway-2) dans un clone d'un site Web de France Télécom a été terminée. Ce système a fait l'objet d'une expérimentation auprès de deux groupes d'utilisateurs avant l'été 2000 afin d'évaluer l'aide apportée. Pour cela, en collaboration avec B. Senach (Ergomatics), nous avons conçu une méthode d'expérimentation identifiant huit tâches ciblées dans le site choisi et les supports correspondants. L'expérimentation a permis de valider l'aide apportée par ce système basé sur l'approche BRoadway, ceci via 1) une analyse de critères objectifs issus des traces du système donnant des résultats significatifs de 0,1% à 2% et 2) une analyse de critères subjectifs issus de recueils d'opinion auprès des utilisateurs.

6.3.2 Capitalisation d'objets métiers dans Broadway*Tools

Nous avons poursuivi notre effort visant à capitaliser des objets métiers au sein de notre boîte à outils Broadway*Tools. En particulier, des composants logiciels ont été rajoutés relativement à l'application de nos index neuronaux pour l'aide à la navigation dans un site Web.

Dans le cadre d'un projet ESSI [16], nous avons étudié les aspects performance d'un annuaire thématique intégrant un système de recommandations de catégories et/ou de sites Web via la technologie des servlets : analyse et conception d'expérimentations et outils d'analyse en Perl.

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, n bus logiciel, Web, décision, point de vue, confrontation, négociation, forum de discussion, argumentation, conception.

Participants : Thierry Despeyroux, Semi Gaieb, Brigitte Trousse, Adrien Zamani.

L'expérience de la conception de Centaur par l'équipe Croap (Inria Sophia Antipolis) ainsi que celle plus récente de la conception d'applications Web par l'action AID (systèmes basés sur Broadway*Tools, Hermès) mettent en évidence le besoin d'environnements adaptés à la programmation asynchrone distribuée qui assistent le concepteur durant toutes les étapes de la construction de telles applications.

Un état de l'art a été fait montrant l'existence de nombreux outils (Sophtalk, Corba/IDL, Java/RMI Olan/OCL, Esterel, Électre), leurs avantages et leurs limites.

Le principal objectif d'un nouvel outil est de regrouper les avantages de chacun de ces outils : celui-ci devra permettre 1) de décrire clairement les interactions entre les objets 2) de rester indépendant du langage hôte, 3) de décrire une architecture dynamique, indépendante de la configuration de l'application, et 4) d'offrir une analyse statique et dynamique de l'architecture ainsi décrite.

Nos résultats dans le cadre de la thèse de S. Gaieb, consistent en (i) un langage de description (appelé ROL pour *Reactive Object Language*) quasi-indépendant du langage hôte (syntaxe proche de Java) [13], (ii) la définition d'une traduction automatique vers Java (nous avons écrit

les bibliothèques nécessaires), (iii) une indépendance totale de la situation géographique des objets et (iv) un début d'interface graphique. Les avantages de ce langage sont la syntaxe intuitive des réactions, la possibilité d'avoir des objets hétérogènes et une architecture dynamique. Cette année, nos travaux ont porté sur une traduction automatique vers un langage cible (Java) et sur une analyse sémantique partielle du code écrit en ROL.

Enfin, dans un but d'évaluation du langage ROL, nous l'avons appliqué cette année au logiciel Hermès (dans le cadre d'un projet ESSI [19] et à un système d'aide à la navigation dans un annuaire thématique basé sur l'approche Broadway.

6.5 Sémantique de sites/services 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, Anshuman Rai, Laurent Rineau, Brigitte Trousse.

6.5.1 Analyse comparative des travaux sur la sémantique de sites Web

Nous avons poursuivi notre analyse des principaux travaux sur la sémantique de sites Web [12], qui peuvent se classer en sensibilités complémentaires : classification et représentation des connaissances (Shoe, On2broker^[FAD⁺98]), outils syntaxiques de représentation de la connaissance (XML et RDF), outils d'expression de contraintes et/ou de manipulation de documents (WebMaster, XML^[MS99], Xduce^[HP00], grammaires attribuées^[PCR99]). C'est dans cette dernière catégorie qu'il faut ranger les travaux menés dans l'équipe.

Anshuman Rai [17] a fait une analyse comparative des approches XSLT, PPML, XML pour la manipulation et la transformation de documents Web ainsi que de Typol et Webmaster pour la vérification.

6.5.2 Aide à la vérification sémantique de sites Web

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 [12].

-
- [FAD⁺98] D. FENSEL, J. ANGELE, S. DECKER, M. ERDMANN, H.-P. SCHNURR, R. STUDER, A. WITT, « On2broker: Lessons Learned from Applying AI to the Web », *rapport de recherche*, Institute AIFB, 1998.
- [MS99] E. MEIJER, M. SHIRLDS, « XML : Afunctional programming language for constructing and manipulating XML documents », *in: Submitted to USENIX 200 Technical Conference, 1999*, 1999.
- [HP00] H. HOSOYA, B. C. PIERCE, « XDuce : A typed XML processing Language », *in: In Proceedings of Third International Workshop on the Web and Databases (ICFP) 2000*, 2000.
- [PCR99] G. PSAILA, S. CRESPI-REGHIZZI, « Adding Semantics to XML », *in: Second Workshop on Attribute Grammars and their Applications, WAGA'99*, 1999.
- [Kah87] G. KAHN, « Natural Semantics », *in: Proceedings of STACS'87*, Lecture Notes in Computer Science n 247, Springer-Verlag, Berlin, février 1987. also Rapport de Recherche de l'INRIA Sophia Antipolis N 601, <http://www.inria.fr/rrrt/rr-0601.html>.

Notre motivation au niveau du “Web Sémantique” concerne ici principalement l’aide à la conception, à la spécification et à la vérification de sites Web plutôt que l’aide à la recherche d’informations. 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.

Pour atteindre ce but, nous avons d’une part utilisé le générateur d’environnement Centaur et CLF pour définir les langages XML et RDF [17, 18] et, d’autre part, utilisé directement Sax et Prolog⁴ afin de construire des maquettes d’outils de vérification sémantique en utilisant la sémantique naturelle, en ayant comme domaine d’application privilégié des sites de type institutionnel (comme celui de l’INRIA) et des annuaires thématiques ou portails (comme *yahoo*, *voila*).

L’utilisation de la sémantique naturelle montre assez clairement la différence entre la vérification syntaxique (par exemple valider une page avec une DTD, comme un validateur XML), qui est indépendante du contexte et une vérification sémantique qui dépend du contexte. L’exemple des annuaires thématiques montre la possibilité d’utiliser des outils ou des ressources externes (thesauri, ontologies).

La sémantique naturelle utilisée comme outil de spécification de contraintes sémantiques est plus facile d’utilisation afin de pouvoir être utilisé par un large public. D’autre part la granularité rencontrée dans les sites Web est une difficulté pour ce qui est de la vérification globale d’un site. Nous essayons donc de trouver un langage d’expression de contraintes sémantiques, reposant sur la sémantique naturelle, mais qui puisse répondre à ces deux principales contraintes.

6.5.3 Aide à la vérification sémantique de services Web adaptatifs

L. Rineau [18] a, d’une part, étudié un problème de formalisation de profils utilisateur et d’adéquation du contenu de pages Web à un profil particulier. Pour cela, il a utilisé Centaur pour donner une définition des langages XML et RDF. D’autre part, il s’est intéressé à la notion de navigation dans un site Web et a pour cela étendu la formalisation de pages Web sous forme de grammaires donnée par [FDS⁺99].

4. Nous utilisons Eclipse d’Imperial College <http://www.icparc.ic.ac.uk/eclipse/>

[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.

[FDS⁺99] J.-M. FARINONE, D. DRECQ, B. SUZARNES, A. SAVOY, J.-M. ABISCOR, «Une formalisation des sites Web sous forme de grammaires», *in: Proceedings of Hypertexte & Hypermedia : Products, Tools and Methods (H2PTM’99)*, p. 311–324, août 1999. Paris.

6.6 Analyse des usages

Mots clés : Web Mining, user behaviour, prediction, filtrage collaboratif, case-based reasoning, RàPC, raisonnement à partir de cas, Web usage mining.

Participants : Gilles Dodinet, Cedric Sempé, Brigitte Trousse.

Il s'agit de capitaliser nos différentes expériences d'analyse de l'usage sur le Web ou dans un site Web réalisées ces deux dernières années (contrat FT CTI, prédiction du comportement utilisateur [14], etc.). Une étude est actuellement menée depuis octobre sur des outils génériques d'analyse de logs HTTP en termes de sessions utilisateurs et en explicitant des critères objectifs de satisfaction utilisateur (selon le point de vue du concepteur).

7 Contrats industriels (nationaux, européens et internationaux)

7.1 Contrat CEA Cadarache

Participante : Brigitte Trousse.

Contrat INRIA/CEA (1/10/1999-31/1/2000).

L'étude proposée s'inscrit dans le cadre du projet READ (Retour d'Expérience Appliqué au Démantèlement) du CEA pour la DGD (Direction de la Gestion des Déchets). L'approche adoptée par le CEA dans le projet READ s'appuie sur le raisonnement à partir de cas (RàPC). Nous avons donc spécifié et réalisé à partir de CBR*Tools une maquette de système RàPC (appelée O*READ) d'aide à la réutilisation d'anciennes expériences liées au choix des outils lors de projets de démantèlement. Enfin, nous avons développé un démonstrateur dans lequel s'inscrit cette maquette afin d'illustrer l'interaction homme-machine envisagée en relation avec le logiciel du CEA Rexdin.

7.2 Contrat France Télécom R&D (CTI)

Participants : Attila Benedek, Semi Gaieb, Michel Jaczynski, Bernard Senach, Brigitte Trousse.

Contrat Ref INRIA 1 98 E 356, Ref FT R&D 98 1B 244.

Les objectifs pratiques de ce CTI visent l'amélioration de la qualité de l'aide fournie pour la navigation sur un site France Telecom en s'appuyant sur la réutilisation de navigations passées d'utilisateurs. Après une analyse des besoins et une étude théorique sur l'intérêt de techniques d'apprentissage symbolique pour la classe de problèmes visée par notre approche Broadway appliquée au Web, nous avons en 1999 spécifié l'assistant Broadway V2 et son intégration dans un clone d'un site de France Telecom, système en cours de réalisation. Nous rappelons que les spécifications du système de recommandations Broadway V2 s'appuient sur l'application de notre modèle d'indexation par situations comportementales, issu des travaux de thèse de

Jaczynski ^[Jac98], à l'application choisie. Cinq rapports de contrat (confidentiels) ont été livrés en 2000 à France Télécom relatifs aux lots 4 et 5 de ce CTI. Le lot 4 concernait la réalisation du système avec CBR*Tools et Broadway*Tools et le lot 5 une évaluation de l'aide apportée par le système auprès des utilisateurs ainsi qu'une étude liée à une éventuelle montée en charge. Une collaboration a eu lieu avec Y. Lechevallier pour ses compétences en statistiques.

7.3 Contacts industriels

Participante : Brigitte Trousse.

Un contrat de recherche est en cours de discussion avec Alcatel (Belgique) sur la conception et la réalisation d'un service Web adaptatif basé sur l'approche Broadway intégrant dans son raisonnement des profils utilisateurs définis à partir du modèle RDF d'Alcatel.

Nous avons présenté les recherches de l'équipe le 5 octobre, à trois membres d'une délégation du laboratoire de recherche de DENSO (Mrs Namba, Ueno et Taguchi).

8 Actions régionales, nationales et internationales

8.1 Actions régionales

Citons les collaborations :

- INRIA Sophia Antipolis: A. Giboin (ACACIA) autour d'Hermès, J-L Szyrka (Semir) autour des problèmes de l'analyse de logs HTTP.
- Ergomatics (Aix): B. Senach dans le cadre d'une coopération sur une tâche d'expérimentation du système livrée à France Télécom (cf. section).
- des contacts ont eu lieu avec la société Technofi (Sophia Antipolis) ainsi qu'avec Scorex (Monaco) pour un co-encadrement de stage ESSI.

De plus, dans le cadre d'une réflexion sur l'utilisation de notre approche Broadway dans le milieu éducatif, nous avons visité le Collège du Beausset (Var) et discuté plus particulièrement avec divers enseignants se rendant à l'atelier informatique (resp N. Hermitte) pour mieux comprendre les difficultés rencontrées lors de recherches d'informations dans le milieu éducatif.

8.2 Actions nationales

8.2.1 GDR-PRC I3

B. Trousse est membre du GDR-PRC I3 (Information - Interaction - Intelligence) sur les axes de travail: systèmes multi-agents (axe 7.3), modèles de raisonnement (axe 1.1), aide à la décision, ingénierie des connaissances (axe 6.1). Notre appartenance aux trois derniers axes

[Jac98] M. JACZYNSKI, *Modèle et plate-forme à objets pour l'indexation des cas par situation comportementales: application à l'assistance à la navigation sur le Web*, thèse de doctorat, université de Nice-Sophia Antipolis, Sophia-Antipolis, décembre 1998.

cités est due au fait que le RàPC est un axe transversal :

<http://www-timc.imag.fr/I3/>.

B. Trousse appartient au groupe national de raisonnement à partir de cas qui organise chaque année un séminaire de recherche :

<http://www.cpe.fr/~cbrfrance/>

8.2.2 Laboratoire des Usages, CNRT, Sophia Antipolis

Dans le cadre d'un projet de laboratoire des usages pluridisciplinaire au sein du CNRT, des premières journées ont été organisées par C. Guéguen et C. Charbit le 24 novembre à EURECOM (Sophia Antipolis). B. Trousse y a présenté les travaux menés dans AID i.e. sur les méthodes et outils d'aide à la conception de sites/services Web adaptatifs et plus particulièrement sur nos travaux sur l'analyse et la réutilisation de l'usage.

8.2.3 Site Web INRIA

Nous suivons, depuis novembre 1998, le groupe de travail sur la restructuration du site public INRIA de la Direction de la Communication, Cette année nous avons initié une coopération avec J-C Le Moal, P. Jancène et J-L Szyrka en vue d'appliquer nos travaux sur l'analyse des logs HTTP à une partie du site national.

8.2.4 Collaborations

Outre les collaborations régionales, nous avons eu des collaborations avec Yves Lechevallier de l'INRIA Rocquencourt au niveau de nos recherches sur l'indexation des sessions utilisateurs avec une approche neuronale (cf. section 6.2.1) et sur l'évaluation de l'approche Broadway dans le cadre du contrat FT-CTI (cf. section 6.3.1).

8.3 Actions internationales

8.3.1 Europe

- IST : 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).

9 Diffusion de résultats

9.1 Animation de la communauté scientifique

9.1.1 Comités de programme

B. Trousse est membre des comités de programme de conférences ou séminaires suivants :

- **au niveau national :** IAD&SMA'00, *7 èmes Journées Francophones Intelligence Artificielle Distribuée & Systèmes Multi Agents*, La Réunion, novembre 1999.
- **au niveau international :**
 - EWCBR'00 : 5th European Workshop on Case-based Reasoning, Trento, Italie, 6-9 septembre;
 - AID'00 : *5th International Conference on Artificial Intelligence in Design*, Worcester (USA), 26-29 juin (en tant que "chair pour la France").

9.1.2 Visites

L'action AID a reçu les visites suivantes :

- Prof Viorel Negru (West Timosoara University);
- Prof Alain Vom Hofe (Université de Rouen, UFR de psychologie, Sociologie et sciences de l'éducation) sur des aspects ergonomiques de la navigation sur le Web;
- Thierry Baccino de l'UNSA (dept Psychologie);
- Mr Mabillet, membre du cabinet du premier ministre (le 9 novembre).

9.1.3 Serveur WWW

L'action AID a poursuivi la mise au point de son site WWW qui permet d'accéder à un certain nombre d'informations relatives à nos recherches et en particulier aux deux logiciels Broadway et Hermès :

<http://www-sop.inria.fr/aid/> Une machine pour nos expérimentations sur le World Wide Web (Hermès et Broadway) est installée à l'Inria :

<http://aid.inria.fr/>

9.1.4 Divers

B. Trousse a été membre du bureau de l'Association Française d'Intelligence Artificielle (AFIA) jusqu'à juin 2000.

T. Despeyroux est président de l'AGOS, membre suppléant 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 faisons partie de l'équipe enseignante du DEA d'Informatique (resp Mr Kounalis) et celle du DESS Ergonomie et NTIC (resps T. Baccino et J. Araszkievitz) de l'université de Nice-Sophia Antipolis.

- Intervention (4h) dans le cours du tronc commun (TC4) “Langages et Modèles à objets” (resp M. Blay) du DEA Informatique à l'UNSA sur les *frameworks*: B. Trousse (travail collaboratif avec Colette Michel, UNSA).
- co-responsables d'un module optionnel (O15) sur *Sémantique et Conception de sites Web* (30h) dans le cadre des modules optionnels du DEA Informatique proposés depuis 1999 : T. Despeyroux et B. Trousse.
- Intervention (3h) sur le *raisonnement à partir de cas* dans un module optionnel (O11) du DEA Informatique (responsable : O. Corby) sur “Modélisation et capitalisation des connaissances” : B. Trousse.
- Intervention (18h) dans deux cours du DESS Ergonomie et NTIC sur la *conception de sites Web évolutifs* et sur la *conception de sites/services adaptatifs d'aide à la recherche d'informations*: B. Trousse.
- TP (92h) d'*algorithmique & programmation en java* à l'ESINSA : S. Gaieb.

9.2.2 Thèses

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

Thèse en cours :

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

Brigitte Trousse a été membre du jury de thèse de Mr V. Popov (université de Poitiers), directeur Mr Lienhart sur le sujet *Contribution à la représentation et à la résolution du problème de conception déclarative de scènes 3D en synthèse d'image*.

9.2.3 Stages

Nous avons accueilli 10 étudiants correspondant aux sept projets/stages suivants (ordre chronologique) :

- **Hakim Maadadi** et **Maamar Maameri**, étudiants de l'ESSI 3ème année, ont effectué leur projet Essi du 15 novembre 1999 à fin Mars 2000 sur la spécification et la réalisation d'outils d'analyse pour la montée en charge des systèmes basés sur l'approche Broadway pour l'aide à la navigation dans un annuaire thématique [16] ;

- **Adrien Zamani** étudiant de l’ESSI a fait son projet ESSI sur les “spécifications d’Hermès en ROL” dont l’objectif visait à évaluer le langage ROL [19] ;
- **Attila Benedek**, étudiant master de West Timisoara University (dans le cadre du programme SOCRATES a effectué son stage de master du 1 avril à mi-août) sur l’utilisation de l’approche neuronale pour l’indexation des cas pour nos application Broadway. Nous avons accueilli cet étudiant à nouveau jusqu’à fin décembre pour poursuivre le travail commencé [15] ;
- **Laurent Rineau**, étudiant de l’ENS (Ulm), a effectué son stage de 1ère année d’une durée de 4 mois à compter de juin sur l’étude des aspects “sémantique des sites Web du point de vue génie logiciel [18] ;
- **Anshuman Rai**, étudiant de l’ITT Kanpur (Inde) a effectué un stage sur l’étude des approches génie logiciel pour l’analyse sémantique et la vérification de sites Web [17] ;
- **Ferdinand Dosser** et **Laurent Perrimond**, étudiants de l’ESSI 3 ème année, effectuent leur projet Essi depuis le 16 octobre 2000 ;
- **Cédric Sempe** et **Gilles Dodinet**, étudiants de l’ESSI 3ème année, effectuent leur projet Essi depuis le 16 octobre 2000 sur des 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 <http://www.inria.fr>

9.3 Participation à des colloques, séminaires

Nous avons présenté nos travaux de recherche dans des séminaires et conférences nationales et internationales : on se rapportera à la bibliographie pour en avoir la liste.

B. Trousse a présenté les recherches menées dans AID au séminaire I3S/INRIA et a donné deux conférences invitées :

- le 22 septembre, à la villa Arson (Nice) dans le cadre des premières rencontres du CEDAR (Coordination des écoles d’Art en Réseau) <http://www.491.org/cedar/> regroupant des professeurs au niveau national issus de structures pédagogiques diverses ;
- le 7 décembre, dans un séminaire du groupe de travail “Logiciel et Statistique” de la Société Française de Statistique à l’IHP Paris.

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, Irlande, 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, à paraître 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 comportementales: 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 AAAI-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 (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>.
- [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.

Communications à des congrès, colloques, etc.

- [12] T. DESPEYROUX, B. TROUSSE, « Semantic Verification of Web Sites Using Natural Semantics », in : *RIA0 2000, 6th Conference on "Content-Based Multimedia Information Access"*, College de France, Paris, France, 1, p. 86–95, avril 2000.
- [13] M. GAIEB, « Reactive Object language », in : *Actes des Journées des jeunes Chercheurs en Système JCS'00*, 2000.
- [14] B. TROUSSE, « Evaluation of the Prediction Capability of a User behaviour Mining Approach for Adaptive Web Sites », in : *RIA0 2000, 6th Conference on "Content-Based Multimedia Information Access"*, College de France, Paris, France, 2, p. 1752–1761, avril 2000.

Divers

- [15] A. BENEDEK, « Analysis of Different Memory Indexing Models for Case-Based Reasoning, particularly the Artificial Neural Network-based Approach », décembre 2000, rapport de stage DEA, West University of Timisoara (Romania), ESSI et Action AID (Inria), 61 pages.
- [16] M. MAAMERI, H. MAADADI, « Outils d'analyse des performance d'un système basée sur l'approche Broadway », avril 2000, mémoire de stage ESSI, UNSA et Action AID (Inria).
- [17] A. RAI, « A Study of software Engineering Approaches for Semantic Analysis and verification of Web Sites », septembre 2000, Internship Report, Action AID (Inria), 63 pages.
- [18] L. RINEAU, « Étude des aspects dynamiques des sites Web du point de vue génie logiciel », septembre 2000, rapport de stage ENS (Ulm) et Action AID, 65 pages.
- [19] A. ZAMANI, « Spécification d'Hermès en ROL », 2000, rapport de projet ESSI et Action AID (Inria).