

MAREL

Models and Reuse Engineering, Languages

Effectifs au 30/06/2013 :
7 permanents (3,5 ETP)
12 doctorants (12 ETP)

Nombre de thèses soutenues
entre le 01/01/2008
et le 30/06/2013 : 8

Responsable :
Abdelhak-Djamel Seriai

Page Internet de l'équipe :
<http://www.lirmm.fr/recherche/equipes/marel>

INGÉNIERIE DIRIGÉE PAR LES MODÈLES, LANGAGES DE PROGRAMMATION, RÉUTILISATION, RÉINGÉNIERIE, ANALYSE FORMELLE DE CONCEPTS, COMPILATION, OBJETS, COMPOSANTS, SERVICES WEB, LIGNES DE PRODUITS LOGICIELS

Présentation

Les travaux de MaREL s'insèrent dans le cadre du Génie Logiciel et s'intéressent à l'automatisation des étapes du cycle de vie du logiciel, de la conception à la maintenance en passant par la compilation et l'optimisation du code. Ces études couvrent la totalité de la démarche, de l'étude théorique aux expérimentations en vraie grandeur, en passant par le développement d'algorithmes et de programmes. L'activité de l'équipe se concentre sur trois axes :

- l'Ingénierie Dirigée par les Modèles (IDM), en association avec une approche formelle originale, l'Analyse Relationnelle de Concepts (ARC), qui est expérimentée sur des données réelles et qui sert pour l'apprentissage de transformations de modèles ;
- l'ingénierie de la réutilisation et de la variabilité, par l'extraction de composants ou de lignes de produits à partir de systèmes orientés objets existants en se basant sur des métriques adaptées et des algorithmes d'optimisation combinatoire ;
- la sémantique et l'implémentation des langages de programmation, à objets ou à composants, pour concevoir un langage «pur composant», comme Smalltalk est «pur objet», ou pour expérimenter des techniques d'implémentation des langages à objets.

Ces trois axes sont à la fois bien identifiés et étroitement associés. Les notions d'objet ou de modèle constituent un socle commun, et les composants sont étudiés à la fois du point de vue des langages de programmation, des modèles et de leur extraction à partir du code. L'ARC est un objet d'étude, aussi bien qu'une base formelle pour le développement d'outils.

Evolution de l'équipe

L'évolution principale de l'équipe se caractérise par son recentrage sur les thématiques du Génie Logiciel à la suite de départs vers des structures voisines :

- Thérèse Libourel et Isabelle Mougenot sont parties (en 2009) fonder l'UMR Espace-Dev de l'IRD : la thématique Système d'Information n'est donc plus représentée dans l'équipe ;
- Suite à la retraite de Claude Laurenço, DR CNRS, la thématique d'application à la chimie a été abandonnée et Philippe Vismara a rejoint (en 2010) l'équipe COCONUT.

En conséquence, de DOC (Données, Objets, Composants), l'équipe est devenue MaREL. En parallèle, Abdelhak-Djamel Seriai a pris la succession de Christophe Dony à sa tête.

Organisation et Vie de l'équipe

- Réunion hebdomadaire pour séminaires, discussions recherche et organisationnelles, et depuis septembre 2012, pour préparer l'organisation des conférences ECOOP/ECSA/ECMFA.
- Séminaires réguliers, avec exposés des membres de l'équipe, en particulier des doctorants, d'anciens doctorants, et de chercheurs invités : Jan Vransy (Czech Technical University in Prague), J-P Rigault (Univ. Nice Sophia-Antipolis), J. Malenfant (Univ. Paris 6), M. Acher (Univ. Lille), J. Laval (Univ. Bordeaux), L. Deruelle (Univ. Littoral).
- Pour la gestion budgétaire, 50% des crédits non dédiés (hors ANR par ex.) sont affectés à un compte commun.

Activités scientifiques

Les résultats majeurs de MaREL concernent plusieurs thèmes.

Implémentation des langages à objets. Technique à base de hachage parfait et plateforme de tests PRM

Les langages à objets reposent sur des mécanismes coûteux pour lesquels il n'existait pas d'implémentation vérifiant des conditions de passage à l'échelle (temps constant, espace linéaire) dans un contexte de chargement dynamique et d'héritage multiple. Nous avons conçu une technique à base de hachage parfait (hachage sans collision), avec étude théorique et simulation sur de grosses hiérarchies de classes [voir publication majeure n°1]. En parallèle, nous avons développé un compilateur autogène (bootstrappé) du langage PRM (Program with Refinement and Modules) qui est le langage de programmation à objets développé dans l'équipe à partir de 2002. Ce compilateur a été utilisé comme banc de test pour évaluer et comparer toutes les techniques connues d'implémentation des objets dans un cadre de compilation statique (sans recompilation) [DMPoops1a09].

Le hachage parfait est la première technique passant à l'échelle dans un contexte de typage statique et d'héritage multiple ; elle est applicable à Java par exemple. Les expérimentations sur le compilateur PRM sont les premières à comparer systématiquement les différentes implémentations connues sur un programme réel, pour analyser leurs coûts respectifs.

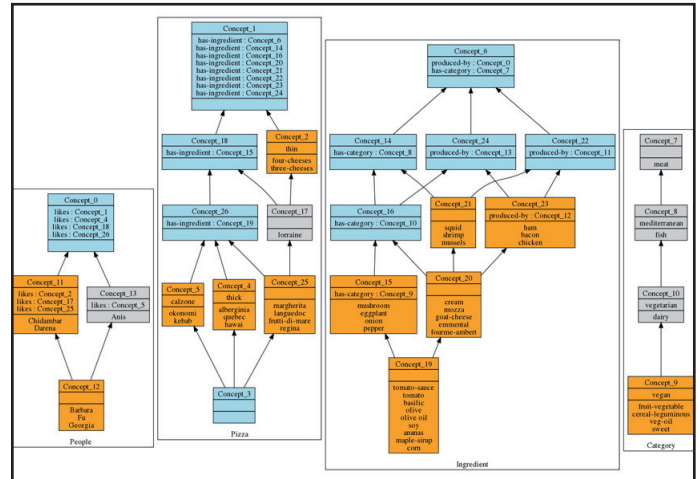
Dans le cadre d'une convention CIFRE avec CORTUS, nous avons développé des optimisations pour l'implémentation de la généricité en .NET, pour des systèmes embarqués [SDlctes12].

La perspective est de passer à une utilisation du hachage parfait dans un contexte de machine virtuelle et compilation dynamique dans une collaboration UQAM-LaBRI avec dépôt d'un PICS.

Aspects théoriques de l'ARC et applications au génie logiciel

L'Analyse Relationnelle de Concepts (ARC) est une méthode d'analyse de données, extension originale de l'Analyse Formelle de Concepts (AFC) et applicable à des données relationnelles. Nous avons approfondi certains aspects théoriques de l'ARC en collaboration avec des chercheurs du LORIA, de l'UQAM et du LIMOS [voir publication majeure n°2]. Nous avons proposé une nouvelle formalisation par une suite convergente de contextes, une méthode de navigation des familles de treillis par des requêtes et nous avons conçu un nouvel algorithme pour construire des AOC-posets (sous-ordre caractéristique du treillis). Nous avons également développé un corpus d'applications de l'ARC au génie logiciel. Parmi les plus abouties, on peut noter la réingénierie de modèles de classes (avec IRSTEA) [AOMLHNiceis12], la réingénierie de diagrammes de

cas d'utilisation [DHNRFi12], la classification de services web (utilisant leur position dans un workflow de tâches) [ADHMTicws11] et l'apprentissage de transformations de modèles, mentionné plus loin. Des travaux en cours concernent la remodularisation de logiciel (ANR CUTTER). La collaboration interdisciplinaire avec IRSTEA nous fournit des données réelles en vraie grandeur pour la réingénierie de modèles de classes.



Treillis de concepts construits par ARC sur le schéma relationnel « des personnes mangeant des pizzas aux ingrédients de diverses catégories produits par des personnes »

Deux outils sont disponibles : **AOC-poset Builder** et **ERCA**.

La perspective est l'étude du passage à l'échelle et la conception de méthodes incrémentales, sachant que dans le pire des cas la méthode crée un nombre exponentiel de concepts.

Génération de transformations de modèles à partir d'exemples de modèles sources et de modèles transformés

L'Ingénierie Dirigée par les Modèles (IDM) promeut dans les processus de développement l'utilisation massive de modèles, structurés par d'autres modèles nommés métamodèles. Les modèles peuvent alors être manipulés par des programmes appelés transformations de modèles. L'équipe a diverses contributions dans ce domaine, telles que la traçabilité des transformations [ALCNedoc10] ou l'alignement de modèles issus de transformations [DDFHNpecmf11].

L'équipe s'est également engagée dans un champ de recherche qui a émergé récemment et qui a pour objectif de faciliter la conception des transformations de modèles, par apprentissage de ces transformations à partir d'exemples.

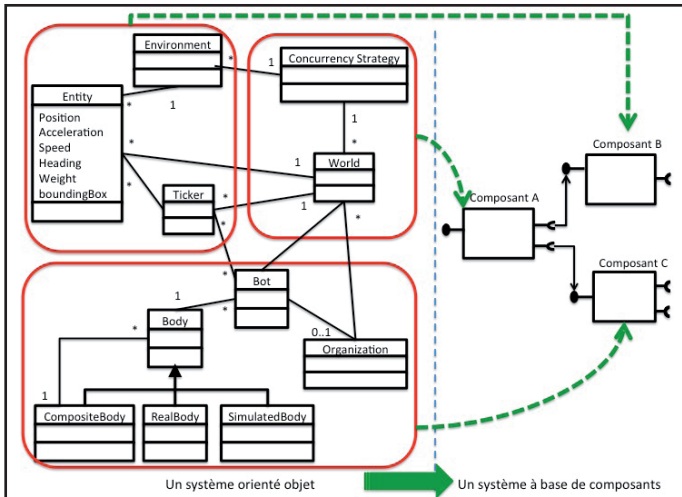
L'approche que nous avons proposée [voir publication majeure n°5] repose sur un apprentissage basé sur l'Analyse Relationnelle de Concept (ARC), une extension de l'Analyse Formelle de Concepts, qui permet une classification des différents éléments de modèle intervenant dans les exemples.

C'est l'approche de génération de transformation de modèles à partir d'exemples la plus automatisée. Elle

repose sur des bases formelles pour l'apprentissage (ARC), et elle est indépendante vis-à-vis des métamodèles et des outils de modélisation.

La perspective est une collaboration avec les équipes COCONUT-MAORE-MAB sur la génération automatique de modèles permettant de présenter de nouveaux exemples (ou contre-exemples) à l'utilisateur.

Réingénierie des Systèmes Orientés Objets (SOO) vers les composants



Identification des composants comme ensemble de classes.

Le paradigme à base de composants constitue une évolution importante de l'orienté objet. Il permet de mieux maîtriser la complexité et la maintenance des logiciels à un niveau de granularité plus important que les objets. Pour répondre aux besoins de réingénierie de SOO vers les composants, nous avons proposé une approche permettant l'extraction de l'architecture à base de composants d'un SOO via une analyse statique de son code source [voir publication majeure n° 3] et des documents de conception [CSecca10]. Cette approche est basée sur :

- Un modèle de mesure de la qualité des composants orientés objets. Ce modèle est basé sur l'identification des caractéristiques principales d'un composant orienté objet (composabilité, spécificité et autonomie) et le raffinement de ces caractéristiques en propriétés et métriques (telles que le couplage et la cohésion des classes qui constituent la structure interne d'un composant).
- Une modélisation de l'identification de composants à partir d'un SOO comme un problème d'optimisation. La fonction objectif à optimiser représente la qualité d'une architecture à base de composants. Nous avons utilisé principalement deux méta-heuristiques (algorithmes génétiques [SCseke11] et recuit simulé). L'utilisation d'autres techniques telles que l'analyse formelle de concepts [HScla10] et le regroupement hiérarchique [KSc3s2e12] a été également explorée.
- L'approche a été expérimentée sur 11 SOO de différentes tailles tels que ArgoUML (+2000 classes) et Jigsaw (+300 classes) et les résultats ont été comparés à ceux d'une réingénierie manuelle.

Dans la réingénierie des SOO vers les composants, c'est la seule approche basée sur un modèle de mesure de la qualité des composants orientés objet. Ce modèle de mesure est inspiré de la norme ISO 9126 définissant des indicateurs pour la qualité logicielle. C'est aussi la seule approche basée sur l'exploitation conjointe de l'analyse statique du code source et des algorithmes méta-heuristiques.

Les perspectives concernent l'identification de composants réutilisables dans les API OO dans une collaboration avec H. Sahraoui (Univ. de Montréal), la réingénierie de variantes de SOO en lignes de produits à base de composants, l'étude de l'automatisation du processus de développement basée sur des lignes de produits conçues dans une approche IDM (thèse CIFRE avec l'entreprise Acelys).

Développement de langages « pur composant »

Le paradigme des composants repose sur les notions de composant (brique logicielle) communiquant au travers de ports spécifiés par des interfaces décrivant des services requis ou fournis par le composant. Ces composants s'assemblent au sein d'architectures par le biais de connecteurs. L'équipe effectue divers travaux dans ce domaine et a également effectué des incursions dans le domaine des systèmes multi-agents et de leur fiabilisation [DTUVweh08] (ANR FACOMA).

Une direction principale consiste à développer un langage à composants « pur », c'est-à-dire construit autant que possible autour des principales abstractions du paradigme. Le premier résultat a été la conception du langage SCL (Simple Component Language) [FBDHclss12] qui s'est poursuivie autour de COMPO un langage de développement intégrant un système d'héritage complet de niveau composant, autorisant la modélisation d'architectures et la programmation dans le même univers conceptuel, et réflexif via une réification originale en composants [SDTFgpce12]. Il permet la réalisation d'applications (éventuellement auto-adaptables), de programmes de vérification d'architectures ou de transformation de modèles (y compris à l'exécution). Autour de COMPO ont été réalisées des études connexes dont le développement de solutions pour exprimer les contraintes d'architecture sous la forme de composants [voir publication majeure n°4] et pour migrer les applications à base de composants en services Web [TKicws12]. Nous proposons globalement une solution unifiée permettant la programmation, la description d'architecture, l'ingénierie dirigée par les modèles, et offrant un environnement opérationnel original de type «models@runtime» (modèles exécutables) qui ne passe pas par une compilation vers une couche objet. COMPO le premier langage de modélisation et programmation «pur composant» avec héritage et contraintes d'architectures, premier langage réflexif avec réification par composants. Les perspectives concernent des traits majeurs des ADL (Architecture Description Language) orientés composants dans COMPO, supprimant de facto les problèmes de conformité «spécification-code», ainsi que la possibilité d'expression de toutes sortes de transformations de modèles.

Valorisation

Valorisation industrielle

MaREL se caractérise par une grande implication dans la valorisation à finalité interdisciplinaire (avec IRSTEA/TETIS ou le LHYGES par exemple) ou industrielle :

- Avec Orange Labs (et l'équipe TEXTE) : un contrat de recherche externalisée sur la réingénierie des modèles de classes.
- Avec Cortus (microélectronique, Montpellier) : 2 CIFRE successives sur le développement d'un compilateur .NET adapté aux processeurs embarqués low-end, puis sur l'application au multi-thread et multi-coeur.
- Avec Thales (Paris-Palaiseau) : 1 CIFRE sur l'étude de la variabilité dynamique des systèmes logiciels complexes adaptables.
- Avec Acelys (SSII, Montpellier) : 1 CIFRE sur l'automatisation de lignes de produits logiciels.
- Dans le cadre du projet FUI-RIDER coordonné par IBM : des contributions sur les lignes de produits logicielles (notamment adaptation d'un modèle de features au contexte).

Collaboration interdisciplinaire

- Nous travaillons avec IRSTEA à des expérimentations sur le suivi de l'évolution et la réingénierie de modèles de classes par l'approche AFC/ARC. IRSTEA dispose de 15 versions d'un modèle de classes pour un système d'information sur les pesticides.
- Avec le CIRAD (et l'équipe TEXTE), la collaboration porte sur des modèles pour la représentation et le traitement des savoirs socio-écologiques et socio-agronomiques.
- Avec le LHYGES (et l'équipe TATOO), nous travaillons avec l'ARC sur l'analyse de données issues de cours d'eau en vue d'évaluer et de suivre leur qualité hydrobiologique (ANR FRESQUEAU).

Participation à des comités d'expertise de travaux industriels

- Participation au Conseil d'Orientation Scientifique, Technique et Industriel (COSTI) de Transfert-LR.

Faits marquants

Organisation de **ECOOP/ECSA/ECMFA** à Montpellier en juillet 2013.

Ces trois conférences européennes majeures dans les domaines connexes de la programmation par objets (ECOOP), de la modélisation (ECMFA) et des architectures logicielles (ECSA) ont rassemblé plus de 400 personnes. Les thématiques de ces trois conférences couvrent assez précisément les thématiques de l'équipe MaREL.

Organisation de **7 évènements scientifiques, ateliers internationaux ou conférences nationales** (ICOOLPS 2008-2013, CAL 20012, WEH 2012, SAVA 2011, SATTtoSE 2010, MASPEGHI 2010, SASE 2008).

ACM SIGSOFT Distinguished Paper Award à CBSE 2011 [voir publication majeure n°4]

Essaimage :

- 4 anciens docteurs recrutés comme enseignant-chercheur (UQAM, Univ. Bordeaux, EMD) ;
- 2 anciens docteurs recrutés en post-doc aux USA (Purdue, IBM TJ Watson) ;
- 2 anciens docteurs recrutés en post-doc en France ;
- 14 anciens stagiaires de Master en thèse.

Collaborations externes

- Université de Montréal (laboratoire GEODES) : co-encadrement en cours de deux thèses sur les thèmes de l'IDM et de la réutilisation.
- Ecoles des Mines d'Alès et de Douai : collaboration suivie depuis 2003 avec plusieurs co-encadrements de thèses dans le domaine des composants logiciels, de l'IDM, avec application à la domotique et participation conjointe à l'ANR FACOMA.
- LORIA, Nancy et LATECE, UQAM : classification de services web, projet PICS CADOE (ARC et ontologies).
- RMOD, INRIA Lille (et l'équipe MAORE) : projet ANR CUTTER (remodularisation de logiciels).
- LHYGES (et l'équipe TATOO) : projet ANR FRESQUEAU.

Publications majeures

[1] R. Ducournau, F. Morandat **Perfect class hashing and numbering for object-oriented implementation.** *Software Practice & Experience*, 41(6) 661-694, John Wiley & Sons, 2011.

[2] M. Rouane Hacene, M. Huchard, A. Napoli, P. Valtchev **Relational concept analysis: mining concept lattices from multi-relational data.** *Ann. Math. Artif. Intell.* 67(1): 81-108, 2013.

[3] S. Kebir, A.-D. Seriai, S. Chardigny, A. Chaoui **Quality-Centric Approach for Software Component Identification from Object-Oriented Code.** The 10th Working IEEE/IFIP Conf. on Software Architecture (WICSA) and the 6th Eur. Conf. on Software Architecture (ECSA) - WICSA/ECSA 2012. pp. 181-190, IEEE, 2012.

[4] C. Tibermacine, S. Sadou, C. Dony and L. Fabresse **Component-based Specification of Software Architecture Constraints.** The 14th Int. ACM SIGSOFT Symp. on Component-Based Software Engineering (CBSE'11), pp. 31-40, 2011.

[5] H. Saada, X. Dolques, M. Huchard, C. Nebut, H. Sahraoui **Generation of operational transformation rules from examples of model transformations.** The 15th ACM/IEEE Int. Conf. on Model Driven Engineering Languages & Systems (MODELS'12). LNCS 7590, p. 546-561, 2012.