

Compte rendu réunion SIDO du jeudi 12 février 2009

Présents : A. Bouju, G. Hillairet, MC Lafaye, JY. Lafaye, G. Louis, J. Malki, W. Mefteh, V. Noyon. - Exc. : C. Demko, R. Mullot

Le point essentiel de l'ordre du jour concerne la définition des éléments structurants de la recherche SIDO et les contributions à la capitalisation.

Le CR ne rend pas compte des débats, mais tente de proposer une synthèse des faits, des idées et des conclusions.

Thématiques SIDO

L'acronyme SIDO signifie « Sémantique et Intermédiation de Données ». Actuellement, nous œuvrons plutôt à l'intégration de données qu'à l'intermédiation mais le programme ainsi évoqué préserve l'avenir et représente honnêtement le cœur de nos centres d'intérêt.

Selon un point de vue très général, SIDO traite de l'ensemble des représentations et des processus de transformation de données et d'informations à travers trois espaces techniques : persistance (CWM), structure (UML), publication dans le cadre du web sémantique (OWL). Dans chaque espace, on s'intéresse aux langages, méthodes et outils de représentation et de manipulation des données et des contraintes d'intégrité. Pour chaque couple d'espaces on s'intéresse de la même manière aux langages, méthodes et outils permettant la transformation des données et des contraintes d'un élément du couple à l'autre.

La philosophie SIDO est de placer l'espace de modélisation UML des informations métier au centre de l'ensemble des stratégies étudiées. Ainsi on étudiera les ponts entre l'espace de persistance et l'espace des modèles métiers d'une part, et entre l'espace des modèles métiers et l'espace ontologique d'autre part. La correspondance entre persistance et ontologie est un résultat dérivé par composition des deux premiers.

Nos travaux se placent donc dans les champs de l'ORM (Object Relational Matching) et OOM (Object/Ontology Matching) avec par transitivité un produit dans le champ de l'ROM (Relational/Ontology Matching).

L'approche que nous appliquons relève de l'Ingénierie Dirigée par les Modèles, du tissage et de la transformation de modèles. Les domaines d'applications développées sont variés mais les aspects spatio-temporel, géographique, document numérique, image & comportement sont des exemples privilégiés.

Activités de SIDO aujourd'hui

L'activité est de fait centrée sur les sujets de thèse et les contrats. Des évolutions sont naturellement envisageables et nécessaires, mais la cartographie actuelle des recherches indique les priorités.

▪ *Thème_Web_Sémantique (FB, JYL, GH)*

Deux principaux cas d'utilisation sont envisagés :

- Un acteur développe une application qui utilise d'une part des données conformes à un modèle métier et stockées dans des bases de données locales (relationnelles, XML,...), et d'autre part des données externes accessibles sur le web dans des bases de connaissances décrites par des ontologies OWL. Une variante de ce cas d'utilisation est celle où l'application

peut non seulement accéder à des données ontologiques, mais également modifier des bases distantes par insertion ou mise à jour d'individus.

- Le second cas d'utilisation est une extension du premier, l'application étant dédiée à l'import/export de données locales. On traite ainsi de la migration/transformation de données entre une structure de persistance relationnelle ou XML et une structure ontologique. Selon notre point de vue, le modèle Objet métier est au centre de la procédure, et nous nous intéressons uniquement à la migration entre données Objet et données Ontologiques, la partie ORM étant ou bien une application de techniques disponibles dans la littérature, ou bien étudiée dans le thème suivant (MCL, GL, AW).

Les sujets soulevés par l'implémentation de ces CU concernent la spécification de solutions génériques dont il convient de définir les modèles considérés en entrées, ceux considérés en sortie et les modèles de mappings correspondants. Les procédures concernent à la fois la traduction des structures abstraites (Objets/Ressources, Contraintes/Règles) que des données concrètes (Instances/Individus) et enfin que des requêtes exprimées dans différents langages (HQL/SPARQL) par exemple.

Une étape nécessaire et importante concerne l'aide à la génération d'ontologies. Pour ce faire on utilise des modèles d'annotation des modèles UML et les principes de traduction issus de QVT & ODM.

▪ *Thème_Persistance_Objets_Contraintes_Métier (MCL, GL, AW)*

Le Matching Objet Relationnel tel que proposé par l'environnement Hibernate connaît un succès indéniable en apportant des solutions efficaces selon un point de vue particulier sur la façon d'assurer la persistance des objets dans un SGBD. Il s'agit en fait de sérialisation d'objets et non de conception de bases de données.

Nous nous plaçons dans le cas où l'on cherche à concevoir une structure de persistance relationnelle robuste qui puisse être partagée par plusieurs applications métier différentes. Le cadre de notre travail impose que cette structure explicite toutes les contraintes de références, de clés, d'inclusion a priori implicites dans le modèle relationnel et héritées du modèle objet du domaine.

Nous souhaitons également que le nommage reflète le vocabulaire métier de manière à rendre évidente la sémantique et ainsi faciliter la maintenance et l'interrogation des SGBD.

Ces objectifs imposent d'annoter le modèle objet initial, de procéder à des transformations de modèles intermédiaires pour arriver finalement à un modèle relationnel conforme à nos spécifications qui raffinent le modèle CWM. le processus d'annotation fait intervenir et assiste une expertise humaine, après quoi le processus de calcul des clés, des autres contraintes et de l'architecture finale de la base est automatique grâce à un algorithme original que nous avons spécifié. Notre processus nous amène à relaxer certaines contraintes propres à la théorie classique de la normalisation, mais nous montrons ne pas introduire de redondance en dehors des clés, ce qui n'est pas pénalisant.

Une part importante de nos recherches concerne la spécification de modèles, méthodes et algorithmes pour prouver les qualités des solutions que nous proposons. Une autre est de contribuer au développement de prototypes pour les mettre en œuvre de façon ergonomique.

Un dernier point original dans notre approche est de nous intéresser à l'expression et au traitement de contraintes dynamiques (en cela qu'elles sont attachées aux états des objets considérés).

▪ ***Thème_Gestion_Information_SpatioTemporelle (AB, JM, WM, ZW, et al.)***

Notre travail porte initialement sur des informations géographiques, mais par extension, nous avons été amenés à nous intéresser de manière plus générale aux informations spatio-temporelles et comportementales. Dans un contexte applicatif, il nous est apparu que les informations géographiques sont souvent interrogées par des utilisateurs mobiles, et que l'adaptation de l'aspect graphique et de la gestion des données est primordial.

En conséquence, nous travaillons selon deux axes complémentaires :

- proposer des modèles ou des patrons spécifiques pour représenter et traiter les objets mobiles ainsi que les données afférentes (trajectoires, comportement)
- proposer des modèles de publication et de communication adaptés à la mobilité et aux différents supports de visualisation.

La problématique que nous abordons pose des questions spécifiques de persistance, d'indexation, de fusion de données hétérogènes multi-source, multi-résolution. Nous sommes donc amenés à traiter à la fois de la sémantique des données grâce à des ontologies *ad hoc* et aussi des contraintes techniques de bande passante de flux et d'adaptation de la visualisation.

Les solutions que nous apportons cherchent à apporter des solutions génériques, pour des systèmes d'information adaptés à la mobilité spatiale. Ceci passe par la conception d'ontologies à partir de modèles objet, le mapping et la transformation de modèles

Factorisation – Synergie

Chacun des trois thèmes évoqués ci-dessus s'inscrit dans une démarche concertée qui s'adresse à l'un des aspects du noyau de recherches de SIDO. A savoir

- le matching Objet/Ontologie avec comme application le web sémantique,
- le matching Objet/Relationnel dans une optique de qualité des schémas produits en terme d'ergonomie et de rigueur et d'exhaustivité dans l'expression des contraintes intrinsèques au modèle.
- l'application au domaine géographique et spatio-temporel visant à la modélisation du comportement d'objets dynamiques (objets et contraintes évolutifs).

L'approfondissement des connaissances dans le domaine des ontologies, de la modélisation, de l'interrogation et de l'expression de la sémantique forme le socle commun.

La proposition de solutions novatrices dans chacun des domaines spécifiques et la réutilisation des résultats de chaque thématique SIDO par les deux autres est le moyen de progresser de façon efficace.

Capitalisation

Plusieurs types de problèmes particuliers sont traités au sein de SIDO qui concernent de façon uniforme chacune de ses thématique. Beaucoup peuvent aussi intéresser ImagIN et IMedoc.

En interne à SIDO :

- Définition et exploitation de langages d'annotation de modèles
- Définition et exploitation de langages de mapping
 - Objet/Ontologie (+ génération d'ontologie à partir d'un modèle Objet)
 - Objet/Relationnel (+ génération d'un mod. relationnel à partir d'un mod. Obj.)
 - Ontologie/Ontologies
- Langages et outils de transformation et de tissage de modèles
- Langages d'expression de contrainte

En interne au L3i (collaboration avec ImagIN et IMedoc) :

- Production de modèles de comportement (objet et ontologie)
- Aide à la conception d'ontologies (document, comportement, objets géographiques,...)
- Aide au mapping et à la transformation de modèles

En externe :

- Présentation de vidéos/animations démontrant l'utilisation des techniques et prototypes développés sur des exemples didactiques

Le développement d'une synergie et la capitalisation supposent par voie de conséquence de préciser les conditions d'échange de modèles, de données, le choix des langages communs. KM3, XML, ATL...

L'essentiel du travail de SIDO porte sur les modèles et les méthodes. Les outils sont nécessaires pour étayer les propositions et prouver la faisabilité mais restent souvent un simple sous-produit des activités. La réalisation d'un prototype utilisable de façon « publique » (même experte) – et qui constitue de fait un élément de capitalisation - est d'un autre degré de difficulté et demande des moyens humains.

La volonté de capitalisation est présente et le besoin avéré. Des expériences ont déjà été menées (réutilisation du modèle de génération d'ontologies sur des données géographiques). D'autres seront relativement immédiates (utilisation de l'ORM pour compléter la chaîne de transformation avec l'OOM). De même, les échanges avec IMedoc (aide à l'indexation) ou ImagIn (raffinement des modèles de comportement, génération de code) semblent tout trouvés. Il reste à systématiser la pratique... et dégager les ressources évoquées plus haut.

En conclusion, les éléments de capitalisation de SIDO sont :

- *des modèles (modèles de mapping, de transformations, patterns,...)*
- *des langages (d'annotation, de mapping)*
- *des méthodes de conception de SGBD, d'ontologies*
- *des prototypes implémentant chacun des éléments des points précédents et d'autre visant à leur intégration*
- *des démonstrateurs sur des exemples (vidéos didactiques)*