

Approche sémantique pour la capitalisation des connaissances métier en conception de produits industriels

Mohamed-Foued Sriti⁽¹⁾, Phillippe Boutinaud⁽¹⁾, Nada Matta⁽²⁾ and Manuel Zacklad⁽²⁾

(1) *CADeSIS R&D – 142/176 avenue de Stalingrad – 92700 Colombes*

(2) *Laboratoire Tech-CICO – Université de Technologie de Troyes (UTT) – 12, rue de Marie Curie – 10010 Troyes cedex*

Résumé

Dans cet article nous présentons une approche de capitalisation de connaissances élaborée dans le contexte d'un projet de « documentarisation des systèmes de gestion des données techniques ». Le but est de faciliter le transfert des connaissances afin de les partager entre les ingénieurs-concepteurs et les échanger entre les agents logiciels. Dans le cadre de la conception de produits industriels, notre préoccupation principale est de doter les connaissances métier d'une représentation formelle pour rechercher et réutiliser plus pertinemment ces connaissances. Notre projet propose une infrastructure (I-Semantec) réutilisable intégrant les concepts du Web Socio-Sémantique et les ressources disponibles dans les systèmes de gestion des données techniques.

Keywords: Web Sémantique, W2S, Gestion des connaissances, Représentation des connaissances, SGDT, CAO.

1 Introduction

Le projet de documentarisation des systèmes de gestion des données techniques vise à intégrer dans une infrastructure réutilisable les concepts du Web Socio-Sémantique (W2S) et les ressources disponibles dans les Systèmes de Gestion des Données Techniques (SGDT).

Dans l'industrie, le processus de conception prend une part toujours plus importante dans les coûts de production. Le travail dans ce secteur utilise massivement l'informatique avec les logiciels de CAO et les SGDT. Des gisements de productivité importants se trouvent dans la réutilisation des documents de conception, fabrication, tests, etc. Au démarrage d'un nouveau projet, il serait critique de pouvoir trouver les fichiers et les informations les plus pertinents qui ont été produits et stockés dans le SGDT lors de projets antérieurs. Mais pour ceci, il est nécessaire de pouvoir accéder à l'ensemble des SGDT de l'entreprise et de ses partenaires. C'est aujourd'hui impossible tant la diversité des outils manipulés, des modèles conceptuels et des fichiers utilisés est grande. Dans ce contexte, le processus de documentarisation peut associer aux documents des attributs facilitant leur réutilisation en s'appuyant sur des modèles heuristiques de l'activité coopérative. Ces modèles permettent la définition de métadonnées basées sur l'approche du Web Socio-Sémantique (cf. infra & [CZM04]). Ce processus facilite la transférabilité des connaissances entre différents groupes de travail. La documentarisation intervient dans le contexte d'un usage collectif de documents qui constituent des Documents pour l'Action [Zac04], un concept émergent dans le champ du W2S.

Dans la suite, nous présentons l'apport incontournable des SGDT dans le processus industriel de développement des produits. Après avoir montré les limites des SGDT pour mieux gérer les connaissances, nous présenterons l'apport de l'approche du W2S pour dépasser ces limites. Nous présenterons ensuite, l'état d'avancement de nos travaux. En conclusion, nous discutons les résultats partiels et les perspectives.

2 SGDT à « mi-chemin » de la gestion des connaissances

2.1 La réutilisation en conception

La réutilisation des connaissances dans le processus de conception d'un produit est une stratégie majeure pour diminuer les coûts, réduire les délais et améliorer la qualité du produit.

Il existe actuellement des outils logiciels de qualité industrielle permettant de répondre aux différents problèmes liés à la réutilisation des quantités importantes de documents générés par une étude (document textuels, documents de conception, etc). Ces outils se sont les SGDT, grâce auxquels on identifié facilement des réponses adéquates aux questions suivantes : Quels sont les documents relatifs à un produit ? Qui a créé/modifié ces documents ? Quels sont les différents stades d'évolution des documents d'un produit en cours d'étude ? Comment partager ces documents dans l'entreprise et gérer les droits d'accès ? Quelle est la nomenclature d'un produit donné et éventuellement quelles sont ses variantes, diversités et effectivités. L'historique des modifications des documents relatifs à un produit est conservé tout au long de la vie du produit. Certaines fonctionnalités des SGDT sont très évoluées et permettent une collaboration forte autour d'un produit ; discussion en ligne, vidéo-conférence, gestion des cycles de vie, etc.... Nous pouvons donc structurer, centraliser et sécuriser les données relatives aux phases de développement et de conception de produits. Le SGDT évolue et ambitionne de couvrir tout le cycle de vie du produit, c'est-à-dire d'inclure les phases d'industrialisation et de distribution. Ceci implique de gérer la co-existence avec les applications de CAO, de gestion des documents, de gestion de ressources industrielles, de gestion de la chaîne logistique, etc.

Afin d'améliorer l'efficacité des bureaux d'études, d'optimiser la réutilisation d'études existantes et d'améliorer la formation de leurs collaborateurs, les industriels sont maintenant demandeurs de solutions qui les aideront à répondre principalement à la question : « une étude est démarrée, quelles sont les études existantes qui ont déjà apporté des solutions partielles ou totales au problème actuel ? » Il s'agit donc de gérer la connaissance métier dans l'entreprise.

2.2 Les limites fonctionnelles et organisationnelles

D'un point de vue industriel, les SGDT sont considérés parfois comme des systèmes de gestion des connaissances quand il s'agit d'optimiser les liens et les relations entre ressources produites par les différents collaborateurs [Cat01]. Les SGDT les plus utilisés sont : Enovia-VPM, MatrixOne, TeamCenter Engineering ou Enterprise et SMARTEAM. Tous ces systèmes sont basés sur le même principe où les métadonnées sont gérées par une base de données relationnelle, indexant les fichiers utilisés par différents logiciels de conception de l'entreprise. L'obstacle habituel lorsque l'on veut faire de la gestion des connaissances avec ces outils est que l'organisation des métadonnées est figée dans le modèle de la base de données. C'est, en général, une opération de grande envergure de modifier ne serait-ce que le type d'un attribut de la base.

Dans des travaux antérieurs¹ nous avons essayé de répondre à la question posée précédemment et nous avons identifié les obstacles techniques liés aux SGDT : rigidité, difficulté de migration et d'interopérabilité, recherches basées sur des attributs existants et/ou sur mots-clefs.

Mais les problèmes ne sont pas que d'ordre technique. En parallèle à l'évolution technique des SGDT et de la complexité accrue des produits, les organisations industrielles évoluent vers plus de versatilité ; la mobilité grandissante des collaborateurs d'une entreprise, l'utilisation systématique de prestataires de service, la délocalisation des activités de haute technologie, etc. On parle donc de *l'ingénierie à grande échelle* [Des02] qui souligne la dimension organisationnelle. Le développement de produits complexes nécessite une coordination forte entre les différents collaborateurs dispersés, or pour Bautzer [Bau05] la distribution géographique des partenaires entraîne une division dans les connaissances (*division of knowledge*). Bautzer [Bau05] cherche à trouver des solutions pour pouvoir rassembler collectivement les connaissances sur le cycle de vie du produit pour favoriser la collaboration.

¹ Ces projets avaient une finalité industrielle et n'ont pas fait l'objet de publications scientifiques pour des raisons de confidentialité.

2.3 L'apport de l'ingénierie des connaissances

Au niveau méthodologique, de nombreux travaux dans le domaine de l'ingénierie des connaissances ont abordé la problématique de capitalisation et la de la réutilisation du capital immatériel de l'entreprise ; CommonKADS [Kin97], MOKA [OKC98], MASK [Erm01], etc. Toutes ces approches offrent des modèles conceptuels pour capitaliser les connaissances industrielles ou pour réaliser des systèmes à base de connaissances contenant les bonnes pratiques, les retours d'expérience, des données relatives à une famille de produit, etc. Cependant, ces démarches ne permettent pas une intégration directe des modèles dans les documents gérés par les concepteurs, par exemple, sous la forme de balises XML.

3 Le Web Socio-Sémantique (W2S) en support des SGDT

Dans le cadre de la conception de produits industriels, notre préoccupation principale est de doter les connaissances métier d'une représentation formelle pour rechercher et réutiliser plus pertinemment ces connaissances. Le contenu des ressources SGDT représentant les connaissances à réutiliser doit être ainsi interprétables par les outils informatiques pour qu'ils soient capables de répondre aux requêtes des utilisateurs. Notre recours aux approches issues du champs du Web Sémantique est du à notre besoin pour la représentation flexible et commune des connaissances, le partage et l'échange d'information et l'automatisation de certains traitements, par exemple, pour l'indexation des ressources ou même pour faire des inférences.

3.1 La portée du Web Sémantique

Le Web Sémantique (WS) vise selon Tim Berners-Lee [BHL01], rapporté par [DCG04], à rendre le contenu sémantique des ressources du Web interprétables non seulement par l'homme mais aussi par des programmes, pour une meilleure coopération entre humains et machines. Pour ceci, le WS ambitionne à une formalisation des contenus sur un vaste espace d'échange de ressources entre utilisateurs et machines pour une meilleure exploitation de ces ressources. Tout comme pour I-Semantec (l'infrastructure proposée), l'infrastructure visée par le WS doit permettre selon Laublet [LRC02] de localiser, d'identifier et de transformer des ressources de manière robuste et saine. Elle doit s'appuyer sur un certain niveau de standardisation des langages de représentation, par exemple, ou sur des ontologies. Elle doit aussi garantir un bon niveau d'automatisation de l'interopérabilité et des transformations entre les différents formalismes et les différentes ontologies. Cependant, l'approche formelle du WS ne permet pas la représentation des connaissances métier heuristiques (dans lesquelles les relations entre les concepts ne sont pas toujours précisément définies) qui sont par ailleurs soumises à un rythme d'évolution rapide. C'est-à-dire que la représentation actuelle des données techniques est très rigide et que ces données sont très dynamiques, le WS pourra ainsi apporter une représentation souple mais le problème de la dynamique du méta modèle reste toujours posé.

3.2 Le W2S en entreprise

Notre étude se restreint à une organisation limitée en nombre de contributeurs, de ressources de données ou matérielles, ce qui simplifié plus la tâche. Dans ce cadre, les auteurs de [DCG04] proposent l'approche du *Web sémantique d'entreprise* (ou *organisationnel*) pour capitaliser les connaissances d'entreprise en reposant sur le concept de la *mémoire d'entreprise* (ou *organisationnelle*), qui est la matérialisation des connaissances cruciales d'une organisation en vue de faciliter leur accès, partage et réutilisation. Cependant, nous jugeons que cette approche est très générique pour mieux capitaliser les connaissances métier et moins structurée pour gérer la collaboration.

Par ailleurs, l'approche du Web Socio-Sémantique (W2S) définit dans [CZM04] est plus adaptée à notre démarche. Cette approche permet « aux acteurs, notamment au sein d'une communautés, de co-crée et enrichir en permanence les sémantiques qui les concernent ». Le W2S propose un modèle de connaissances (*HyperTopic*) pour la représentation des connaissances et la navigation et une méthode (la *documentarisation*) pour simplifier la transférabilité des connaissances et faciliter la communication. Ces aspects peuvent soutenir les activités de coopération dans lesquelles les interactions s'appuient également sur des informations ou des documents parta-

gés. Le W2S nous aide mieux à construire une représentation structurée et commune tant du domaine que du collectif.

Dans cette perspective, nous allons essayer de décharger le plus possible l'utilisateur du travail effectué pour la recherche d'information en exploitant les capacités accrues des machines avec des techniques d'inférence, par exemple. Donc, nous proposons l'infrastructure I-Semantec comme base à des services autour de la gestion des connaissances métier. Il s'agit d'une part de développer un outil logiciel offrant une interopérabilité sémantique, entre les différents SGDT, de l'autre, de fournir de nouveaux éléments de validation à l'approche du W2S. Sur la base de l'interopérabilité sémantique une représentation unifiée des données permettra de mettre en œuvre des inférences logiques et heuristiques pour faciliter la gestion des connaissances métier aux différentes phases du processus de conception.

Le candidat le plus prometteur pour une représentation unifiée est RDF (Resource Description Framework) [Las99]. Un langage réputé en WS, RDF se base sur des métadonnées pour décrire les ressources sans faire référence à un domaine d'application particulier tout en garantissant l'interopérabilité entre des machines et l'automatisation des traitements des ressources. Le mécanisme des métadonnées, qui sont facilement utilisable pour l'indexation et la classification des ressources, rend plus robuste la recherche d'information et l'application des inférences. Ceci rend RDF le plus conforme aux exigences de la plateforme I-Semantec (notre plateforme) que d'autres formalismes de représentation. Avec l'approche du Web Socio-Sémantique, nous allons mettre en place des modules enrichissant la représentation RDF avec les *Documents pour l'Action* [Zac04] et structurant les données avec le modèle *HyperTopic* [CZP04]. Ces approches nous permettront de renforcer l'indexation des données pour des meilleurs résultats de recherches et de mieux présenter les données au niveau de l'interface utilisateur.

Dans ce qui suit, nous présentons l'état d'avancement de nos travaux pour la réalisation de I-Semantec ; l'infrastructure industrielle basé sur les technologies du WS et du W2S pour la gestion des connaissances métier.

4 Extraction et réorganisation des connaissances du SGDT pour une ré-exploitation pertinente

4.1 Un outil d'instrumentation de la coopération

La plateforme I-Semantec est un Système d'Information collectif pour permettre la capitalisation et la recherche de données métier. I-Semantec propose de réaliser un entrepôt de données métier pour permettre :

- Le rafraîchissement et l'accumulation des connaissances. Les connaissances issues des informations contenues et gérées par les SGDT sont centralisées et réorganisées de manière dynamique de façon à pouvoir être réutilisées dans des contextes différents de ceux pour lesquelles elles ont été créées.
- De rechercher les connaissances de manière transparente. L'outil I-Semantec permettra de créer dynamiquement et collectivement de nouvelles ontologies qui architecturent la connaissance en prenant en compte à tout moment les besoins des concepteurs. Donc les possibilités de recherche ne seront plus limitées par la structure initiale imposée par les administrateurs du SGDT. Actuellement les seules possibilités de recherche disponibles pour un utilisateur d'un SGDT sont :
 - soit celles qui ont été prévues et modélisées par le concepteur du SGDT,
 - soit celles qui peuvent être indexées par un système de recherche par mots clefs.
- De partager les connaissances. Les données actuellement gérées par un SGDT sont organisées de façon à permettre leur utilisation dans le cadre d'un projet précis, il est rarement prévu une organisation permettant la réutilisation de ces données dans des projets futurs. Ces données sont aussi souvent organisées en fonction d'un besoin précis : besoin fonctionnel, d'étude, de fabrication, de qualité... ces différentes vues du produit sont rarement interopérables. I-Semantec permet de partager les connaissances en fournissant une infrastructure et des fonctions qui :
 - unifient le format de représentation ;
 - créent dans un même référentiel des organisations souples, dynamiques et collectives de ces différentes sources d'informations ;

- donnent un outil de recherche unique.
- L'utilisation de métadonnées et d'ontologies. Le projet I-Semantec se base sur les métadonnées et données déjà présentes dans les SGDT, le but est de les enrichir et les réorganiser par des ontologies souples, dynamiques, multipoint de vue et construites collaborativement.

4.1 Une approche sémantique pour la réutilisation des connaissances

Notre étude nous a permis de concevoir l'architecture de la plateforme I-Semantec (cf. Fig.1) adoptant une méthode de capitalisation de connaissances générique (identifier, recueillir, formaliser, valider, stocker, distribuer et maintenir). Les principes sur lesquelles nous avons bâti notre plateforme résident au niveau de sa flexibilité (représentation en RDF), extensibilité (architecture modulaire) et le degré d'interopérabilité avec d'autres outils. Dans cette perspective, I-Semantec intègre les aspects du W2S et doit répondre à des critères précis de robustesse, performance et être support à une forte collaboration. Pour ceci, nous avons sélectionné des logiciels libres, matures et robustes permettant la manipulation des ressources RDF ; Jena [Mbr01], Redland [Buc01], Protégé [NSD00], etc. Nous proposerons la plateforme I-Semantec comme logiciel libre.

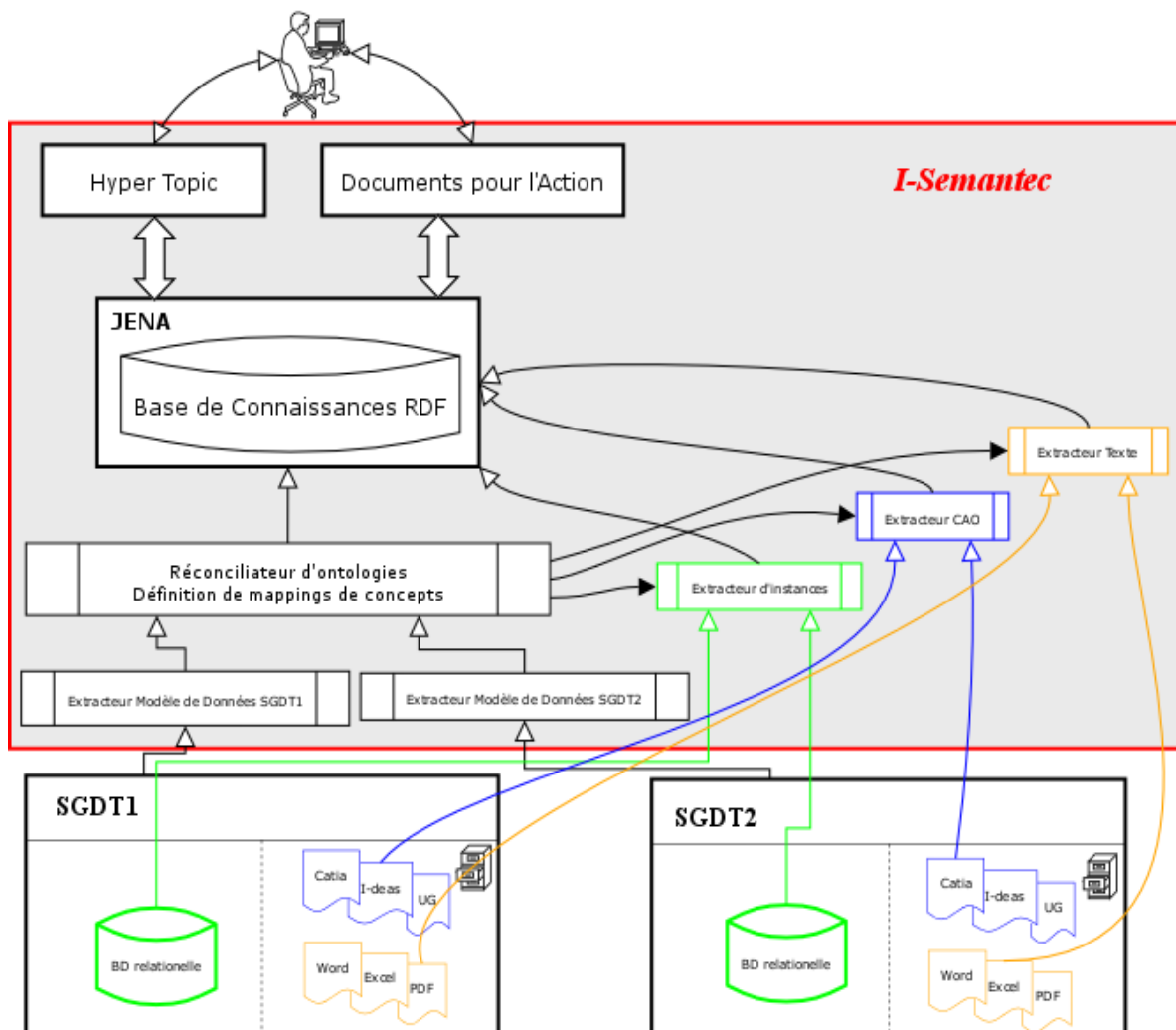


Fig. 1 – Architecture de la plateforme I-Semantec

4.1.2 Extraction et formalisation des connaissances

Le SGDT est composé d'une base de données relationnelle, d'un coffre (ou « vault », un espace sécurisé de stockage de fichiers) et d'une interface par laquelle l'utilisateur peut manipuler les données et les documents. A travers I-Semantec, les ressources issues d'un SGDT parcourent trois phases essentielles pour être à la disposition des utilisateurs : extraction, formalisation et réconciliation. Le principe de l'extraction est d'utiliser un API spécialisé pour accéder au contenu du SGDT : le modèle de données, les instances du modèle et les documents. L'extraction peut s'étendre à plusieurs SGDT d'une même entreprise ou de ces partenaires. Dans ces circonstances, la diversité des SGDT et de leur représentation des données implique la réalisation d'extracteurs spécifiques à chacun des SGDT. L'implémentation de ces extracteurs est conforme aux API fournies avec les SGDT, en d'autres termes, un extracteur par type de SGDT et par version de son API pour extraire la sémantique exacte (propriétés et relations) des concepts. En suite, les données extraites sont transformées en format RDF en utilisant Redland ou Jena. Cette phase de formalisation consiste en général à représenter les modèles de données par une ontologie définie en RDFS et les instances de ce modèle par un ensemble de ressource RDF dans *la base de connaissances*.

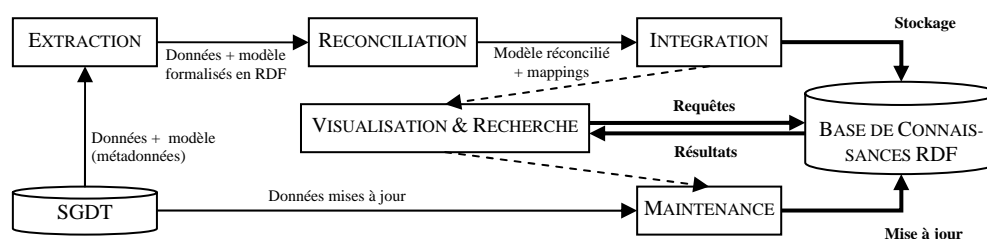


Fig. 2 – Processus de capitalisation en utilisant la plateforme I-Semantec

4.1.3 Réconciliation des modèles

Les modèles de données extraits des différents SGDT sont rarement similaires. Afin d'homogénéiser les concepts des modèles de données dans la base de connaissances, le module *réconciliateur d'ontologies* intervient pour offrir la possibilité de définir des « mappings » entre les concepts. Dans ce cas nous avons adapté PROMPT [Noy03] pour ce besoin. PROMPT est un outil basé sur Protégé et qui permet de gérer plusieurs ontologies, fusionner des ontologies distinctes pour créer une seule ontologie cohérente et extraire une partie d'une ontologie. Il faut noter que la phase de réconciliation à partir d'un SGDT et d'un modèle de donnée particulier n'intervient qu'une seule fois dans le processus d'enrichissement de la base RDF, le modèle de données d'un SGDT varie peu dans le temps. Naturellement si le modèle de données utilisé par un SGDT change, il faudra compléter la réconciliation. Ensuite les données issues du SGDT sont mises à jour régulièrement en respectant le « mapping » défini lors de la phase de réconciliation.

4.1.4 Visualisation et recherche d'informations

On utilise actuellement Protégé comme interface pour rechercher et visualiser le contenu extrait des SGDT. Le logiciel Protégé permet de naviguer facilement dans le modèle de données et ces instances. Les ressources disponibles dans la base RDF sont des liens vers les documents gérés par les SGDT, ces documents ne sont pas dupliqués mais restent gérés par le SGDT qui est propriétaire du document. L'accès à un document est donc indirect et doit donner lieu à une connexion vers le SGDT propriétaire.

Nous envisageons de réaliser une interface plus riche pour la recherche d'informations et effectués des inférences pour avoir des meilleurs résultats.

Actuellement, I-Semantec couvre les processus décrits à partir de SMARTEAM (un SGDT de Dassault Systèmes). En effet, nous envisageons prochainement de rendre exploitable le contenu des documents extraits à l'aide des outils de traitement automatique de langage (TAL). Pour les documents CAO il est important de mettre à disposition du concepteur des informations sur les cartouches de dessins, l'historique de conception, les annotations, etc. Pour les documents textuels, la recherche plein texte ne suffit pas, il faut prendre en compte le profil de l'utilisateur et les métadonnées décrivant le contexte de l'objet recherché.

5 Conclusion

Nous rappelons que le but de notre projet est de mettre à la disposition des ingénieurs-concepteurs un accès intelligent aux connaissances métier. Le support de ces connaissances est représenté par les données et les documents contenus dans le SGDT qui assure leur sécurité, cohérence, intégrité et disponibilité. Pour atteindre notre but, nous avons choisi de suivre l'approche du Web Socio-Sémantique, basée sur le Web Sémantique, pour représenter et enrichir les ressources du SGDT afin d'améliorer leur ré-exploitation.

Aujourd'hui I-Semantec, notre plateforme en cours de développement, implémente quelques principes du Web Socio-Sémantique pour la ré-exploitation des ressources en conception de produits. Les principales fonctionnalités sont : la recherche pertinente, l'échange de données avancé, simple et rapide et la modification de la structure du modèle de données. La réalisation d'I-Semantec nous met face à trois classes de problèmes liées respectivement aux domaines 1) de la modélisation des connaissances 2) du génie logiciel (implémentation et interface avec les API) et 3) de l'ergonomie cognitive et IHM.

Dans un futur proche nous envisageons d'expérimenter l'outil I-Semantec en « grandeur nature ». Nous mettrons à la disposition des concepteurs impliqués des connaissances extraites des anciens projets à travers une interface simple. Afin de pouvoir évaluer l'interface et l'utilité des fonctionnalités fournies par la plateforme et mesurer le niveau de collaboration réalisée à travers la plateforme, nous allons étudier le comportement et les pratiques des utilisateurs.

Références

- [Bau05] Bautzer R.L., 'Bringing together knowledge about the product life cycle to improve inter-firm collaboration capabilities'. in Int. J. of Automotive Tech. and Management, 5:184-198, 2005.
- [BHL01] Berners-Lee T., Hendler J., Lassila O., 'The Semantic Web', in Scientific American, 2001.
- [Buk01] Bucket, D., 'The Design and Implementation of the Redland, RDF Application Framework'. The Tenth International World Wide Web Conference, 2001.
- [CZM04] Cahier, J.P., Zacklad M., Monceaux A., 'Une application du Web socio sémantique à la définition d'un annuaire métier en ingénierie'. Ingénierie des Connaissances (IC'04), Lyon, 2004.
- [Cat01] Cattan, M., 'L'entreprise et ses données techniques : méthodes pour une documentation maîtrisée', Afnor, Paris La Défense 2001, chap. 11, p. 111.
- [CZM04] Cahier, J.-P., Zacklad M., Monceaux A., 'Une application du Web Socio-Sémantique à la définition d'un annuaire métier en ingénierie', in Actes de IC'2004.
- [Des02] Deshmukh, S.D., Patil M.D., 'PDM Requirements for Large Scale Engineering', Product Data Technology Europe Conference, 2002.
- [DCG04] Dieng-Kuntz, R., Corby O., Gandon F., Golebiowska J., 'Ontologies pour la construction d'un Web sémantique d'entreprise', in: Gestion dynamique des connaissances industrielles, B. Eynard, M. Lombard, N. Matta, J. Renaud, Hermès 2004, chap. 1, p. 27-43.
- [Erm01] Ermine J.L., Matta N., 'Gestion de connaissances et conception mécanique', Journée AIP - PRIMECA de Grenoble, 2001
- [Las99] Lassila, O., Swick R.R., 'Resource Description Framework Model and Syntax Specification'. Recommendation W3C, 1999.
- [LRC02] Laublet, P., Reynaud C., Charlet J., 'Sur quelques aspects du Web Sémantique'. Les assises du GdR-I3, Nancy, 2002.
- [Kin97] Kingston, J., 'Designing Knowledge Based Systems: The CommonKADS Design Model', Proceedings of Expert Systems '97, Cambridge: Churchill College, 1997.
- [Mbr01] McBride, B., 'Jena: Implementating the RDF Model and Syntax Specification'. Second International Workshop on the Semantic Web, 2001.

- [NSD00] Noy, N. F., Sintek M., Decker S., Crubezy M., Ferguson R.W., Musen M.A., 'Creating Semantic Web Contents with Protege-2000', IEEE Intelligent Systems, 2000.
- [Noy03] Noy N. F., Musen M. A., 'The PROMPT suite: Interactive tools for ontology merging and mapping'. International Journal of Human-Computer Studies, 59(6):983–1024, 2003.
- [OKC98] Oldham, K., Kneebone S., Callot M., Murton A., Brimble R., 'MOKA - A Methodology and tools Oriented to Knowledge-based engineering Applications', Conference on Integration in Manufacturing, Göteborg, Suède, 1998.
- [Zac04] Zacklad, M., 'Processus de documentarisation dans les Documents pour l'Action : statut des annotations et technologies de la coopération associées', in Colloque Le Numérique : Impact sur le cycle de vie du document pour une analyse interdisciplinaire, Montréal, 2004.