

## Échanges de connaissance structurée médiatisés par ordinateur

### Computer-mediated exchange of structured knowledge

[EXMO]

Jérôme Euzenat

*INRIA Rhône-Alpes*

Cette proposition de recherche a pour but d'indiquer et de motiver nos directions de recherches pour les années à venir.

#### 1. PROBLÉMATIQUE ET MOTIVATION

Exmo promet l'utilisation de la connaissance formalisée dans la communication entre personnes médiatisée par ordinateurs. Nous posons que cela vaut la peine d'exprimer la connaissance de manière formelle sur ordinateur, non pour les besoins exclusifs de l'ordinateur, mais aussi pour communiquer à d'autres cette connaissance. Ce faisant, l'ordinateur peut apporter une valeur ajoutée à son rôle de mémoire ou de médium en accomplissant des fonctions plus avancées (détection de similarité, de cohérence, généralisation, filtrage et autres transformations).

Ceci est mis en œuvre dans le cadre de mémoires techniques ou de bases de connaissance scientifiques et poursuivra son développement dans tous les aspects des systèmes d'information, en particulier dans le contexte d'une généralisation du travail collaboratif assisté par ordinateur (CSCW).

La figure 1 présente une organisation dans laquelle coopèrent des agents (humains et logiciels). Les agents sont à la fois producteurs et consommateurs de connaissance représentée informatiquement (l'organisation elle-même peut-être considérée comme un agent). Pour être plus concret, on peut envisager que les individus (a) et (b) collaborent à la conception (C) d'un document sur un produit et une source extérieure (un sous-traitant) y apporte les informations concernant les caractéristiques sous-traitées. Cette information, suivant les canaux commerciaux peut être traduite (T) en diverses langues, complétée par l'information des prix de ventes pratiqués (P) et formatée (A) pour diverses utilisations, y compris des utilisations informatiques extérieures à l'entreprise. Ce type de schéma n'a rien de futuriste ; il

s'incarne déjà dans la gestion de connaissance, la gestion de ressources intégrée, la gestion électronique de documents (ou la maquette numérique) et l'ingénierie concurrente.

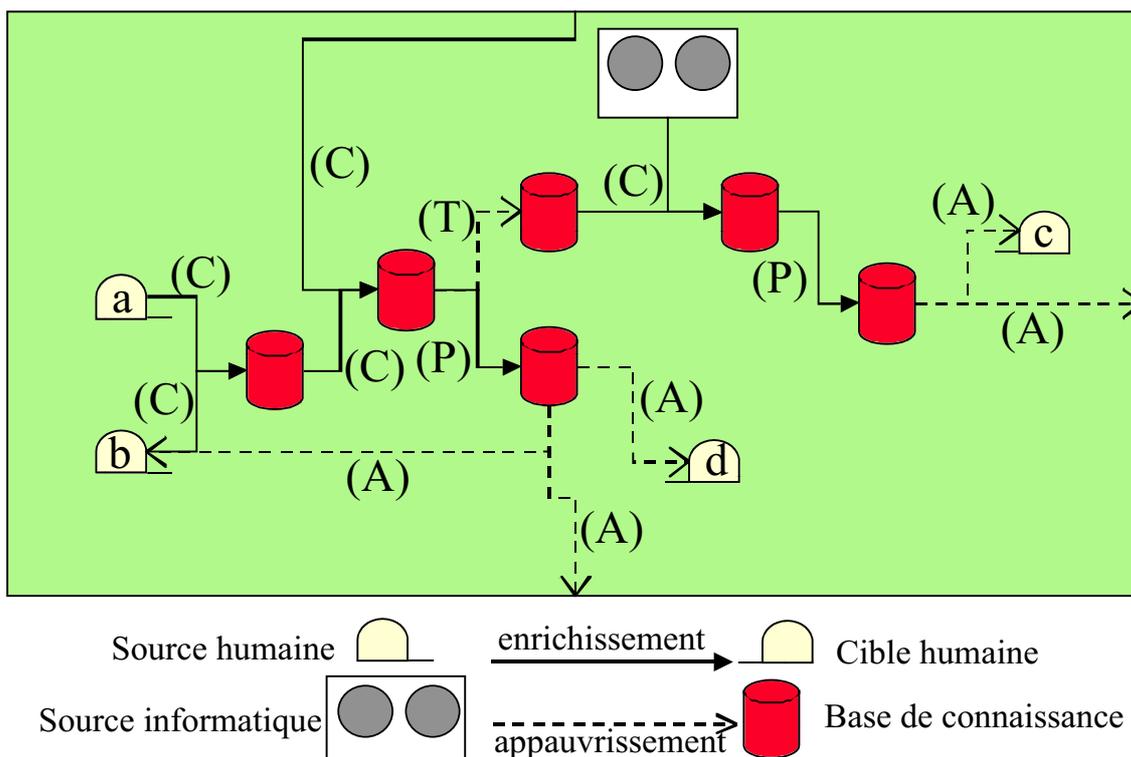


Figure 1 : Transformation de la connaissance structurée dans une organisation.

Nous voulons mettre en évidence les règles générales qui sont à l'œuvre dans une telle organisation. Pour cela, les différents stades du schéma sont considérés comme des bases de connaissance et les transitions entre deux bases comme des transformations (qui peuvent être des combinaisons de plusieurs bases). La prise en compte de la connaissance formalisée, permet d'aller plus loin que la transformation de documents en tenant compte non seulement de la structure mais aussi de la signification des objets transformés.

Il est vital d'étudier les propriétés conservées par ces transformations et surtout la composition de celles-ci. En effet, il peut être utile de définir une transformation (appauvrissement) qui fournisse aux clients une documentation sur le produit tout en garantissant qu'un certain type d'information ne s'y trouvera pas. Dans le cadre d'un travail collaboratif, il est judicieux de connaître les propriétés de ce type de filtres afin de savoir où le placer dans la chaîne (il est inutile, si après lui, l'information ajoutée permet de retrouver l'information masquée). Dans ce même cadre, si l'on considère la protection de parties du documents, il faut s'assurer que les transformations qui y sont appliquées ne peuvent modifier les parties protégées.

À l'inverse, lorsque l'élaboration d'une représentation est un processus collaboratif et continu (sur le modèle de l'ingénierie concurrente), il est nécessaire que les traitements qui sont appliqués ne remettent pas, inopportunément, en cause les développements précédents. Il est alors utile de pouvoir garantir la préservation de la structure ou du contenu de la base.

La problématique d'Exmo est le développement d'outils (théoriques et informatiques) pour la représentation de connaissance permettant l'organisation, la manipulation, la communication, la présentation et la combinaison d'éléments de connaissance structurée de telle sorte que le résultat soit adapté à et intelligible par les interlocuteurs. Ceci requiert de développer une façon abstraite de considérer les représentations et les systèmes de transformations qui leur seront appliqués.

## **2. PROGRAMME DE RECHERCHE**

Notre but est de contribuer sur deux axes dénommés ici transformation et communication. Pour cela, il faut s'appuyer sur un formalisme de représentation.

### **Représentation**

Concernant les formalismes de représentation de connaissance, nous voulons nous libérer du langage précis utilisé pour constituer les bases de connaissance. Cependant, afin de borner et de concrétiser le travail de recherche, on se concentrera sur deux types de langages qui sont les deux extrémités d'un spectre d'applications:

- Les langages de représentation de connaissance par objets et spécialement celui que nous avons conçu et implémenté : Troeps. C'est un langage très précis doté d'une sémantique dénotationnelle.
- XML, plus polyvalent (c'est un métalangage), peut difficilement être considéré comme un langage de représentation de connaissance car sans sémantique.

On prolongera notre réflexion sur l'échange de connaissance formalisée et, en particulier, sa communication à travers le Web et sa contribution au Web. Sur ce dernier point, une action coopérative en cours a pour but d'étudier comparativement les apports de différents langages de représentation de connaissance (graphes conceptuels, représentations par objets, logiques de description) dans la représentation du contenu de documents pour l'indexation.

### **Transformation**

Une transformation est un moyen algorithmique d'engendrer une représentation à partir d'une autre (pas forcément dans le même langage). Comment transformer une représentation lors de la communication, que ce soit pour l'enrichir ou l'appauvrir ? Le but est de bâtir une « théorie générale de ces transformations » fondée non pas sur les représentations ni sur les transformations en tant que telles, mais sur les propriétés qu'elles doivent satisfaire (préservation du contenu ou de la structure, identifiabilité de la source — traçabilité —, confidentialité). L'intérêt est, lors de la mise au point d'un système de transformations, d'assurer en sortie les propriétés souhaitées. Ce devrait être un outil précieux pour les futurs architectes des systèmes d'information.

Dans le schéma de la figure 1, il est possible de distinguer deux types de transformations d'un état à un autre suivant que la transformation conserve l'information présente dans la base initiale (enrichissement) ou qu'au contraire elle ne la conserve pas (appauvrissement). Cette problématique recouvre deux aspects déjà explorés et complémentaires :

- Combinaison et enrichissement de représentations ;
- Adaptation et appauvrissement de représentations ;

Nous envisageons d'abord comment une représentation peut être manipulée (en considérant que l'objet à représenter reste le même). Ceci requiert de rendre compte de la manière d'enrichir la représentation ou d'assembler plusieurs représentations (qui peuvent être exprimées dans différents langages, par différents auteurs et/ou peuvent considérer différents aspects de l'objet à représenter).

Une application évidente de ce travail concerne les outils de représentation de connaissance et les méthodologies pour le travail collaboratif. On peut y considérer des applications de type :

- Construction collaborative de bases de connaissance consensuelles,
- Édition collaborative guidée par le contenu (un sujet de départ concerne de l'édition concourante de documents XML),
- Assemblage automatique de documents à partir de sources hétérogènes,
- Comparaison des formalismes de représentation de connaissance pour représenter le contenu de documents (une action coopérative est en cours sur ce sujet).

Nous avons déjà étudié les propriétés de consensus lors de la construction collaborative ; nous envisageons de considérer la cohérence du processus d'ingénierie concourante.

De manière duale, on considère la génération de représentations plus abstraites à partir d'une représentation initiale. De telles représentations seront principalement engendrées automatiquement. Les représentations obtenues auront à satisfaire des propriétés particulières : classement, privilèges, pertinence, niveau de granularité, traçabilité de l'information. L'illustration évidente de cet aspect concerne la construction de serveurs de connaissance pour des cibles particulières.

Ici, il est intéressant d'étudier la préservation des propriétés en particulier lors de la composition de transformations de différents types (par exemple, comment préserver la traçabilité face aux exigences de classement de l'information ?).

Des travaux systématiques sont nécessaires pour prendre en compte ces transformations. Afin d'accomplir les tâches décrites ci-dessus chaque représentation peut être considérée comme un objet et chaque transformation comme un morphisme entre deux représentations préservant des propriétés bien définies. Ce type de structure se définit naturellement pour les représentations par objets ou les documents XML.

## **Communication**

Comment s'assurer de la bonne compréhension du contenu entre deux utilisateurs? La sémantique formelle utilisée dans les systèmes de représentation de connaissance est valable pour un utilisateur en présence d'un système car il interprète de manière cohérente les termes (identificateurs) qu'il a utilisés. Mais lorsque plusieurs utilisateurs communiquent, cette bonne compréhension est difficile à obtenir. Les travaux, menés ailleurs, sur la construction d'« ontologies » consensuelles cherchent à atteindre ce but.

Mais d'autres problèmes interviennent. Par exemple, un utilisateur peut exprimer sa connaissance à l'aide de hiérarchies de classes et de clauses. Puis communiquer celle-ci en la traduisant vers un langage destiné à assurer l'interopérabilité. Mais si ce langage exprime la totalité de la connaissance sous forme de clauses (tout en préservant la sémantique des assertions), l'émetteur risque de ne plus reconnaître le résultat pourtant sémantiquement équivalent. Ainsi, lorsqu'il y a traduction entre plusieurs systèmes de représentations, la bonne compréhension ne peut être garantie. Il faut ajouter un traitement sémiologique (ou pragmatique dirait-on en linguistique) pour faciliter la reconstruction du sens. Ce traitement complète le traitement purement sémantique utilisé habituellement.

Pour contribuer à la résolution de ce dernier problème, on s'intéresse à la possibilité de communiquer non plus seulement la syntaxe des langages utilisés dans les représentations mais aussi leur sémantique. Ainsi, la forme sera conservée autant que possible en utilisant la forme initiale (la compréhension humaine est tributaire de la forme) et l'interopérabilité est assurée grâce à la possibilité d'accéder à la sémantique (le traitement informatique ne dépend que de la sémantique).

### **3. OBJECTIFS À DEUX ANS**

Dans le court terme, l'objectif d'Exmo consistera à développer ces idées sur quelques propriétés, quelques transformations et un (ou deux) langage(s) particulier(s). Puis on examinera l'applicabilité des résultats sur quelques autres langages afin de choisir un formalisme adéquat de description des phénomènes. Dans le même temps on se penchera sur un langage d'expression de la sémantique qui soit utilisable dans le contexte du Web.

L'une des pistes est proposée par Joseph Goguen, dans la continuité de ses travaux sur les institutions, qui développe une approche baptisée sémiologie algébrique visant à considérer les représentations en liaison avec leur usage. On considère également d'autres sources d'inspiration telles que les flux d'informations de Jon Barwise et Jerry Seligman.

Par ailleurs, on cherchera à élargir la base de collaboration.