



INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

## *Action EXMO*

*Échanges de connaissance structurée médiatisés par ordinateur*

*Rhône-Alpes*

THÈME 3A

*R*apport  
*d'Activité*

2002



## Table des matières

<b>1</b>	<b>Composition de l'équipe</b>	<b>2</b>
<b>2</b>	<b>Présentation et objectifs généraux</b>	<b>2</b>
<b>3</b>	<b>Fondements scientifiques</b>	<b>3</b>
3.1	Sémantique de la representation de connaissance . . . . .	3
3.2	Transformations et propriétés . . . . .	4
3.3	Interprétation des flux d'information . . . . .	5
<b>4</b>	<b>Domaines d'applications</b>	<b>5</b>
4.1	Panorama . . . . .	5
4.2	Ingénierie des systèmes de transformations . . . . .	6
4.3	Technologies pour un « web sémantique » . . . . .	7
<b>5</b>	<b>Logiciels</b>	<b>8</b>
5.1	Panorama . . . . .	8
5.2	TRANSMORPHER : spécification et exécution de flux de transformations XML . . . . .	8
<b>6</b>	<b>Résultats nouveaux</b>	<b>10</b>
6.1	Interopérabilité sémantique . . . . .	10
6.1.1	Hierarchie de propriétés sémantiques . . . . .	10
6.1.2	Familles de langages . . . . .	11
6.2	Treillis et information . . . . .	11
<b>7</b>	<b>Contrats industriels (nationaux, européens et internationaux)</b>	<b>12</b>
7.1	Fluxmedia . . . . .	12
<b>8</b>	<b>Actions régionales, nationales et internationales</b>	<b>12</b>
8.1	Actions nationales . . . . .	12
8.1.1	Action spécifique Web sémantique (CNRS) . . . . .	12
8.2	Actions européennes . . . . .	12
8.2.1	Réseau thématique OntoWeb : ontology-based information exchange for knowledge management and electronic commerce . . . . .	12
8.3	Actions internationales . . . . .	12
8.3.1	Groupe de travail WebOnt (W3C) . . . . .	12
<b>9</b>	<b>Diffusion de résultats</b>	<b>13</b>
9.1	Animation de la communauté scientifique . . . . .	13
9.2	Enseignement . . . . .	13
9.3	Participation à des colloques, séminaires, invitations . . . . .	13
<b>10</b>	<b>Bibliographie</b>	<b>14</b>

## 1 Composition de l'équipe

### Responsable scientifique

Jérôme Euzenat [CR Inria]

### Assistante de projet

Marie-Anne Dauphin [jusqu'au 1er septembre 2002]

Marion Ponsot [depuis le 1er septembre 2002]

### Personnel Inria

Jean-François Baget [CR, depuis le 1er septembre 2002]

### Chercheurs doctorants

Olivier Brunet [allocataire Ministère]

Raphaël Troncy [CIFRE (Institut National de l'Audiovisuel)]

### Ingénieur

Fabien Triolet [Ingénieur associé, depuis le 1er juillet 2002]

## 2 Présentation et objectifs généraux

**Mots clés :** représentation des connaissances, sémantique des représentations, transformations, ontologies, treillis, préservation de propriétés, web sémantique, représentation du contenu, sémiologie, Transmorpher, WWW, DLML, XML, XSLT.

EXMO part du principe que, dans les futurs systèmes d'information, la connaissance formalisée sera échangée de façon routinière. Dans le processus de communication, l'ordinateur peut introduire une plus-value à son rôle de médium et de mémoire en accomplissant des tâches comme le formatage, le filtrage, la catégorisation, le test de consistance ou la généralisation de la connaissance. Ces manipulations peuvent être vues comme des transformations. Cette approche commence à se développer avec la généralisation de l'utilisation de langages d'échange standardisés (XML) lors de la communication en réseau. En contrepartie, les utilisateurs devront exiger une plus grande assurance sur les transformations opérées. EXMO a pour but le développement d'outils théoriques et logiciels pour aider à l'échange d'éléments de connaissance formalisée.

Notre angle d'attaque principal consiste à examiner les propriétés que doivent satisfaire les transformations appliquées aux représentations formalisées. Parmi ces propriétés, on trouve la préservation du contenu ou de la structure, la traçabilité des sources ou, au contraire, la confidentialité. On cherche à élaborer une « théorie générale des transformations » fondée sur les propriétés satisfaites par les transformations plutôt que sur les transformations elles-mêmes.

### Axes de recherche

D'une part, nous nous attachons à montrer que cette problématique s'applique à un ensemble varié de contextes dépendant des langages, propriétés et transformations impliqués. Ainsi, nous étudions la préservation de l'information modélisée par des treillis ou la préservation de scénario de documents multimédia durant l'adaptation.

D'autre part, nous nous intéressons plus particulièrement aux propriétés sémantiques lors de la transformation de la connaissance d'un langage à un autre. La question essentielle est : les conséquences d'une représentation initiale sont-elles les transformées de celles de la représentation transformée ? Pour cela nous établissons les relations entre plusieurs types de propriétés sémantiques, connues ou non, en rapport avec celles-ci. Nous avons aussi étudié des cadres plus restreints (famille de langages, patrons) permettant d'utiliser ces relations et d'établir plus facilement les propriétés des transformations.

Enfin, sur un plus long terme, nous avons l'ambition d'explorer des propriétés appelées « sémiologiques », c'est-à-dire qui concernent l'interprétation par un utilisateur humain de la représentation communiquée, et leur préservation.

### Applications

Les applications anticipées sont l'ingénierie des systèmes de transformation (où le système d'information est compris comme un flux de transformations) et l'infrastructure du « web sémantique » (serveurs de connaissance, représentation du contenu, transformation de représentations).

## 3 Fondements scientifiques

### 3.1 Sémantique de la représentation de connaissance

Après les années de développement empirique, une rationalisation du domaine de la représentation de connaissance a eu lieu, principalement liée à l'étude de la sémantique des formalismes utilisés. Depuis une quinzaine d'années, la sémantique des langages de représentation de connaissance (logiques de descriptions, graphes conceptuels et représentations de connaissance par objets) a été étudiée<sup>[Neb90]</sup>. Elle est définie en général à l'aide de la théorie des modèles, issue elle-même de la logique.

On considérera ici un langage  $L$  comme un ensemble d'expressions syntaxiques, souvent définies inductivement en appliquant des constructeurs sur d'autres expressions. Une représentation ( $r$ ) est un ensemble d'expressions. On définit inductivement sur la structure du langage une fonction d'interprétation ( $I$ ) vers un ensemble nommé domaine d'interprétation ( $D$ ). Cette fonction reflète la construction du sens d'une expression en fonction de ses composants. Les assertions de ce langage sont satisfaites par une interprétation si elles remplissent une condition (en général être interprétées dans un sous-ensemble particulier du domaine). Dans ce cadre un modèle est une interprétation dans laquelle tous les axiomes sont satisfaits. Une expression ( $\delta$ ) est alors une conséquence d'un ensemble d'axiomes ( $r$ ) si elle est satisfaite par tous leurs modèles (ce qui est noté  $r \models \delta$ ).

Pour un ordinateur, le but consiste à déterminer si une expression particulière (par exemple une requête) est conséquence des axiomes (par exemple, la base de connaissance considérée). Pour cela on développe des systèmes exécutables, appelés démonstrateurs, qui peuvent être fondés sur l'interprétation de règles d'inférence ou sur des programmes plus classiques. Ils

---

[Neb90] B. NEBEL, *Reasoning and revision in hybrid representation systems, Lecture Notes in Artificial Intelligence 422*, Springer Verlag, Berlin (DE), 1990.

permettent de déduire des théorèmes (ce qui est noté  $r \vdash \delta$ ). Ces démonstrateurs vérifient la propriété de correction s'ils ne répondent positivement qu'en cas de conséquence et la propriété de complétude s'ils répondent positivement pour toutes les conséquences. Cependant, suivant le langage et sa sémantique, la décidabilité – c'est-à-dire la possibilité de créer de tels démonstrateurs – n'est pas garantie. Même dans le cas de langages décidables, leur complexité algorithmique peut être prohibitive.

Pour cela on est amené à réaliser un compromis entre l'expressivité d'un langage et la complexité ou la complétude des démonstrateurs associés. Ce choix s'est traduit par la définition de langages à l'expressivité limitée – comme les graphes conceptuels ou les représentations par objets – ou de familles de langages modulaires dont les démonstrateurs peuvent être étendus – comme les logiques de descriptions.

L'action EXMO s'appuie d'abord sur des langages dont la sémantique est ainsi définie afin d'établir les propriétés des manipulations informatiques appliquées aux représentations.

## 3.2 Transformations et propriétés

Dans notre approche, les traitements appliqués aux représentations sont qualifiés de transformations. Une transformation est un moyen algorithmique d'engendrer une représentation à partir d'une autre (pas forcément dans le même langage). On se concentrera sur des transformations obtenues par assemblage de transformations atomiques, considérées comme des boîtes noires, dont on ne connaît que entrées et sorties ainsi que des propriétés liant les entrées aux sorties.

On définit un système de transformations par l'ensemble de ses transformations élémentaires et l'ensemble de ses assembleurs de transformations. Un flux de transformations est un ensemble d'instances de transformations élémentaires assemblées et reliées par des canaux. Un flux de transformations est lui-même une transformation. Plus concrètement, nos travaux concernent des transformations syntaxiques de documents XML (« Extensible Markup Language ») encodant des langages de représentation de connaissance. Nous cherchons à exploiter le langage de transformations XSLT (« XML Stylesheet Language Transformations »)<sup>[Cla99]</sup> recommandé par le W3C, pour lequel nous avons proposé un langage d'assemblage de transformations.

Le but d'EXMO est d'étudier les propriétés des transformations et leur combinaison. Une propriété est simplement un prédicat booléen portant sur une transformation (par exemple, « préserver l'information » est un tel prédicat – il est vrai ou faux d'une transformation – et se vérifie s'il existe un procédé algorithmique permettant, pour toute représentation  $r$ , de la retrouver à partir de sa transformée  $\tau(r)$ ). La question intéressante est alors de vérifier, pour un flux de transformations donné qu'il vérifie bien une propriété. Nous identifions deux types de propriétés : des propriétés de préservation et des propriétés à motifs qui consistent à tester si la présence d'un motif dans la représentation de départ entraîne la présence de ce motif dans la représentation d'arrivée.

Les propriétés de préservation nous intéressent plus particulièrement. Elles peuvent permettre de garantir la préservation d'un ordre entre la représentation d'entrée ( $r$ ) et la représentation

---

[Cla99] J. CLARK, (ÉD.), « XSL Transformations (XSLT) », *Recommendation*, W3C, Cambridge (MA US), 1999, <http://www.w3.org/TR/xslt>.

de sortie ( $\tau(r)$ ) comme son anti-préservation. Pour cela on peut identifier des propriétés :

- syntaxiques : fondées sur l’organisation des éléments syntaxiques (ordre, encodage...);
- sémantiques : fondées sur les notions de modèles ou de conséquence ;
- sémiologiques : fondées sur l’interprétation en tant que signe des objets manipulés.

Dans un premier temps, nous travaillons plus spécifiquement les propriétés sémantiques.

### 3.3 Interprétation des flux d’information

L’analyse des flux d’information est centrée autour d’une formalisation à base de demi-treillis (un ensemble  $E$  muni d’un ordre partiel  $\leq$  et dans lequel tout couple d’éléments a un plus petit majorant pour l’ordre). Les treillis permettent d’avoir une notion d’imprécision dans la description de l’information. L’information disponible en un point sur un objet est un élément de ce treillis. L’ordre permet de préciser qu’un élément du treillis contient plus d’information qu’un autre. On peut ainsi définir l’absence totale d’information (l’élément maximal de l’ordre) et un élément plus grand qu’un autre est une approximation de celui-ci (car il contient moins d’informations, mais ne le contredit pas). Ainsi, considérant deux systèmes échangeant de l’information, on peut écrire de façon rigoureuse qu’un système connaît précisément telle information d’un autre système, ou qu’il ne connaît cette information que partiellement, ou encore pas du tout.

Si l’on envisage les langages que nous utilisons comme des treillis ( $\langle L, \leq \rangle$ ), une transformation  $\tau : L \rightarrow L'$  d’un tel treillis dans un autre permet de modéliser la modification subie par l’information lors de sa circulation (la transformation) d’un espace de représentation à un autre. Une transformation perdra de l’information si sa transformation adjointe (définie par le plus petit majorant des éléments qui ont pour image un élément de  $L'$ ) peut retourner un élément plus grand que l’élément de départ (elle n’en retournera jamais de plus petit). Cette approche est à la base de travaux concernant l’échange et l’organisation de l’information comme la théorie des flux d’information (*Information Flow Theory* <sup>[BS98]</sup>) et l’analyse de concepts <sup>[GW99]</sup>.

Nous étudions pour notre part, la façon dont se composent et s’assemblent les transformations et les systèmes logiques particuliers que cela induit.

## 4 Domaines d’applications

### 4.1 Panorama

Les efforts de l’équipe EXMO sont dirigés vers deux types d’applications particuliers : l’ingénierie des systèmes de transformations (§

---

[BS98] J. BARWISE, J. SELIGMAN, *Information Flow, the Logic of Distributed Systems*, Cambridge university press, Cambridge (UK), 1998.

[GW99] B. GANTER, R. WILLE, *Formal Concept Analysis*, Springer Verlag, Berlin (DE), 1999.

## 4.2 Ingénierie des systèmes de transformations

**Résumé :** *La diffusion et la transformation de documents structurés induisent une ingénierie des systèmes de transformations qu'il faudra pourvoir en logiciels et en méthodes formelles qu'EXMO cherche à construire.*

De plus en plus, l'informatisation et la mise en réseau des organisations les conduisent à échanger de l'information sous forme électronique. Par exemple, le commerce électronique engendre un échange permanent de documents structurés, les dates de réunions s'échangent dans toutes sortes de formats, les données de conception d'un artefact sont transmises aux sous-traitants et aux unités de production.

Cependant, l'intégration des données est loin d'être parfaite. Elle est freinée par de nombreux facteurs comme l'hétérogénéité des outils de manipulation de l'information, les formats de données inaccessibles ou la confidentialité de certains éléments d'information. Ainsi, les documents traités par un tableur ne sont pas disponibles dans les agendas de la force de vente, l'unité chargée d'imprimer le matériel promotionnel ne peut lire les données de CAO et les statisticiens ne peuvent avoir accès à certains tableaux car ils contiennent des éléments confidentiels.

Certaines briques technologiques ont été récemment développées afin de résoudre ces problèmes. C'est notamment le cas du langage XML destiné à servir de base à l'encodage de nombreux langages. Le langage XSLT permet pour sa part d'exprimer des transformations de documents XML. Ces éléments sont de natures à résoudre beaucoup des problèmes cités plus haut. Ainsi, XML permettra d'appréhender plus facilement la syntaxe des données et des transformations utilisant XSLT permettront d'adapter ces données à un contexte cible. Les fichiers de tableur et de CAO exportés en XML pourront être transformés dans un format admissible par les agendas ou les outils de mise en pages, les données confidentielles pourront être expurgées des fichiers clients avant d'être traitées par l'unité de statistique.

Cette présentation constitue une instance du cadre plus général auquel EXMO entend contribuer : elle décrit bien la transformation de données structurées lors de sa communication entre personnes ou entre applications.

Cependant, cette approche descriptive des transformations n'est pas satisfaisante. Elle ne permet pas de définir les propriétés attendues de celles-ci et par conséquent de vérifier que ces propriétés sont satisfaites. Cette lacune devra être comblée dans les prochaines années si l'on veut construire une véritable ingénierie. Nous identifions et discutons ci-dessous trois problèmes particuliers:

- l'absence de prise en compte de manière globale de ces transformations;
- la nécessaire prise en compte des propriétés des transformations et, en particulier, de leurs propriétés sémantiques;
- la conception des systèmes de transformations à partir de ressources ouvertes.

Il est peu raisonnable d'imaginer développer une transformation en XSLT à chaque fois que l'on désire communiquer de l'information entre deux nouvelles cibles. D'autre part, il est peu imaginable de considérer chaque transformation indépendamment des autres. Il semble inévitable que, dans l'avenir, il faudra gérer des systèmes de transformations complexes. Les transformations doivent donc être appréhendées sous forme de systèmes de transformations intégrant chaque transformation individuelle dans un tout cohérent (voir §

On a besoin de pouvoir assembler facilement des transformations : par exemple, « donnez-moi la liste des publications d'EXMO depuis 2000, dont on a supprimé celles qui ne sont pas encore diffusables, classées par catégories au format HTML ». Pour cela, à partir d'une source de données bibliographique, on aimerait pouvoir assembler des transformations élémentaires (sélection, filtrage, classement et formatage) tout en s'assurant que l'assemblage satisfait certaines propriétés (par exemple, puisque les informations ont été triées, on aimerait que le formatage ne change pas l'ordre). Un système de transformations organise explicitement l'ensemble de ces transformations de flux de données dans un schéma unique permettant de réaliser les tâches de changement de formats à l'aide de transformations standard, souvent réutilisées, les tâches de filtrage à l'aide de transformations simples dont on s'assurera qu'elles ne laissent pas passer ce qui doit être filtré. L'ensemble de ces transformations s'exécute de manière coordonnée afin de s'assurer que l'information fournie est toujours à jour.

Les systèmes de transformations s'appliquent donc à l'échelle du fonctionnement d'un système d'information et non plus à l'échelle de chaque individu. Il y a, à notre avis, une ingénierie à développer autour de la conception de tels systèmes. Nous appelons ingénierie des systèmes de transformations l'activité qui s'attache à développer des transformations, à les assembler, à définir et à vérifier les propriétés qui doivent être satisfaites par ces transformations.

L'ingénierie de ces systèmes de transformations nécessitera des outils, des méthodologies et des méthodes formelles. En effet, il sera nécessaire de vérifier qu'un système particulier, et non plus une transformation individuelle, ne laisse pas échapper plus d'information qu'autorisé ou que le flux dans sa globalité termine. Pour cela, il faudra disposer d'une description analysable de ce flux de transformations et de caractérisations formelles des propriétés escomptées. EXMO est concerné par les outils et les méthodes formelles et cherche à les associer en proposant des solutions pour la création de véritables ateliers de conception de systèmes de transformations (voir §

### 4.3 Technologies pour un « web sémantique »

**Résumé :** *Dans un univers où les documents du web seront annotés formellement, il sera nécessaire d'importer et de manipuler ces annotations en accord avec leur sémantique et en accord avec leurs utilisations. Nous travaillons aux technologies permettant cela.*

L'utilisation des technologies d'Internet est l'opportunité pour les entreprises d'accéder à et de partager la connaissance bien souvent difficilement accessible sous forme documentaire. Cependant, les limites de cette approche apparaissent rapidement : l'organisation des sites se révèle une tâche coûteuse et la recherche en texte intégral peu efficace. La recherche et l'interrogation d'un site en s'appuyant sur le contenu «sémantique» des documents sont une nécessité et les formalismes de représentation de connaissance sont de bons candidats pour représenter ce contenu. La représentation du contenu permettra de le manipuler pour faire de la recherche par analogie, par spécialisation, par similitude, etc.

L'idée d'un « web sémantique » [BLHL01] consiste à adjoindre au web actuel (informel) des annotations (rédigées dans des langages exploitables par une machine) liées entre elles comme au web informel [3, 15]. Exploiter ce web sémantique demandera donc d'appréhender ces représentations formelles diversifiées. La problématique de l'action EXMO se trouve donc au cœur de la mise en œuvre du web sémantique.

L'ensemble des travaux d'EXMO ont généralement pour but de concourir à une meilleure appréhension des contenus [10, 8]. Cela concerne, bien entendu, les travaux sur l'intelligibilité des connaissances communiquées, mais aussi les travaux sur la transformation de représentations formelles. Les travaux sur le web sémantique se basent actuellement sur la notion d'ontologie (que l'on peut rapidement décrire comme une axiomatisation d'un domaine). Même s'il existe un jour un langage commun de représentation, il faudra importer les ontologies décrites dans d'autres langages d'une façon qui respecte la sémantique des langages manipulés [6]. Apporter des solutions à ce problème fait partie des ambitions d'EXMO (voir §

Nous avons développé ci-dessus l'aspect local de l'ingénierie des systèmes de transformations. Nous pensons qu'il est nécessaire de la considérer dans un contexte global. C'est pourquoi nous cherchons à développer une infrastructure permettant de prolonger les solutions évoquées plus haut en tirant parti du web.

Au lieu d'utiliser des transformations disponibles localement dans son unité, l'ingénieur des systèmes de transformations aura la possibilité d'intégrer des transformations disponibles sur le réseau mondial. Il est vraisemblable que toutes les transformations de changement de format (par exemple, du format de calendrier vCal vers HTML) ou d'export pour des dispositifs particuliers (WAP) devraient y être disponibles. Ainsi, il est vraisemblable que les transformations d'un format vers un autre pourront être importées telles quelles, d'une entreprise vers une autre. Un atelier permettant cela améliorera grandement la productivité de ces ingénieurs et, espérons le, la qualité des transformations utilisées. Cependant, il devra faire face au problème du contrôle de cette qualité.

## 5 Logiciels

### 5.1 Panorama

Les travaux de l'action EXMO sont susceptibles de donner lieu au développement de logiciels. Nous avons ainsi conçu et développé un système de spécification et d'exécution de flux de transformations (voir §

### 5.2 Transmorpher : spécification et exécution de flux de transformations XML

**Participants :** Jérôme Euzenat [Correspondant], Fabien Triolet.

Comme on a pu le voir plus haut, pour établir ou vérifier des propriétés sur des transformations, il faut en disposer d'une représentation. Le langage XSLT, en exprimant les transfor-

---

[BLHL01] T. BERNERS-LEE, J. HENDLER, O. LASSILA, « The semantic web », *Scientific american* 279, 5, 2001, p. 35-43, <http://www.scientificamerican.com/2001/0501issue/0501berners-lee.html>.

mations en XML, permet de faire cela à peu de frais, mais reste verbeux et délicat à analyser. Afin de résoudre ce problème, nous avons conçu et développons en collaboration avec la société FLUXMEDIA le logiciel TRANSMORPHER qui se présente comme une sur-couche à XSLT permettant d'exprimer des flux de transformations.

TRANSMORPHER [7] est un environnement permettant de définir et d'exécuter des transformations génériques sur des documents XML. Il a pour but de proposer des compléments du langage XSLT afin de :

- décrire simplement des transformations simples (suppressions d'éléments, remplacements de noms d'attributs, assemblage de documents, substitutions d'expressions régulières sur le contenu...);
- assembler des transformations en connectant leurs entrées et sorties (multiples);
- appliquer des transformations jusqu'à leur inapplicabilité;
- intégrer des moteurs de transformations externes.

TRANSMORPHER permet de décrire en XML des flux de transformations, c'est-à-dire des ensembles de transformations connectées par des canaux d'entrée-sortie. Les canaux véhiculent l'information à transformer, principalement du XML. Les transformations peuvent être d'autres flux de transformations ou des transformations élémentaires. TRANSMORPHER définit un ensemble d'abstractions de transformations élémentaires dotées d'un modèle d'exécution et des instanciations de ces transformations. On peut décliner ainsi les transformations élémentaires : appel externe (incluant XSLT), dispatcheur, agrégateur, sérialiseur, générateur, moteurs de requêtes, itérateurs et systèmes de règles.

Les systèmes de règles permettent de décrire des transformations dans un langage plus simple qu'XSLT. Actuellement, il est exécuté en le transformant directement en une transformation XSLT (étendue aux substitutions d'expressions régulières).

Cette année la version 1.0 de TRANSMORPHER réalisant toutes les fonctionnalités citées, a été achevée. Elle est disponible depuis <http://transmorpher.inrialpes.fr>. Un certain nombre de bibliothèques sont disponible avec cette première version, en particulier pour manipuler des documents encodant l'information personnelle (vCard pour les répertoires et vCal pour les agendas). À l'aide des ces bibliothèques plusieurs applications ont été développées pour engendrer les programmes de conférences sous de nombreux formats ou pour compléter les informations de répertoires en allant chercher les informations sur le web [17].

Par ailleurs, nous avons aussi spécifiée une extension de ce logiciel permettant de déclarer des propriétés et leur comportement par rapport aux constructeurs de transformations (préservées ou non). Ces comportements peuvent être spécialisés dans les implémentations des constructeurs génériques et les propriétés peuvent être attachées à chaque transformation (élémentaire, externe ou composée) permettant de signifier si une propriété est prouvée, supposée, à vérifier ou à démontrer.

## 6 Résultats nouveaux

### 6.1 Interopérabilité sémantique

**Participants :** Jérôme Euzenat [Correspondant], Heiner Stuckenschmidt [TZI Bremen].

Dans le cadre de la communication de représentations formalisées, on cherche à garantir la préservation du sens entre deux représentations. Plus précisément, entre deux langages  $L$  et  $L'$ , une représentation  $r$  de  $L$  va être transformée en une représentation  $\tau(r)$  dans  $L'$  (où  $\tau : L \rightarrow L'$ ). On adopte donc une approche consistant à étudier l'interopérabilité sémantique via une transformation.

Assurer l'intelligibilité sémantique peut être décrit à l'aide des deux équations complémentaires (que l'on peut comprendre comme la complétude et la correction de la transformation):

$$\forall r \subseteq L, \forall \delta \in L, r \models_L \delta \implies \tau(r) \models_{L'} \tau(\delta) \text{ et } \forall r \subseteq L, \forall \delta \in L, \tau(r) \models_{L'} \tau(\delta) \implies r \models_L \delta$$

Garantir l'interopérabilité sémantique dans toute sa généralité est hors de portée. Par conséquent, nous nous intéressons à des cas particuliers de ces équations.

#### 6.1.1 Hiérarchie de propriétés sémantiques

Nous avons entrepris d'inventorier des manifestations de l'équation (1) que l'on peut utiliser dans les transformations de langages à langages. Cela permet d'envisager différents cas de figure, notamment:

- $L'$  contient tous les constructeurs de  $L$ , dans ce cas, la traduction  $\tau$  est l'identité;
- $L$  contient strictement plus de constructeurs que  $L'$  (il est plus expressif). Dans ce cas, la seule solution n'est pas comme on pourrait penser l'appauvrissement syntaxique. Il faut, en toute théorie, mettre en oeuvre un démonstrateur pour déduire de  $r$  tout ce qui est exprimable en  $L'$ .
- Il existe une transformation de  $L$  vers  $L'$  préservant les interprétations des assertions. Nous avons exploré le cas où l'on disposait de preuves d'équivalence entre  $L$  et  $L'$  d'un type assez particulier : pour chaque constructeur d'un langage, on montre que toute expression l'utilisant est exprimable avec les constructeurs de l'autre. On a alors la définition constructive d'une transformation préservant les interprétations et que l'on peut tenter de développer en XSLT.
- Il existe une transformation de  $L$  vers  $L'$  mettant les modèles en correspondance. Ce cas correspond à la définition de l'expressivité proposée par [Baa96] appliquée à notre cadre (on a une transformation et non une fonction de renommage).
- Il existe une transformation de  $L$  vers  $L'$  telle que tout modèle de la représentation image inclue un modèle de la source modulo la transformation. Ceci généralise la propriété précédente et permet d'autoriser des représentations avec un nombre de modèles plus importants tout en préservant les conséquences.

---

[Baa96] F. BAADER, « A formal definition of the expressive power of terminological knowledge representation languages », *Journal of logic and computation* 6, 1, 1996, p. 33–54.

Ces propriétés se retrouvent dans des contextes divers (preuve d'équivalence entre logiques de descriptions, « interprétation » en logique du premier ordre, réductions polynomiales) et ont bien souvent été étudiées de manière séparée. À partir de cet inventaire, nous étudions les relations entre les propriétés ce qui permet de déterminer la propriété satisfaite par une succession de transformations utilisées pour passer d'un langage à un autre [13].

### 6.1.2 Familles de langages

Dans le cas d'une famille de langages, il existe un langage  $L \vee L'$  tel que toute formule de  $L$  ou de  $L'$  soit une formule de  $L \vee L'$ . De plus, l'interprétation des opérateurs de  $L$  sont les mêmes dans  $L \vee L'$  (et, par conséquent, les opérateurs présents dans les deux langages ont la même interprétation). Un cas de famille de langages bien connu est celui des logiques de descriptions où toute une hiérarchie de langages a été définie [Neb90].

Nous l'avons exploitée dans la construction du langage DLML (voir [projet EXMO, année 2001, section logiciels, module dlml]). Nous avons expérimenté avec elle l'établissement de propriétés sémantiques en utilisant des transformations entre deux langages dont on connaît les propriétés sémantiques. Affecter les propriétés aux transformations permet de rechercher un chemin (matérialisé par la composition d'une séquence de transformations) dans les langages de la famille de telle sorte de préserver une propriété particulière.

Le travail de l'année passée a été implémenté au sein de TRANSMORPHER (voir §

## 6.2 Treillis et information

**Participant :** Olivier Brunet [Correspondant].

On part du principe que les langages  $L$  sont munis d'un ordre  $\leq$  caractérisant la perte d'information entre deux représentations (voir §

On peut définir le cas particulier où les transformations sont des applications dans le même langage ( $\tau : L \rightarrow L$ ), comme le domaine d'interprétation d'une logique modale. Dans ce cas, l'interprétation d'une formule  $\phi$  est l'ensemble des représentations (les nœuds du treillis) dans lesquels elle est vraie et l'interprétation de la modalité  $K_i$  (connu de  $i$ , liée à une transformation  $\tau_i$ ) comme l'ensemble des représentations qui, une fois transformées par  $\tau_i$ , satisfont la formule à connaître ( $[K_i\phi] = x \in L; \tau_i(x) \in [\phi]$ ). Comme l'information croît de manière monotone en descendant le treillis, l'interprétation d'une formule est un ensemble clos pour  $\leq$ .

L'axiomatisation de cette structure révèle qu'il s'agit de la logique modale IS4 (qui ne satisfait ni le tiers exclus  $\phi \vee \neg\phi$ , ni l'axiome 5  $\neg K_i\phi \implies K_i\neg K_i\phi$  – du fait de la spécificité de la négation sur des sous-ensembles clos pour  $\leq$ ) [12, 5].

D'autre part, si l'on se restreint aux informations se rapportant à des agents (de la forme  $K_i\psi$ ), on obtient une autre logique (nommée OL [5]) légèrement différente de IS4. Nous en avons développé une formulation en calcul des séquents bénéficiant de la propriété d'élimination des coupures, ce qui permet d'envisager de façon efficace d'implémentation de méthodes automatiques de preuves concernant les propriétés d'échanges d'informations.

---

[Neb90] B. NEBEL, *Reasoning and revision in hybrid representation systems, Lecture Notes in Artificial Intelligence 422*, Springer Verlag, Berlin (DE), 1990.

## 7 Contrats industriels (nationaux, européens et internationaux)

### 7.1 Fluxmedia

**Participants :** Jérôme Euzenat, Fabien Triolet.

Le logiciel TRANSMORPHER (voir §

FLUXMEDIA a développé cette année une première version d'un environnement graphique au dessus de TRANSMORPHER.

## 8 Actions régionales, nationales et internationales

### 8.1 Actions nationales

#### 8.1.1 Action spécifique Web sémantique (CNRS)

**Participants :** Jérôme Euzenat [correspondant], Jean-François Baget.

Le département STIC du CNRS a lancé une série d'« actions spécifiques » dans le but de proposer des lignes d'actions. EXMO est impliqué dans celle liée au « web sémantique » pilotée par Jean Charlet, Philippe Laublet et Chantal Reynaud. À cette action collaborent aussi les membres d'ACACIA et d'ORPAILLEUR. Le but du travail de prospective mené à l'heure actuelle consiste à dégager les points forts des intervenants français et leur apport dans le « web sémantique » ainsi que leur insertion dans les travaux en cours au niveau international.

### 8.2 Actions européennes

#### 8.2.1 Réseau thématique OntoWeb: ontology-based information exchange for knowledge management and electronic commerce

**Participants :** Jérôme Euzenat [correspondant], Jean-François Baget, Raphaël Troncy, Bruno Charre, Heiner Stuckenschmidt.

EXMO est impliqué dans le réseau thématique ONTOWEB (“Ontology-based information exchange for knowledge management and electronic commerce”) financé par l'union européenne. Le nœud INRIA d'ONTOWEB est constitué du projet ACACIA et des actions AXIS, EXMO et ORPAILLEUR. Il a en charge le module dévolu à promouvoir la collaboration internationale sur les thématiques d'ONTOWEB. EXMO représente l'INRIA au sein du bureau du projet.

On trouvera plus d'informations sur ONTOWEB à <http://www.ontoweb.org>.

### 8.3 Actions internationales

#### 8.3.1 Groupe de travail WebOnt (W3C)

**Participants :** Jérôme Euzenat [correspondant], Jean-François Baget.

Nous avons commencé de représenter l'INRIA auprès du groupe de travail WEBONT de

l'activité « semantic web » du W3C. Ce groupe développe le langage de représentation de connaissance pour le « web sémantique » dénommé OWL (Langage d'ontologie pour le web). On trouvera plus d'informations sur WEBONT à <http://www.w3.org/2001/sw/webont> .

## 9 Diffusion de résultats

### 9.1 Animation de la communauté scientifique

- Jérôme Euzenat est membre élu du bureau de l'Association Française d'Intelligence Artificielle (AFIA).
- Jérôme Euzenat est membre fondateur de la « Semantic web science association » (comité de pilotage des conférences ISWC) et membre du comité de pilotage de la conférence RFIA 2004.
- Comité de rédaction de la revue « L'objet » et « Journal électronique d'intelligence artificielle (JEDAI) » (Jérôme Euzenat).
- Comité de programme des éditions 2002 des colloques « World-wide web conference », « International semantic web conference », « International conference on knowledge engineering and knowledge management (EKAW) », « Langages et modèles à objets », de la session « semantic web » de FLAIRS 2002, des séminaires « Ontology and information sharing » de l'ECAI 2002, « semantic web » de WWW 2002, et « Knowledge management and organizational memory » de l'ECAI 2002 (Jérôme Euzenat).
- Membre du bureau et coordinateur du package « Promoting world-wide collaboration » du réseau thématique européen OntoWeb (voir §

### 9.2 Enseignement

- « Sémantique des représentations de connaissance », DEA d'informatique, communication et système (ISC), université Joseph Fourier et INPG, Grenoble, 2×18h, Jérôme Euzenat et Jean-François Baget.

### 9.3 Participation à des colloques, séminaires, invitations

- Knowledge representation formalisms: presentation, comparison, and their integration in the web : Séminaire « sémantique » Xerox research centre europe, Meylan (FR), 18 janvier 2002, exposé oral (Jérôme Euzenat);
- Visite INTAP, Sophia-Antipolis, 15 janvier 2002, Exposé oral (Jérôme Euzenat);
- Le web sémantique : Séminaire action spécifique « web sémantique » CAMS, Paris (FR), 29 janvier 2002, exposé oral (Jérôme Euzenat et Amedeo Napoli);
- Les langages pour le web sémantique : Séminaire action spécifique « web sémantique » CAMS, Paris (FR), 29 janvier 2002, exposé oral (Jérôme Euzenat et Mohand Saïd-Hacid);
- New instruments for the semantic web : « Knowledge technologies » workshop, Commission européenne, Luxembourg (LU), 15-16 mai 2002, exposé oral (Jérôme Euzenat);
- Participation à la table-ronde « Research Directions for the Semantic Web », 3e Ontoweb meeting, Chia Laguna (IT), 13 juin 2002 (Jérôme Euzenat);

- XML, transformations (web sémantique) et réécriture, présentation au projet Prothéo, Nancy (FR), 4 décembre 2002 (Jérôme Euzenat);
- Web sémantique, représentation de connaissance et raisonnement: présentation groupe « modélisation du raisonnement » GDR I3, Nancy, 6 décembre 2002 (Jérôme Euzenat).
- Le web sémantique: séminaire du « Groupement des éditeurs de services en ligne », Paris (FR), 13 décembre 2002 (Jérôme Euzenat);

## 10 Bibliographie

### Ouvrages et articles de référence de l'équipe

- [1] R. DUCOURNAU, J. EUZENAT, G. MASINI, A. NAPOLI (éditeurs), *Langages et modèles à objets : état et perspectives de la recherche, collection "Didactique", 19*, INRIA, INRIA Rocquencourt (FR), 1998, <http://www.inria.fr/rrrt/d-019.html>.
- [2] J. EUZENAT, *Représentations de connaissance : de l'approximation à la confrontation*, Habilitation à diriger des recherches, Université Joseph Fourier, Grenoble (FR), janvier 1999.

### Livres et monographies

- [3] I. CRUZ, S. DECKER, J. EUZENAT, D. MCGUINNESS (éditeurs), *The emerging semantic web*, IOS press, Amsterdam (NL), 2002, <http://www.inrialpes.fr/exmo/papers/emerging/>.
- [4] J. EUZENAT, A. GÓMEZ PÉREZ, N. GUARINO, H. STUCKENSCHMIDT (éditeurs), *Proc. ECAI 2002 workshop on Ontologies and semantic interoperability, Lyon (FR)*, <http://ceur-ws.org/Vol-64/>.

### Thèses et habilitations à diriger des recherches

- [5] O. BRUNET, *Étude de la connaissance dans le cadre d'observations partielles : la logique de l'observation*, Thèse d'informatique, Université Joseph Fourier, Grenoble (FR), 2002, <ftp://ftp.inrialpes.fr/pub/exmo/thesis/these-brunet.pdf>.

### Articles et chapitres de livre

- [6] D. DENEUX, C. LERCH, J. EUZENAT, J. BARTHÈS, « Pluralité des connaissances dans les systèmes industriels », *in : Coopération et connaissance dans les systèmes industriels : une approche interdisciplinaire*, R. Soënen et J. Perrin (éditeurs), Hermès Science publisher, Paris (FR), 2002, ch. 4.
- [7] J. EUZENAT, L. TARDIF, « XML transformation flow processing », *Markup languages : theory and practice 3*, 3, 2002, p. 285–311.
- [8] J. EUZENAT, « Eight questions about semantic web annotations », *IEEE Intelligent systems 17*, 2, 2002, p. 55–62, <http://computer.org/intelligent/ex2002/x2055abs.html>.
- [9] J. EUZENAT, « An infrastructure for formally ensuring interoperability in a heterogeneous semantic web », *in : The emerging semantic web*, I. Cruz, S. Decker, J. Euzenat, et D. McGuinness (éditeurs), IOS press, Amsterdam (NL), 2002, p. 245–260.

**Communications à des congrès, colloques, etc.**

- [10] R. AL-HULOUE, O. CORBY, R. DIENG-KUNTZ, J. EUZENAT, C. MEDINA RAMIREZ, A. NAPOLI, R. TRONCY, « Three knowledge representation formalisms for content-based representation of documents », in : *Proc. KR 2002 workshop on Formal ontology, knowledge representation and intelligent systems for the world wide web (SemWeb)*, Toulouse (FR), 2002.
- [11] B. BACHIMONT, R. TRONCY, A. ISAAC, « Semantic Commitment for Designing Ontologies : A Proposal », in : *Proc. 13th international conference on knowledge engineering and knowledge management (EKAW)*, Siguenza (ES), 2002.
- [12] O. BRUNET, « A modal logic for observation-based knowledge representation », in : *Proc. FLoC workshop on Intuitionistic modal logic and applications (IMLA 2002)*, Copenhagen (DK), R. Goré, M. Mendler, V. de Paiva (éditeurs), p. 69–81, 2002, <ftp://ftp.inrialpes.fr/pub/exmo/publications/brunet2002a.pdf>.
- [13] J. EUZENAT, H. STUCKENSCHMIDT, « The ‘family of languages’ approach to semantic interoperability », in : *Proc. ECAI 2002 workshop on Knowledge Transformation for the Semantic Web*, Lyon (FR), B. Omelayenko, M. Klein (éditeurs), p. 92–99, 2002, <ftp://ftp.inrialpes.fr/pub/exmo/publications/euzenat2002e.pdf>.
- [14] R. TRONCY, A. ISAAC, « DOE : une mise en oeuvre d’une méthode de structuration différentielle pour les ontologies », in : *Actes 13e journées francophones sur Ingénierie des Connaissances (IC)*, Rouen (FR), p. 63–74, 2002.

**Divers**

- [15] « Research challenges and perspectives of the Semantic web », Sophia Antipolis (FR), 2002, <http://www.ercim.org/EU-NSF/semweb.html>.
- [16] « Semantic web special issue », ERCIM, Sophia-Antipolis (FR), 51, 2002, ERCIM News n°51, [http://www.ercim.org/publication/Ercim\\_News/enw51/](http://www.ercim.org/publication/Ercim_News/enw51/).
- [17] B. CHARRE, « Web sémantique et recherche d’informations personnelles », Paris (FR), 2002, DESS d’intelligence artificielle, Université Pierre et Marie Curie, Paris (FR), <ftp://ftp.inrialpes.fr/pub/exmo/reports/dessia-charre.pdf>.
- [18] J. EUZENAT, « Research challenges and perspectives of the semantic web », IEEE Computer society, New-York (NY US), 17, 5, 2002, IEEE Intelligent systems 17(5) :86-88, <ftp://ftp.inrialpes.fr/pub/exmo/reports/euzenat2002h.pdf>.