

Rapport INRIA 1994 — Programme 3
Bases de connaissances à objets

PROJET SHERPA

3 mai 1995

PROJET SHERPA

Bases de connaissances à objets

Localisation : *Grenoble*

Mots-clés : base de connaissances (1), catégorisation (1), classification (1), contrainte (1), diagnostic (1), modèle à objets (1), modèle de tâches (1), modélisation (1), modélisation en biologie (1), représentation de connaissances (1), résolution de problème (1), typage (1).

1 Composition de l'équipe

Responsable Scientifique

François Rechenmann, DR Inria

Personnel Inria

Jérôme Euzenat, CR

Pierre Fontanille, IE

CNRS

Gilles Bisson, CR

Université

Patrice Uvietta, IE MESR-INPG

Danielle Ziébelin, MC UJF

Chercheurs doctorants

Nathalie Beauboucher, allocataire MESR

Cécile Capponi, allocataire MESR

Jérôme Gensel, allocataire MESR

Pierre Girard, allocataire MESR
Florence Lemaire, allocataire MESR
Bruno Orsier, allocataire MESR
Olivier Schmeltzer, allocataire MESR
Nina Tayar, MAE, complément Inria
Jutta Willamowski, boursière Inria

Secrétariat

Danièle Herzog, Inria

Autres personnels

Alain Garreau, stagiaire CNAM, Grenoble

2 Présentation du projet

Sherpa est un projet commun entre l'Inria et le CNRS, localisé à Grenoble et intégré dans le laboratoire LIFIA (Laboratoire d'Informatique Fondamentale et d'Intelligence Artificielle) de l'Institut Imag. Dans de nombreux domaines scientifiques, l'accumulation des données et des connaissances nécessite de recourir à la constitution de grandes bases dans lesquelles les connaissances sont explicitées et exprimées à l'aide de modèles à la sémantique bien définie. Ces connaissances sont descriptives (sur les entités du domaine), comportementales (sur la dynamique de ces entités), méthodologiques (sur les méthodes employées pour découvrir et compléter ces entités) et terminologiques (sur les correspondances entre les noms utilisés dans la base et les termes du domaine).

Le projet *Sherpa* a pour premier objectif de contribuer à la définition de modèles de connaissances à objets susceptibles de satisfaire ces exigences, en réalisant les meilleurs compromis entre leur capacité d'expression et le bien fondé de leur sémantique. Ces modèles sont associés à des mécanismes d'exploitation conformes à leur sémantique, tels que la classification d'instances sur des points de vue multiples ou la décomposition de tâches. Le projet a ainsi spécifié et réalisé le modèle Tropes qui sert maintenant de base à plusieurs travaux de recherche plus spécifiques, sur les types, les contraintes ou les tâches.

De plus, le projet s'intéresse à la construction incrémentale et concurrente des bases de connaissances à objets, plus particulièrement aux moyens de confronter formellement les connaissances en provenance de plusieurs sources.

La biologie moléculaire constitue le terrain d'inspiration et d'expérimentation privilégié, mais pas exclusif, de ces travaux de recherche. À titre d'illustration, les connaissances descriptives portent alors sur les entités telles que les gènes, les signaux de régulation de leur expression ou encore sur des structures composites telles que les opérons. Les connaissances comportementales modélisent en particulier les mécanismes de régulation de l'expression génétique. Les connaissances méthodologiques portent sur les nombreuses méthodes d'analyse de séquences. Enfin, les connaissances terminologiques permettent de lier les descriptions de ces entités à la littérature scientifique qui les justifie.

3 Actions de recherche

Les recherches menées au sein du projet Sherpa s'organisent donc autour des modèles de connaissances à objets. Dans les modèles étudiés, les connaissances sont organisées dans autant de structures hiérarchiques distinctes que de concepts identifiés dans le domaine modélisé. Les structures hiérarchiques sont induites par la relation de spécialisation entre classes décrites à l'aide d'attributs typés. Les classes possèdent des instances, modèles d'objets du monde réel, dans lesquelles les attributs reçoivent des valeurs conformes à leurs types.

3.1 Classification

Participants : Gilles Bisson, Jérôme Euzenat

La classification est le mécanisme privilégié d'exploitation d'une base de connaissances à objets. Il permet de rechercher les classes d'un concept auxquelles une instance donnée est susceptible d'être rattachée. Ce mécanisme a été l'objet de travaux intensifs dans le projet Sherpa, travaux qui ont en particulier abouti à la classification multi-points de vue du système Tropes. Les travaux portent maintenant, d'une part sur l'abstraction des activités de classification à travers la notion de système classificatoire, d'autre part sur l'étude de la similarité en classification.

Systemes classificatoires

L'importance des mécanismes de classification dans les modèles à objets a conduit à la notion de système classificatoire, une abstraction

des actions de classification telles qu'elles sont exécutées dans les représentations de connaissances. Au lieu de partir de la structure de la représentation pour en inférer les mécanismes de classification, elle part de l'activité elle-même. Ceci permet de rendre compte de la classification dans un grand nombre de formalismes (logiques terminologiques, représentations par objets, analyse de données) et d'étudier leurs propriétés de manière générale. L'introduction d'opérateurs de produit et de projection permet de donner une définition constructive des algorithmes de classification à partir de types primitifs [11].

Mesure de similarité et classification

Les mécanismes de classification reposent essentiellement sur la notion de subsomption qui permet de comparer le degré de généralité de deux descriptions A et B données. De la sorte, la relation "A subsume B" exprime le fait que l'ensemble des individus décrits par A, c'est-à-dire l'extension de A, inclut l'ensemble des individus décrits par B. Ce travail vise à remplacer ce critère de subsomption par un critère de *similarité*, c'est-à-dire une mesure numérique permettant de quantifier les *ressemblances* et les *dissemblances* qui existent entre un couple d'objets.

Dès lors, les opérations de classification et de catégorisation d'objets complexes, c'est-à-dire de graphes composés d'objets simples connectés entre eux, se ramènent à la recherche d'un appariement qui soit "globalement optimal" au sens de la mesure de similarité utilisée, entre les différents composants de ces deux graphes. L'intérêt de la mesure qui a été développée est de rendre compte naturellement de la structure relationnelle du graphe. En d'autres termes, la similarité qui est mesurée entre deux objets reflète simultanément leur ressemblance propre et la ressemblance des objets qui leur sont connectés. De ce fait, il devient possible de réaliser un appariement des graphes qui préserve au mieux la structure relationnelle sans qu'il soit nécessaire d'effectuer une recherche arborescente ([7] et DEA de Jean-Marc Gabriel [24] sous la direction de Gilles Bisson).

3.2 Objets, types et contraintes

Participants : Jérôme Euzenat, Cécile Capponi, Jérôme Gensel

Le modèle de connaissances Tropes résulte des travaux antérieurs du projet Sherpa. Il se caractérise fondamentalement par la distinction on-

tologie / taxonomie et par la prise en compte explicite des points de vue multiples sur un même concept. Sur chaque point de vue, un graphe de spécialisation différent peut être développé, une même classe pouvant alors apparaître dans plusieurs points de vue simultanément. Ce modèle est le cadre de travaux fondamentaux sur la représentation de connaissances, plus particulièrement sur les types et les contraintes.

Types

L'adjonction d'un système de types à un modèle de connaissances à objets comme Tropes [10] présente de nombreux avantages liés à l'extensibilité, la connexion avec des modules procéduraux, ainsi que l'efficacité et la sécurité de mécanismes d'inférence tels que la classification.

Le système de types développé, appelé Meteo, comprend deux niveaux de typage :

- D'une part, des classes de types (appelés C-types) représentent des structures de données particulières (spécification d'ensembles de valeurs et d'opérations applicables à ces valeurs). De nouveaux C-types, tels que *matrice* ou *séquence génomique*, peuvent être intégrés par l'utilisateur ; ils seront considérés de la même façon que les C-types primitifs comme *entier*. Un C-type permet la description d'opérations de manipulation de ses valeurs, ainsi que la spécification d'opérations de manipulation de ses parties : c'est au niveau du C-type qu'est décrite, par exemple, la relation de sous-typage qui tient compte des propriétés de ses valeurs. De même, c'est au niveau du C-type que peut être décrite la mesure de similarité entre deux valeurs.
- D'autre part, à chaque C-type est associé un ensemble de Δ -types qui représentent les parties de l'ensemble caractérisé par le C-type. Les Δ -types sont organisés en treillis par la relation de sous-typage traduisant l'inclusion ensembliste. Par exemple, dans le cas du C-type *entier*, les Δ -types sont des listes d'intervalles fermés de valeurs entières ; la relation de sous-typage dérive de la relation d'inclusion entre listes d'intervalles.

Toute entité abstraite du modèle de représentation est typée :

- Les classes se voient associer un C-type si la spécification d'opérations particulières sur des éléments de cette classe est nécessaire,

ou un Δ -type issu du C-type *record* sinon. Les attributs sont typés par des Δ -types.

- Les instances elles-mêmes sont typées de façon à prendre en compte statiquement les contraintes qui portent sur elles. L'opération de typage d'une entité, d'une part prend en compte sa définition statique, d'autre part intègre dynamiquement les résultats donnés par le système de gestion des contraintes.

Le système de types considère de façon générique les opérations des C-types et les utilise pour l'interface entre le modèle de connaissances et la base de types. Il est constitué principalement d'une opération de typage (comprenant une phase de normalisation), d'une opération de test de sous-typage, d'opérations génériques telles que l'union, la disjonction ou l'intersection de types. Ces opérations génériques utilisent les opérations particulières de chaque C-type. Meteo réalise en outre la gestion dynamique des treillis de Δ -types. En conséquence, toute opération d'ajout, de suppression ou de modification d'un Δ -type est dynamiquement prise en compte par le système de types.

Les primitives offertes par le système de types sont alors directement utilisables par les mécanismes d'instanciation et d'inférence du modèle de connaissances tels que le filtrage, la classification de classe ou d'instance, la gestion de contraintes ou encore l'activation d'attachements procéduraux.

La propriété d'extensibilité permet l'intégration de nouvelles structures de données de façon à personnaliser la base de types pour une application particulière. Enfin, le modèle de représentation des connaissances manipule indifféremment des structures de données extraites, ou non, des entités de représentation, et ceci du fait de l'homogénéité des C-types qui sont une implémentation de ces abstractions.

Contraintes

La déclarativité et les capacités d'inférence d'un modèle tel que Tropes peuvent être étendues par la définition et la maintenance de relations mathématiques entre les attributs. Les contraintes peuvent être définies aux différents niveaux que sont les concepts, les classes et les instances. D'autre part, ces réseaux de contraintes sont dynamiques : des contraintes peuvent être ajoutées ou supprimées à tout moment.

Un Module d'Intégration de Contraintes et de Relations aux Objets (Micro) a été conçu. Couplé à Tropes, il y est chargé de la maintenance des réseaux de contraintes. Micro propose une panoplie de contraintes prédéfinies classiques (numériques et booléennes), mais aussi spécifiques aux attributs multivalués de Tropes (ensemblistes, sur des listes), ou plus originales (conditionnelles, d'évolution). Micro utilise une représentation par objets des contraintes et gère les liens entre les attributs contraints des instances de Tropes et les attributs des instances de contraintes associées. Les techniques de propagation reposent sur des méthodes génériques de maintien de la consistance associées aux classes de contraintes. La maintenance des contraintes numériques exploite les principes de l'arithmétique des intervalles. La résolution des contraintes est basée sur un algorithme de "forward-checking" paramétrable pour les domaines finis et sur une méthode de fractionnement dynamique pour les domaines infinis ou continus.

Les contraintes s'avèrent un puissant outil de maintien de la cohérence pour le modèle. La maintenance des contraintes assure la consistance du domaine d'un attribut contraint et, par là même, est amenée à modifier son type. Le type effectif d'un attribut contraint (dans le concept, la classe, ou l'instance) est alors calculé par Micro après la phase de propagation, puis transmis au module de gestion des types. Les contraintes étendent donc la définition des types des attributs.

Enfin, les contraintes ont été utilisées pour l'extension du modèle lui-même [13]. Ainsi, le partage de propriétés entre un objet composite et ses composants, le passage de paramètres entre une tâche et ses sous-tâches, la sémantique des liens de modélisation de relation et les requêtes effectuées à base de filtres, sont décrits et contrôlés par des contraintes.

Implémentation dans Tropes

Une version opérationnelle de Tropes a été développée en Talk (Ilog). Elle inclut les modules de typage et de satisfaction de contraintes décrits ci-dessus.

Au delà de cette implémentation, l'étude des rapports entre langages de programmation à objets et représentations de connaissances a conduit, dans le cadre du projet soutenu par le GREG (§3.6), à une proposition d'implémentation du modèle Tropes sur Icos [12], le langage à objets associé à Talk et proposé par Ilog.

3.3 Tâches et résolution coopérative de problèmes

Participants : Danielle Ziébelin, Nathalie Beauboucher,
Pierre Girard, Jutta Willamowski

La résolution d'un problème revient très souvent à choisir des modules algorithmiques, puis à enchaîner leurs exécutions. Les modèles de tâches permettent de décrire, d'une part les modules en termes des objets qu'ils acceptent en entrée et produisent en sortie, d'autre part des enchaînements de ces modules (stratégies) en vue de la résolution d'une classe de problèmes. La résolution proprement dite d'un problème consiste alors à exécuter la tâche associée en la décomposant récursivement et de façon opportuniste (c'est-à-dire en tenant compte des résultats intermédiaires obtenus) en sous-tâches plus élémentaires jusqu'à l'exécution directe de modules.

L'environnement de résolution de problèmes Scarp [2] met en œuvre un tel modèle. Le mécanisme de résolution repose sur une alternance de phases de décomposition et de caractérisation (par classification) de tâches. Scarp est utilisé à l'extérieur du projet ; il est en particulier au centre des développements de bases de connaissances méthodologiques par le laboratoire de biométrie, génétique et biologie des populations de Lyon (URA CNRS 243 et Université Claude Bernard) et par l'Institut Pasteur à Paris, et ce dans le cadre du projet soutenu par le GREG (§3.6).

Sémantique des tâches décomposables et spécialisables

Dans le cadre du stage de DEA d'Isabelle Crampé [22], sous la direction de Jérôme Euzenat, une sémantique opérationnelle de ce modèle de tâches a été établie. Elle permet de concilier un modèle de tâches aussi expressif que Scarp avec une sélection rationnelle des tâches à exécuter. La sélection rationnelle permet d'arbitrer, pour le même but à atteindre et dans une situation donnée, entre l'exécution immédiate d'une tâche déjà sélectionnée et une recherche coûteuse de tâches plus utiles (au sens de la théorie de la décision). Le choix est fondé sur l'espérance de gain d'utilité entre la méthode courante et des méthodes plus difficiles d'accès. Un moteur de tâches mettant en œuvre ce modèle a été méta-programmé en Prolog.

Ce travail a permis de mettre en évidence des aspects (utilité monocritère et indépendance de l'environnement) qui permettent de compiler statiquement le programme d'évaluation des tâches, par simple ordonnancement des clauses, au lieu d'évaluer dynamiquement l'espérance.

Similarité de tâches et reconnaissance d'intentions

Dans un environnement de résolution coopérative de problèmes tel que Scarp, la reconnaissance d'intentions intervient lors d'un échec de l'utilisateur à résoudre par lui-même une certaine tâche. Il s'agit alors de lui proposer une stratégie de résolution qui tienne compte de la stratégie qu'il a lui-même engagée avant qu'il ne se rende compte de son échec.

Fondamentalement, le problème de la reconnaissance d'intentions est donc un problème de mise en évidence de similarités entre réseaux de tâches : celui construit par l'utilisateur et ceux connus dans la base de tâches. Trois critères de similarité ont été dégagés (DEA d'Elsa Collin [21], sous la direction de Danielle Ziébelin) : une similarité portant sur le contenu des stratégies, une similarité portant sur leur structure, et une similarité portant sur leur résultat. La première est évaluée par application d'une méthode numérique/symbolique de mesure de similarité entre deux graphes de tâches. La seconde l'est en effectuant un test d'isomorphisme (au moyen d'un algorithme de recherche arborescente informée de type "branch and bound") sur des graphes de décomposition re-étiquetés. Enfin, la troisième se définit par extension de la notion de subsomption au cas des plans et des actions élémentaires qui les composent. Cette différenciation ne doit cependant pas masquer le caractère fortement corrélé de ces trois composantes, ni l'intérêt de les unifier en une unique mesure de similarité. Cette étude se concrétisera par l'adjonction de fonctions de similarité dans Scarp.

3.4 Liens avec les connaissances textuelles

Participants : François Rechenmann, Florence Lemaire

La construction d'une base de connaissances est très clairement un processus de modélisation. Son utilisation par des personnes autres que celles qui ont directement participé à sa conception requiert donc que soient explicités les choix et les hypothèses de modélisation. S'agissant de représenter des connaissances d'un domaine scientifique tel que la biologie moléculaire, ces choix et ces hypothèses reposent pour beaucoup

sur la littérature spécialisée. Il est donc nécessaire de relier les connaissances formalisées sous la forme d'objets aux connaissances textuelles telles que les articles de revue ou les monographies. Ce rôle d'interface peut être joué par un lexique.

D'un côté, les entités de la base possèdent des noms : noms de classes, d'instances, d'attributs ; de l'autre, les textes peuvent être indexés par des mots. Un lexique permet, d'une part de retrouver les noms des objets pertinents dans une base, et par là les objets eux-mêmes, à partir des mots employés par un utilisateur ou contributeur, d'autre part d'associer à des descriptions nommées dans la base de connaissances des textes, ou des fragments de textes, qui les commentent ou les justifient.

3.5 Construction incrémentale et concurrente de bases de connaissances

Participants : Gilles Bisson, Jérôme Euzenat, Nina Tayar

Le processus de construction d'une base de connaissances est nécessairement incrémental. Le projet Sherpa a déjà proposé des mécanismes d'aide à la définition et au positionnement de nouvelles classes dans une taxonomie. Les travaux visent maintenant à assister la construction d'une base à partir d'instances caractéristiques du domaine et ce à l'aide d'un mécanisme de catégorisation. D'autre part, le processus de construction est susceptible de faire intervenir simultanément plusieurs personnes. Les travaux portent donc également sur la mise en œuvre d'une architecture et de protocoles permettant la construction concurrente de bases de connaissances.

Explicabilité des processus de catégorisation

L'opération de catégorisation consiste à construire de manière automatique une hiérarchie de classes à partir d'un ensemble d'apprentissage constitué d'instances caractéristiques du domaine que l'on cherche à modéliser. Si dans les domaines simples, il est concevable que les données initialement fournies au système de catégorisation soient suffisamment complètes et correctes pour que celui-ci découvre par ses "propres facultés d'analyse" une modélisation satisfaisante du domaine, ce présupposé ne tient plus dès lors que l'on travaille dans des domaines complexes.

Dans ce cas, l'utilisateur peut vouloir supprimer ou fusionner des classes ou encore en modifier la définition.

Dès lors, pour obtenir une base de connaissances plus pertinente, deux solutions sont envisageables. Tout d'abord, on peut essayer de corriger directement les classes qui ont été produites par le système, mais ce faisant, on perd la correspondance qui existe entre l'ensemble d'apprentissage et la base de connaissances. En d'autres termes, lors de catégorisations ultérieures, toutes les modifications que l'on a pu introduire manuellement dans les règles sont perdues. Aussi, la meilleure façon de procéder consiste à raffiner les informations à partir desquelles la catégorisation s'est effectuée. Par exemple, on peut vouloir rajouter des instances, modifier le langage de description des instances ou encore changer les paramètres de contrôle du système. Dans ce cadre, la catégorisation est généralement constituée par une série d'itérations, au cours desquelles, en fonction des résultats obtenus, l'utilisateur va être amené à modifier les données initiales. Or, le problème, bien connu en analyse des données, est qu'il n'est pas du tout aisé de trouver les données à modifier et les modifications à apporter. De ce fait, les outils de catégorisation automatique ne peuvent être utilisés que par des personnes ayant une bonne connaissance des mécanismes mis en œuvre.

Le but de ce travail [8] est de définir des architectures de systèmes et des méthodes de catégorisation dans lesquelles les différents types de traitement effectués sont découpés de manière telle qu'il devienne possible d'expliquer à des utilisateurs néophytes les résultats de chaque étape de fonctionnement en fonction des informations obtenues lors des étapes précédentes. Ainsi, l'utilisateur peut remonter progressivement des sorties qu'il souhaite modifier vers les éléments en entrée qui ont engendré les structures fautives.

Communication entre bases de connaissances

Une architecture permettant la connexion de bases de connaissances à une base consensuelle a été proposée. Elle est dotée d'un protocole inspiré de celui de la soumission à une revue scientifique. Ce protocole devrait être implémenté entre bases Tropes à l'aide de KQML, un langage d'interaction entre bases de connaissances distribuées indépendant du langage de représentation.

Enfin, un système de gestion de versions de bases de connaissances a été proposé et implémenté pour le modèle Shirka, antérieurement développé au sein du projet et utilisé dans de nombreuses applications. Ce système s'appuie sur une structuration de la base en couches hiérarchiques ordonnées chronologiquement [14].

3.6 Modélisation des connaissances en biologie moléculaire

Participants : Gilles Bisson, François Rechenmann, Daniëlle Ziébelin, Alain Garreau, Florence Lemaire, Olivier Schmeltzer, Jutta Willamowski

Environnement coopératif d'aide à l'analyse de séquences génomiques

Le projet Sherpa est à l'origine d'un projet soutenu par le GIP GREG (Groupement de Recherches et d'Études sur les Génomes) et impliquant, outre le projet, la société Ilog et deux laboratoires de biologie (Biométrie à Lyon et Institut Pasteur à Paris). L'objectif est la réalisation d'un environnement coopératif d'aide à l'analyse de séquences génomiques. Les travaux sont prévus sur deux ans, 1994 et 1995.

Les travaux de cette première année ont consisté, d'une part à établir, en collaboration étroite avec la société Ilog, les spécifications d'un système générique de modélisation d'objets et de tâches apte à recevoir et à exploiter les connaissances méthodologiques sur l'analyse de séquences, d'autre part, en collaboration avec les équipes du laboratoire de biométrie à Lyon et de l'Institut Pasteur à Paris, à développer des bases de connaissances en utilisant l'environnement Scarp. Lors de la deuxième année, ces bases seront retranscrites dans le nouveau système réalisé, puis étendues.

Interface cartographique générique

Afin de visualiser et de manipuler les objets en entrée et en sortie des différentes méthodes d'analyse de séquences, une interface cartographique générique est développée par Alain Garreau dans le cadre de son stage d'ingénieur CNAM. L'interface permet d'afficher des cartes à différentes échelles et de natures différentes, tout en maintenant des liens entre

les différentes vues d'un même objet sur plusieurs de ces cartes. De nombreuses fonctionnalités sont disponibles, par exemple pour "zoomer" sur une partie de carte ou pour cacher certains types d'objets afin de comparer rapidement les résultats en provenance de méthodes concurrentes.

Modélisation de cartes génomiques

Un travail plus générique et formel a été mené par Olivier Schmeltzer dans le cadre de sa thèse. La modélisation de cartes génomiques pose en effet de nombreux problèmes de représentation et de traitement. Les premiers proviennent de l'existence de plusieurs descriptions des entités du génome, les seconds de la complexité algorithmique de la construction des cartes, ajoutée à la nécessité de pouvoir traiter les incohérences issues des expériences. Dans un premier temps, il s'agissait de préciser la notion de carte grâce à une formalisation qui spécifie comment sont construites les cartes génomiques et quelles sont les relations entre les entités qui y apparaissent. Dans un second temps, un algorithme de construction de cartes à partir de la description des relations entre les entités qui la constituent a été proposé, en s'appuyant sur des techniques de raisonnement temporel.

3.7 Aide au diagnostic électromyographique

Participantes : Danielle Ziébelin, Nathalie Beauboucher

Le projet Sherpa est impliqué dans le développement d'un système à base de connaissances, Myosys, permettant d'effectuer un diagnostic électromyographique (EMG) portant sur l'identification des pathologies neuro-musculaires ([19], [18]). Ce projet est mené en étroite collaboration avec le service du Dr. Annick Vila au CHRU de Grenoble. Il s'inscrit dans le cadre du projet européen AIM Esteem ; il a également bénéficié d'un soutien de la Région Rhône-Alpes de février 1993 à février 1995.

Le diagnostic EMG s'effectue à partir d'une acquisition systématique de données numériques et symboliques. Il se décompose en un ensemble d'étapes clairement définies : formulation d'hypothèses, génération d'un protocole d'examen spécifique au cas traité, évaluation des résultats d'examen, validation ou remise en cause de l'hypothèse courante, élaboration d'une conclusion. Chacune de ces étapes est constituée d'un

concept (symptômes, hypothèses, protocoles, résultats d'examens) et d'une action (évoquer une hypothèse, générer un protocole, évaluer des résultats...). Un concept est représenté par une hiérarchie de classes et une action est représentée par une tâche qui se décompose en une liste de sous-tâches de plus en plus simples jusqu'à l'association d'une tâche élémentaire à un mécanisme d'inférence de base opérant sur les concepts précédemment cités (créer une instance, classer une instance dans une hiérarchie de classes, spécialiser une instance...). L'ensemble du système est développé dans l'environnement Scarp.

Dans le prolongement des travaux de thèse de Nathalie Beauboucher [1], il s'agit maintenant d'intégrer à Myosys un module de raisonnement par cas qui s'appuie sur une représentation explicite, sous forme de graphes de tâches, du raisonnement du médecin électromyographe.

4 Action industrielles

Dans le cadre du projet soutenu par le GIP GREG, le projet Sherpa travaille en étroite collaboration avec la société Ilog pour la spécification et le développement d'un système générique de modélisation de tâches et d'objets multi-points de vue. Il s'agit de valoriser les travaux menés ces dernières années par le projet sur le modèle Tropes (modèles d'objets multi-points de vue) et le système Scarp (modèle de tâches).

5 Actions nationales et internationales

- Le projet Sherpa participe au PRC-GdR Intelligence Artificielle. François Rechenmann y anime le groupe de travail "Classification et objets".
- Le projet est membre du GdR 1029 CNRS "Informatique et génomes", dont l'objectif est de permettre la confrontation de méthodes informatiques à des problèmes d'analyse de séquences génomiques et le traitement distribué de problèmes soumis par les experts biologistes du groupement. François Rechenmann en est le directeur adjoint et y anime, avec Christian Gautier (laboratoire de biométrie, Lyon) un groupe de travail sur "Modélisation des connaissances en biologie moléculaire".

- Le GIP GREG (Groupement de Recherches et d'Études sur les Génomes) soutient un projet de développement d'un système coopératif d'aide à l'analyse de séquences génomiques. Ce projet, dont les objectifs sont détaillés au paragraphe 3.6, rassemble sur 2 ans (1994 et 95), outre l'Inria Rhône-Alpes à travers le projet Sherpa : la société Ilog, le laboratoire de biométrie, génétique et biologie des populations de Lyon (équipe de Christian Gautier) et l'Institut Pasteur de Paris (équipe d'Antoine Danchin). François Rechenmann en est le responsable scientifique.
- L'objectif du projet européen AIM n° 2010 "Esteem" (European Standardised Telematic tool to Evaluate EMG knowledge-based systems and Methods) est de développer, puis de mettre à disposition des centres d'examen cliniques électromyographiques, une plate-forme commune regroupant un ensemble de systèmes d'aide au diagnostic en électromyographie (EMG). L'ensemble des médecins et informaticiens de ce projet travaillent à l'évaluation, la validation et l'intégration des différents composants de cette plate-forme : base de données EMG, interface avec l'équipement de traitement du signal, module de télécommunication, systèmes à base de connaissances, boîte à outils d'évaluation.

Le projet Sherpa est impliqué, sous la direction de Danielle Ziébelin, dans le développement d'un système à base de connaissances, Myosys, destiné à assister le médecin novice dans le choix de protocoles d'examens et l'élaboration d'un diagnostic électromyographique.

En plus des équipes grenobloises, les partenaires impliqués dans le projet Esteem sont :

- trois industriels (Computer Resources International, Copenhague ; Dantec Electronic, Copenhague ; SIPRA, Gand) ;
- six laboratoires de recherche (Uninova, Lisbonne ; Hospital de Egas Moniz, Lisbonne ; Hvidovre Hospital et Rigshospitalet, Université de Copenhague ; Newcastle General Hospital ; Universit tkliniken, Mayence ; Universit  de Bologne).

Le consortium a propos  la continuation du projet Esteem pour les trois ann es   venir, afin d'int grer les diff rents modules r alis s dans un environnement hospitalier multim dia.

- Le projet Sherpa a présenté des démonstrations dans le cadre du *Village scientifique* à TEC'94 (Grenoble, du 11 au 14 octobre).

6 Diffusion des résultats

6.1 Enseignement

Enseignement universitaire

- DEA Informatique, INP et Université Joseph Fourier, Grenoble, 12h. sur “Représentation des connaissances”, François Rechenmann
- DEA de Sciences Cognitives, INP, Grenoble, 18h. sur “Représentation des connaissances et modèles du raisonnement”, Gilles Bisson, Jérôme Euzenat et François Rechenmann
- DEA de Recherche Opérationnelle, Université Joseph Fourier, Grenoble, 12h. sur “Introduction aux systèmes à base de connaissances”, Danielle Ziébelin
- DESS Informatique, Université Joseph Fourier, Grenoble, 18h. sur “Intelligence Artificielle”, Danielle Ziébelin
- Troisième année Ensimag, INP, Grenoble, 24h. sur “Systèmes d'aide à la décision”, Danielle Ziébelin
- François Rechenmann est responsable des cours “Intelligence artificielle”, au sein de la “voie d'approfondissement informatique”, à l'ENTPE (École Nationale des Travaux Publics de l'État), à Vaulx-en-Velin, près de Lyon. Jérôme Euzenat y intervient sur la modélisation des connaissances et des raisonnements.

6.2 Participation à des conférences et colloques

- François Rechenmann, organisateur du séminaire d'évaluation du programme 3 de l'Inria (“Intelligence artificielle, systèmes cognitifs et interaction homme-machine”), Grenoble, 3-4 février 1994
- Gilles Bisson, *Apprentissage de base de connaissances en logique des prédicats*, séminaire au Laforia, 10 février 1994
- Gilles Bisson, *Le système KBG*, séminaire chez Thomson-RCM, 10 février 1994

- Gilles Bisson, *Une approche symbolique/numérique de la notion de similarité*, présentation invitée aux 4es Journées sur l'Induction Symbolique/Numérique, Orsay, 14-15 mars 1994
- Gilles Bisson, *KBG: Induction de base de connaissances en logique des prédicats*, séminaire à l'Inria Sophia-Antipolis, action Orion, 16 mars 1994
- Olivier Schmeltzer, *Modélisation de cartes génomiques*, réunion du groupe de travail "Modélisation des connaissances" du GdR "Informatique et Génomes", Lyon, 18 mars 1994
- Gilles Bisson, *Apprentissage symbolique et explication*, séminaire au Laforia pour le groupe "Explication" du PRC-GdR IA, 15 avril 1994
- Florence Lemaire a présenté le poster *Object-Based Knowledge Models for Genomic Information*, lors du workshop de l'ACM sur *Information Retrieval and Genomics*, Bethesda, USA, 2-4 mai 1994
- Séminaire, avec des exposés de Jacques Lebbe (Univ. Paris 6), Philippe Martin (projet Acacia, Inria Sophia-Antipolis) et Florence Lemaire sur *Connaissances formelles et textuelles*, Grenoble, 26 mai 1994
- Gilles Bisson, *Conceptual Clustering + Similarity Measure = KBG: A FOL Learning Tool*, invitation au GMD par Werner Emde dans l'équipe "Intelligence Artificielle", Bonn, 15-16 juin 1994
- Olivier Schmeltzer, *Modélisation de cartes génomiques*, Forum Inter-Disciplinaire "Génome et Informatique" (FIDGI), Aussois, 15-17 juin 1994
- François Rechenmann, intervention sur *Perspectives in knowledge representation: toward concurrent and incremental design of integrated knowledge bases*, lors du séminaire MOKEM-net (MOdelling and Knowledge base for an Ecologically oriented stream Management: towards a European network?), Arles, 17-22 juin 1994
- Jérôme Euzenat, intervention au séminaire de formation permanente "autour de l'objet", organisé par INA-pg, INRA, IUT Nice, 3 juillet 1994

- Olivier Schmeltzer, poster sur *Genomic Map Modeling: A Generic Formalization*, ISMB'94, Stanford, CA, 14-17 août 1994
- Gilles Bisson, cours sur *Classification conceptuelle* à Dourdan (Paris) dans le cadre de l'école européenne organisée par le MLnet : "Machine Learning Summer School", 5-10 septembre 1994
- François Rechenmann, journée Inria - Industrie (prog. 3), intervention sur *Systèmes coopératifs de résolution de problèmes*, Paris, 29 septembre 1994
- François Rechenmann, colloque "Informatique et biologie moléculaire" (7es Entretiens Jacques Cartier), intervention sur *Object-based modeling of biological knowledge*, Lyon, 2 décembre 1994

6.3 Organisation de colloques et de cours

- Dans le cadre de la "27th Hawai International Conference on System Sciences", François Rechenmann était co-organisateur, avec Dong-Guk Shin de l'Université du Connecticut, du "Minitrack on Data and Knowledge Base Issues in Genomics" (Hawaï 4-7 janvier 1994). Une quinzaine de communications ont été sélectionnées et ont fait l'objet de présentations. Les actes du "minitrack" ont été publiés dans le 5e volume (*Biotechnology Computing*) des actes de la conférence.
- Le colloque "Langages et Modèles à Objets" a rassemblé, les 13 et 14 octobre à Grenoble (dans le cadre de TEC'94), plus de 60 participants. François Rechenmann était le président du comité de programme et du comité d'organisation. Ce colloque était patronné par l'Inria Rhône-Alpes, l'IMAG, le groupe "Classification et objets" du PRC-GdR IA et le groupe "Évolution des langages à objets" du PRC Programmation. Les actes ont été édités.
- Dans le cadre des Septièmes Entretiens Jacques Cartier, François Rechenmann a co-organisé, avec Christian Gautier de l'université Claude Bernard de Lyon, Robert Cedergren du département de biochimie et David Sankoff du centre de recherches mathématiques, tous deux de l'université de Montréal, un colloque sur le thème "Informatique et biologie moléculaire", à Lyon les 30 novembre, 1er et 2 décembre. Une trentaine de spécialistes européens, canadiens et nord-américains ont été invités. Ce colloque était patronné par l'Inria Rhône-Alpes, l'IMAG et le GdR 1029

du CNRS “Informatique et génomes”. Les actes rassemblant les résumés étendus des interventions ont été édités.

7 Publications

Thèses

- [1] N. BEAUBOUCHER, *ANAÏS: raisonnement à partir de cas en résolution de problèmes*, Thèse, Université Pierre et Marie Curie, Paris, 28 juin 1994.
- [2] J. WILLAMOWSKI, *Modélisation de tâches pour la résolution de problèmes en coopération système-utilisateur*, Thèse, Université Joseph Fourier, Grenoble, 6 avril 1994.

Articles et chapitres de livre

- [3] C. CAPPONI, «Interactive Class Classification Using Types», *in: New Approaches in Classification and Data Analysis*, E. Diday, Y. Lechevallier, M. Schader, P. Bertrand, et B. Burtschy (éd.), Springer-Verlag, Berlin, 1994, p. 204–211.
- [4] F. RECHENMANN, «Building and sharing large knowledge bases in molecular biology», *in: Knowledge building and knowledge sharing*, K. Fuchi et T. Yokoi (éd.), IOS Press, Amsterdam (NL), 1994, p. 290–311.
- [5] F. RECHENMANN, «Modelling mathematical objects in knowledge-based systems for scientific computing», *in: Artificial intelligence in mathematics*, J. Johnson, S. McKee, et A. Vella (éd.), Clarendon Press, Oxford (GB), 1994, p. 183–194.
- [6] J. WILLAMOWSKI, F. CHEVENET, F. JEAN-MARIE, «A development shell for cooperative problem-solving environments», *in: Mathematics and Computers in Simulation*, R. Vichnevetsky (éd.), Elsevier Science NH, 1994, p. 361–379.

Communications à des congrès, colloques, etc.

- [7] G. BISSON, «Définition de la notion de similarité dans les modèles à objets», *in: actes 1e conférence “Langages et modèles à objets”*, F. Rechenmann (éd.), INRIA-IMAG, p. 53–68, Grenoble, 13-14 octobre 1994.
- [8] G. BISSON, «From Inductive Learning to Knowledge Acquisition through Explanations», *in: ECAI workshop on “Integration of Machine Learning and Knowledge Acquisition”*, C. Nédellec (éd.), p. 1–11, Amsterdam (NL), 8-12 août 1994.

- [9] G. BISSON, «Une approche symbolique/numérique de la notion de similarité», *in: actes 4es journées sur l'“Induction symbolique/numérique”*, E. Diday, Y. Kodratoff (éd.), p. 93–96, Orsay, 14-15 mars 1994.
- [10] C. CAPPONI, «Exploitation des types dans un modèle de représentation des connaissances par objets», *in: Actes 9e RFIA*, p. 171–183, Paris, 11-14 janvier 1994.
- [11] J. EUZENAT, «Classification dans les représentations par objets : produits de systèmes classificatoires», *in: Actes 9e RFIA*, p. 185–196, Paris, 11-14 janvier 1994.
- [12] J. EUZENAT, «KR and OOL co-operation based on semantics non reducibility», *in: Actes 11th ECAI workshop on integrating object-orientation and knowledge representation*, Amsterdam (NL), 8 août 1994. (sans pagination).
- [13] J. GENSEL, P. GIRARD, O. SCHMELTZER, «Intégration de contraintes, d'objets composites et de tâches dans un modèle de représentation de connaissances par objets», *in: Actes 9e RFIA*, p. 281–292, Paris, 11-14 janvier 1994.
- [14] F. LEMAIRE, N. TAYAR, «Gestion de la construction incrémentale d'une base de connaissances consensuelle», *in: Actes 2es rencontres nationales des jeunes chercheurs en intelligence artificielle*, p. 319, Marseille, septembre 1994.
- [15] B. ORSIER, B. AMY, V. RIALLE, A. GIACOMETTI, «A Study of the Hybrid System SYNHESYS», *in: Actes 11th ECAI workshop on combining symbolic and connectionist processing*, p. 1–9, Amsterdam (NL), 9 août 1994.
- [16] J. WILLAMOWSKI, «Coopération système-utilisateur pour la résolution de problèmes», *in: Systèmes à base de connaissances coopératifs*, Actes de l'atelier IA 94, p. 23–26, Paris, 31 mai 1994.
- [17] J. WILLAMOWSKI, «Modélisation de tâches pour la résolution de problèmes en coopération système-utilisateur», *in: Actes 9e RFIA*, p. 305–316, Paris, 11-14 janvier 1994.
- [18] D. ZIÉBELIN, P. VALTCHEV, I. IORDANOVA, A. VILA, «Classification of complex objects in a knowledge-based diagnosis system», *in: Actes 6th international conference on artificial intelligence: methodology, systems, applications*, p. 269–279, Sofia (BU), 21-24 septembre 1994.
- [19] D. ZIÉBELIN, A. VILA, V. RIALLE, «Neuromyosys, a diagnosis knowledge based system for EMG», *in: Actes 12th international congress of Medical Informatics in Europe*, p. 156–161, Lisboa (PT), 22-26 mai 1994.

Rapports de recherche et publications internes

- [20] C. CHEVALIER, «Construction incrémentale de bases de connaissances à objets», *Mémoire d'ingénieur*, CNAM, Grenoble, décembre 1994.
- [21] E. COLLIN, *Étude de similarité de raisonnements représentés à l'aide de graphes de tâches*, Mémoire, INP-Université Joseph Fourier, Grenoble, juin 1994, DEA d'informatique.
- [22] I. CRAMPÉ, *Sémantique des tâches décomposables et spécialisables*, Mémoire, INP-Université Joseph Fourier, Grenoble, juin 1994, DEA d'informatique.
- [23] J. EUZENAT, «Granularité dans les représentations spatio-temporelles», *Rapport de recherche n°2242*, Inria, Grenoble, avril 1994, 62 pages.
- [24] J.-M. GABRIEL, *Définition de la notion de similarité dans les modèles à objets*, Mémoire, INP, Grenoble, juin 1994, DEA de sciences cognitives.
- [25] S. MARTIN-MICHELLOT, *Apport de l'analogie dans un système de résolution de problèmes*, Mémoire, INP, Grenoble, juin 1994, DEA de sciences cognitives.

8 Abstract

In various scientific domains, such as molecular biology, the important flow of experimental data leads to the design and the management of large knowledge bases. Four categories of knowledge can be identified: descriptive knowledge on the entities of the domains; behavioral knowledge on the dynamic behavior of these entities; methodological knowledge on the methods which can be used to identify these entities and to complete their descriptions; terminological knowledge on the relationships between the names used in the knowledge base and the terms used in the domain.

The first objective of the Sherpa research project is the specification of object-based knowledge models which satisfy these requirements and offer a good trade-off between a high level of expressivity and a well founded semantics. The incremental and cooperative building of large knowledge bases is another line of research in the project. More precisely, the problem is the formal comparison, and merge, of contributions from several knowledge sources into a single consensual knowledge base.

Molecular biology is the main domain of experimentation of these research activities. In this context, descriptive knowledge concerns the

entities such as genes, regulation signals or operons. Behavioral knowledge models the dynamics of the expression regulation mechanisms. Methodological knowledge is attached to the various sequence analysis methods. Terminological knowledge allows to link the knowledge which is formalized in the base to the scientific literature which supported the modeling process.

Table des matières

1	Composition de l'équipe	1
2	Présentation du projet	2
3	Actions de recherche	3
3.1	Classification	3
3.2	Objets, types et contraintes	4
3.3	Tâches et résolution coopérative de problèmes	8
3.4	Liens avec les connaissances textuelles	9
3.5	Construction incrémentale et concourante de bases de connaissances	10
3.6	Modélisation des connaissances en biologie moléculaire . .	12
3.7	Aide au diagnostic électromyographique	13
4	Action industrielles	14
5	Actions nationales et internationales	14
6	Diffusion des résultats	16
6.1	Enseignement	16
6.2	Participation à des conférences et colloques	16
6.3	Organisation de colloques et de cours	18
7	Publications	19
8	Abstract	21