Construction concourante de bases de connaissances associant objets et hyper-textes

François RECHENMANN
Directeur de recherche
INRIA Rhône-Alpes
Francois.Rechenmann@inria.fr

Résumé
Comme tout modèle, une base de connaissances est difficilement utilisable par des personnes qui n'ont pas participé à sa construction, sauf à la replacer dans son contexte de conception décrit par des nœuds d'un réseau hyper-média reliés aux descriptions formelles qu'elle rassemble.

Mots-clés
base de connaissances, modèle à objets, hyper-texte, conception concourante, partage de connaissances, biologie moléculaire

Introduction - Les systèmes à base de connaissances

Les systèmes à base de connaissances sont des systèmes informatiques pour lesquels est recherchée une séparation explicite entre les connaissances nécessaires à la résolution d'une classe de problèmes et les mécanismes d'exploitation de ces connaissances pour résoudre un problème donné. Les avantages attendus d'une telle séparation sont, d'une part la possibilité de produire des explications sur le processus de résolution et des justifications des solutions proposées, d'autre part la facilité d'évolution de la base de connaissances.

La production d'explications et de justifications permet de faire reposer les processus de résolution de problèmes sur des connaissances non consensuelles. C'est ainsi que les systèmes à base de connaissances autorisent le développement de systèmes dits "experts" dans le sens qu'ils incorporent des connaissances propres à certains individus, car fruits de leur expérience personnelle.

L'explicitation des connaissances suppose quant à elle l'existence d'un modèle[1] de connaissances (de même qu'une base de données utilise un modèle de données, qu'il soit réseau, hiérarchique, relationnel ou à objets). Les connaissances s'exprimant à l'aide de ce modèle indépendamment de la manière dont elles seront ultérieurement exploitées, leur compréhension par un lecture humain est bien meilleure que lorsqu'elles sont noyées dans un programme et inextricablement liées avec les connaissances de mise en œuvre.

---

[1] Le terme "modèle" est employé ici dans deux sens différents : d'une part dans l'expression "modèle de connaissances" pour désigner le formalisme dans lequel s'expriment les connaissances d'une base, d'autre part pour désigner le résultat d'un processus de modélisation, ici le contenu d'une base de connaissances.

Les bases de connaissances comme modèles

Au delà de ces avantages immédiats et pratiques du principe même de cette catégorie de systèmes informatiques, la modélisation explicite et formelle des connaissances contribue à l'accroissement de la compréhension des domaines concernés. En effet, après qu'il ait été naïvement cru qu'il suffisait, par un processus néanmoins long et délicat, de recueillir les connaissances auprès des personnes qui les détenaient ou des sources bibliographiques qui les contenaient, il est maintenant très largement reconnu que les connaissances doivent de plus faire l'objet d'une élaboration. De fait, la construction d'une base de connaissances s'apparente étroitement à la construction d'un modèle. Des bases de connaissances sont maintenant construites dans un but essentiel de modélisation, la résolution de problèmes étant alors un objectif secondaire, voire absent. L'exemple des bases de connaissances développées en biologie moléculaire est significatif[2].

En tant que modèle, la pertinence et l'utilité d'une base de connaissances dépend ainsi très largement des objectifs qui président à sa conception. De plus, il est difficile et dangereux de faire utiliser une base de connaissances par des personnes autres que celles qui ont participé à son développement. Difficile, parce que les connaissances formalisées, même énoncées à l'aide d'un langage déclaratif, sont délicates à interpréter. Dangereux, car les objectifs qui ont sous-tendu la conception de la base ne correspondent sans doute pas aux attentes des utilisateurs non concepteurs. Enfin, l'utilisateur qui réutilise une base de connaissances est de fait privé de l'accroissement de compréhension du domaine qui résulte de son élaboration.

Il est donc nécessaire de partager le processus même de modélisation afin que la base de connaissances qui en est résultée soit à son tour partageable.

Conception incrémentale et concourante de bases de connaissances

Partager, au sens strict, le processus de modélisation consiste à permettre que plusieurs personnes contribuent simultanément à la construction d'une base de connaissances. Concrètement, il s'agit de mettre en œuvre une architecture distribuée et des protocoles d'échanges[3] afin que les concepteurs puissent mettre en commun et confronter leurs contributions respectives (Figure 1). Si ces contributions sont exprimées formellement (sous la forme de réseaux d'objets complexes), elles peuvent être comparées entre elles et avec le contenu de la base. Il est ainsi possible de décider si deux descriptions sont contradictoires, redondantes, semblables (en utilisant une mesure de similarité pertinente[4]), ou sont en relation de subsumption.

---

2 F. Rechenmann "Building and sharing large knowledge bases in molecular genetics", actes 1st International Conference on *Building and Sharing of Very Large-Scale Knowledge Bases* (KBKS), Tokyo (JP), pp291-301, (1-4 décembre) 1993
`ftp://ftp.imag.fr/imag/SHERPA/articles/rechenmann93.ps.gz`

3 J. Euzenat, "Building Consensual Knowledge Bases : Context and Architecture", dans Nicolaas Mars (éd.), *Towards very large knowledge bases* (actes 2nd international conference on building and sharing very large-scale knowledge bases (KBKS), Enschede (NL), (10-13 avril) 1995) IOS press, Amsterdam (NL), p. 144-155, 1995

4 G. Bisson, "Why and How to Define a Similarity Measure for Object-based Representation Systems", dans Nicolaas Mars (éd.), *Towards very large knowledge bases* (actes 2nd international conference on building and sharing very large-scale knowledge bases (KBKS), Enschede (NL), (10-13 avril) 1995) IOS press, Amsterdam (NL), p. 236-246, 1995

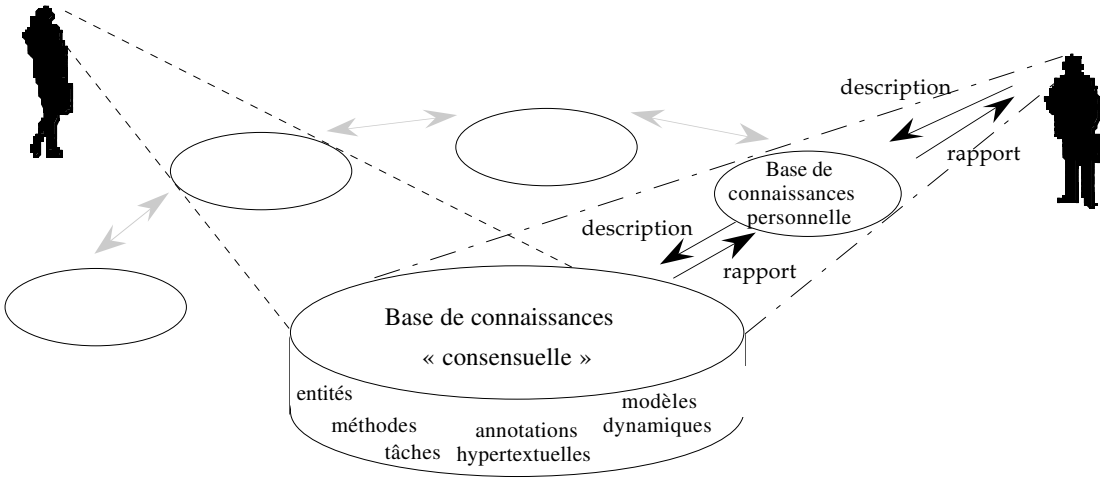

Figure 1 – Construction incrémentale et concourante de bases de connaissances. Les concepteurs soumettent leurs contributions qui sont confrontées au contenu de la base en cours d'élaboration. Les résultats de ces confrontations formelles sont renvoyés aux contributeurs qui peuvent décider de modifier leur contribution ou d'en discuter avec leurs collègues. Un protocole adéquat permet de garantir la convergence du processus vers une base de connaissances consensuelle.

Les bases de connaissances et leurs contextes de conception

Une autre approche du partage d'une base de connaissances consiste à l'envelopper de son contexte, constitué d'une part des sources de connaissances, humaines et bibliographiques, utilisées, d'autre part de l'énoncé des objectifs, des hypothèses et des choix de modélisation. Il s'agit en pratique de relier les énoncés formels de la base à des énoncés en langage naturel. Très naturellement, une structuration de ce contexte en hyper-texte (au sens large, c'est-à-dire pouvant inclure des figures, des images fixes ou animées, et du son) s'impose.

À titre d'exemple, les classes de plus haut niveau de la base de connaissances à objets ColiGène[5], qui décrit le génome de la bactérie *E. coli* du point de vue des mécanismes de régulation de l'expression génétique, ont été reliés aux nœuds d'un réseau hyper-texte. L'utilisateur de la base de connaissances peut ainsi naviguer aussi bien dans le réseau des objets de la base que dans le réseau hyper-textuel. Il peut également passer de l'un à l'autre dès lors qu'un même nom désigne simultanément une entité de la base (classe, instance ou attribut) et un nœud de l'hyper-texte (Figure 2). Ce processus de navigation permet au lecteur de prendre connaissance de la base de connaissances.

L'articulation entre les deux réseaux peut être assouplie par l'insertion d'un lexique rassemblant tous les termes significatifs pour la base et son contexte[6]. Ce lexique offre par ailleurs de nombreux avantages liés au processus de construction de la base et à l'interaction avec l'utilisateur.

---

5 G. Perrière, C. Gautier "ColiGene: Object-centered representation for the study of *E. coli* gene expressivity by sequence analysis", *Biochimie*, vol. 75, 5, p. 415-422, 1993

6 F. Lemaire "Knowledge bases, texts and lexicon", dans Nicolaas Mars (éd.), *Towards very large knowledge bases* (actes 2nd international conference on building and sharing very large-scale knowledge bases (KBKS), Enschede (NL), (10-13 avril) 1995) IOS press, Amsterdam (NL), p. 281-288, 1995

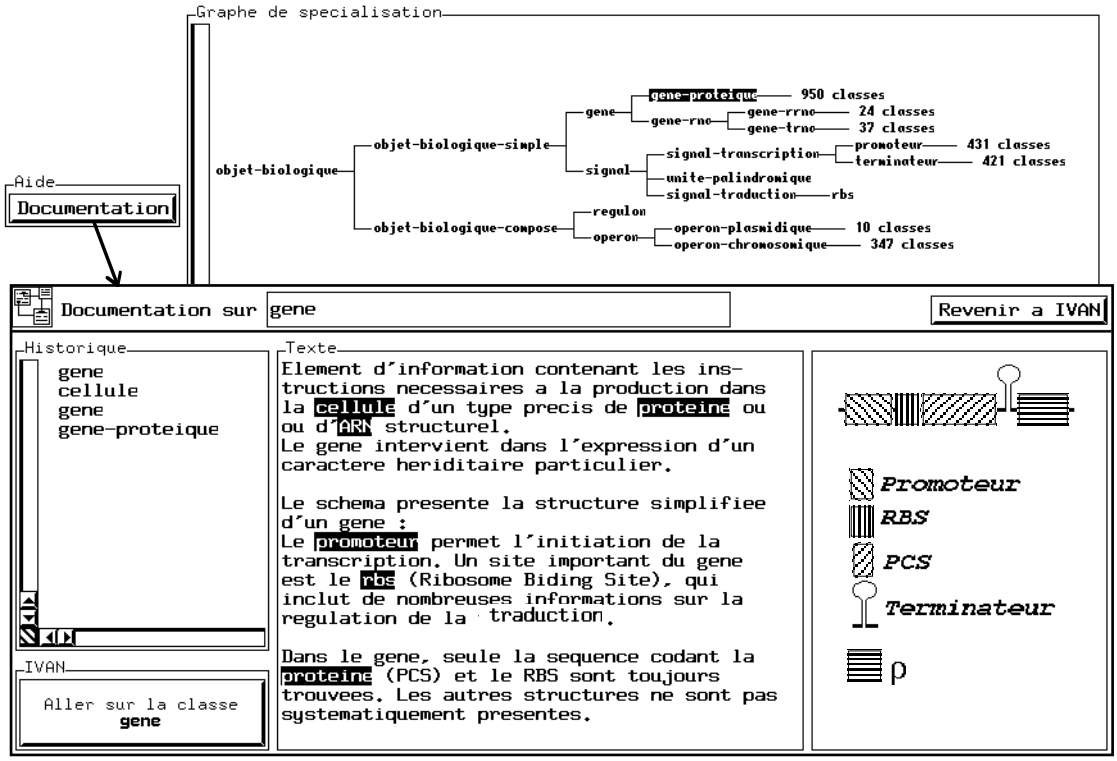

Figure 2 – Navigation dans une base associant des descriptions formelles à un réseau hyper-texte. Il s'agit de la base de connaissances ColiGène sur le génome de la bactérie *Escherichia coli*. À la classe `gene` est associé un nœud de l'hyper-texte qui la commente et l'explique. À partir de ce nœud, d'autres nœuds du réseau hyper-texte sont accessibles. L'utilisateur peut ainsi naviguer alternativement dans l'hyper-texte et dans la base de connaissances.

Conclusion - Base de connaissances et hyper-média

Les relations entre base de connaissances et réseaux hyper-média sont en fait symétriques. D'une part, comme expliqué ci-dessus, les nœuds hyper-média permettent de replacer les connaissances formalisées dans leur contexte d'acquisition et de modélisation. Mais à l'inverse, les connaissances formalisées sont susceptibles de servir d'ossature à une base hyper-média. Les modèles de connaissances à objets permettent en effet le développement de réseaux sémantiques qui offrent d'excellentes capacités de structuration, et, dans une moindre mesure, d'interrogation, qui manquent cruellement aux systèmes multi-média.