

# Proposition d'un méta-modèle basé sur les Topic Map pour la structuration et la recherche d'information

Manuel Zacklad, Jean Caussanel, Jean-Pierre Cahier\*

\* Tech-CICO Laboratoire Technologies de la Coopération pour l'Innovation et le  
Changement Organisationnel, Université de Technologie de Troyes (UTT)  
{Jean-Pierre.Cahier, Jean.Caussanel, Manuel.Zacklad }@utt.fr

## Résumé

Nous présentons ici un état d'avancement de nos travaux dans le domaine du Web Sémantique à la lumière d'une application industrielle. Après avoir replacé notre approche parmi celles du Web Sémantique, nous présenterons HyperTopic, un méta-modèle générique inspiré des Topic Maps pour la réalisation d'applications permettant de structurer et de rechercher un ensemble de ressources présentant des similarités « fonctionnelles » pour des utilisateurs partageant des préoccupations communes.

## 1. Introduction

Dans un précédent travail [CAU 02], nous avons caractérisé deux approches du Web Sémantique : celle du Web Computationnellement Sémantique et celle du Web Cognitivement Sémantique. Dans cette dernière, l'accent n'est pas mis prioritairement sur la problématique d'une sémantique opérationnelle destinée à l'inférence automatique ou à l'exploitation par des agents logiciels, mais sur l'utilisation du Web par des utilisateurs humains engagés dans des activités de navigation. Par ailleurs, dans l'optique du Web Cognitivement Sémantique, les ressources sont considérées comme évoluant rapidement à l'initiative de multiples contributeurs non dotés de compétences poussées en formalisation. Les outils mis à la disposition des contributeurs/concepteurs et des lecteurs/utilisateurs doivent donc permettre des « interactions cognitives » aisées, ce critère étant alors jugé comme prépondérant dans cette approche du WS.

Dans la première phase de notre travail, nous avons identifié le standard des Topic Maps comme étant un des plus adaptés pour la réalisation d'applications répondant aux critères du Web Cognitivement Sémantique. En s'inspirant assez librement de cette norme, nous avons développé un premier prototype permettant la construction et l'exploitation d'applications conformes aux principes du Web Cognitivement Sémantique et proposant des modalités de coopération assez riches entre contributeurs et utilisateurs de l'information. Ce prototype visait à développer des « Places de Marché à Base de Connaissances » (Knowledge-Based Market Place ou KBM) [CAH 02] un dispositif technico-organisationnel en cours de développement

au sein du laboratoire Tech-CICO et pouvant s'appliquer, moyennant des adaptations, aussi bien au contexte de la gestion coopérative de documents en environnement intranet que dans celui d'une place de marché à vocation commerciale en environnement internet.

La réalisation d'une deuxième application dans un contexte industriel plus exigeant (cf. infra) nous a conduit à formaliser d'avantage notre méthode de représentation et ses liens avec la norme Topic Map, sous la forme d'un méta-modèle, HyperTopic. HyperTopic spécialise les TM pour la réalisation d'applications permettant de structurer et de rechercher un ensemble de ressources présentant des similarités « fonctionnelles » pour des utilisateurs partageant des préoccupations communes.

## 2. Contexte de l'application et méthode de conception

La première KBM (Knowledge Based MarkePlace) réalisée au sein du laboratoire Tech-CICO avait permis, outre l'expérimentation du formalisme Topic Maps, le développement d'éléments de méthode permettant de faciliter la structuration de connaissances pour faciliter leur recherche ultérieure par différents types de « clients ». Ces connaissances, les stages de formation, évoluent rapidement et peuvent être abordées selon des points de vue variés, dépendant des finalités des clients (savoir-faire précis dans le contexte d'un projet, cursus qualifiant, culture générale...) [CAH 02].

A partir d'interviews avec des experts du domaine, la méthode permet de définir un certain nombre « d'attributs heuristiques » correspondant à l'organisation des « produits » selon différents points de vue représentés par des hiérarchies de thèmes. Ces hiérarchies de thèmes permettent de guider la navigation de l'utilisateur. D'autres fonctionnalités permettent de mettre en relation les clients, clients et fournisseurs, et fournisseurs entre eux.

Après la réalisation de ce premier prototype, le Laboratoire Tech-CICO a été retenu au printemps 2002 par un groupe industriel du domaine des TIC pour la réalisation, dans le contexte d'un projet de gestion des connaissances, d'une base d'information

permettant la mise en relation de fournisseurs d'information, maîtrise d'ouvrage de projets de système d'information, et de leurs clients, entités métiers utilisatrices ou chargées de la commercialisation des résultats du projet. Destiné à une utilisation en Extranet multilingue, le système devait répondre à des besoins de navigation selon de multiples points de vue, dans une base répertoriant de nombreux livrables de projets relevant d'un domaine de compétences très étendu.

La version initiale du système devait être développée en quelques mois avec une équipe légère et recevoir ensuite des enrichissements continus des « produits » et des classifications de thèmes. Le système devait respecter la terminologie très spécialisée utilisée dans le domaine et l'approche « place de marché » était adaptée pour dynamiser les échanges entre les utilisateurs du système. Ces derniers devaient pouvoir intervenir selon des rôles de « fournisseurs » ou de « clients », le résultat attendu étant une meilleure synchronisation, à la fois au sein de l'offre et entre l'offre et la demande.

Après une série d'entretiens approfondis avec une douzaine d'experts des projets considérés, le réseau sémantique de cette application industrielle a rapidement inclus une abondance de thèmes (près de 2000 pour indexer la première centaine de projets), structurés en plusieurs points de vue métier. Par exemple (cf. aussi infra), pour augmenter ses chances d'intéresser des « clients », un projet doit être indexé :

- selon les technologies mises en œuvre,
- selon les usages potentiels qu'il permet à ses utilisateurs ;
- selon les avantages pour le client de l'entreprise qui achèterait les résultats du projet transformés en produit (qui n'est pas, en l'occurrence, le client de l'application KBM destinée à un usage interne).

La version initiale du système est aujourd'hui opérationnelle et comporte plus d'une centaine de projets, chacun étant indexé par 50 thèmes en moyenne. Le système met en œuvre les fonctions associées aux rôles de client, de fournisseur, de gestionnaire de thèmes et d'administrateur de la base, ainsi qu'aux principales interactions entre ces rôles,

comme la mise en relation du client et du fournisseur ou les demandes d'évolution faites par les fournisseurs aux gestionnaires de thèmes.

### **3. Des Topic Maps au méta-modèle HyperTopic**

Au travers des différentes normes (ISO[BIE 99] et XTM [TMO 01]), les Topic Maps (notées encore TM) spécifient un langage de représentation des connaissances assez proche des réseaux sémantiques. Dans le contexte de la conception d'une application de structuration et de recherche d'information qui nécessite de concevoir un modèle du domaine adapté, il semble pertinent de spécialiser les types de Topic et les types d'associations offertes aux différents contributeurs de manière à éviter une trop grande dispersion dans la représentation du domaine. Dans ce contexte, les TM apparaissent donc comme un méta-modèle à partir duquel seront définies des primitives de représentation adaptées à un domaine et à un type d'activité de navigation particulier. La norme est elle-même muette sur la manière de modéliser un domaine pour atteindre une représentation adaptée.

Nous avons suivi une approche similaire en cherchant à offrir une spécialisation du langage TM qui soit conforme aux principes de représentation des connaissances que nous avons initiée dans l'approche KBM. Notre choix de représentation vise à supporter à la fois la méthode d'ingénierie des connaissances préconisée lors de la conception initiale d'une KBM et la méthode de résolution du problème de recherche d'information sous-jacente aux activités de navigation de l'utilisateur dans les différents points de vue dégagés lors de la phase de conception.

En se voulant générique, notre spécialisation du modèle TM, HyperTopic, apparaît donc elle-même comme un méta-modèle devant guider les utilisateurs souhaitant concevoir une application de structuration et de recherche d'information conforme aux principes du Web Cognitivement Sémantique dont les KBM nous semblent de bons exemples. Selon cette démarche, les TM apparaissent donc comme le méta-modèle d'HyperTopic et deviennent un « méta-méta-modèle » pour les applications finales.

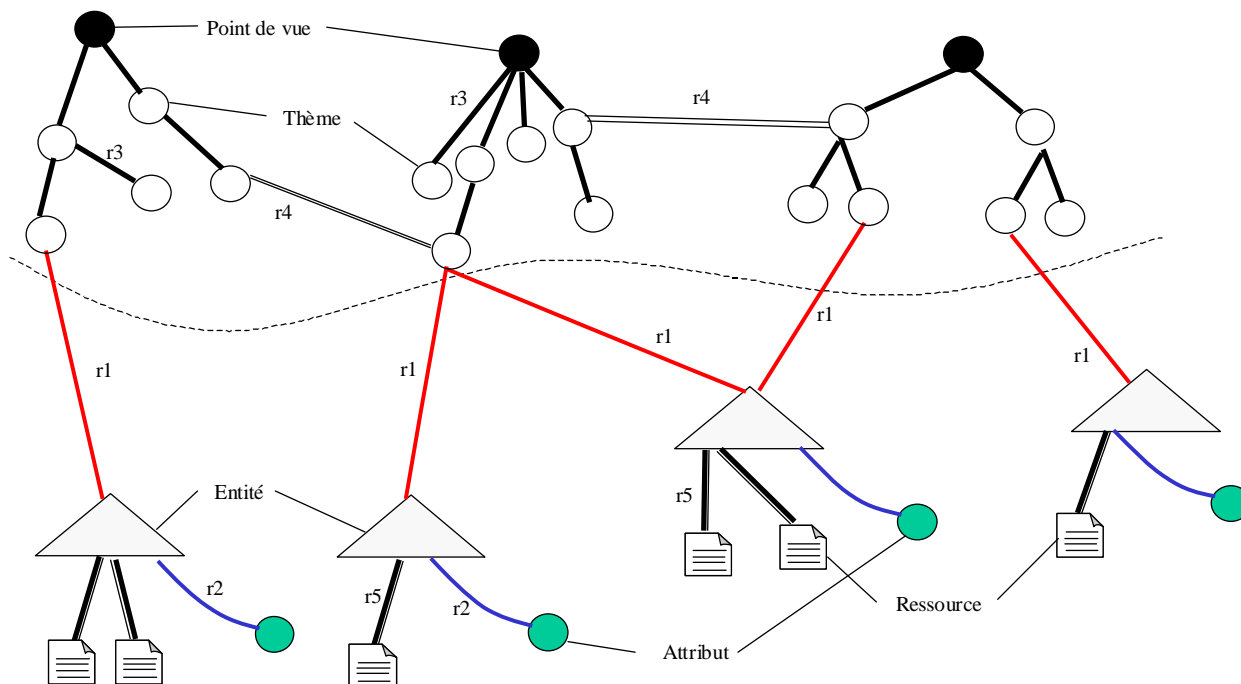


Figure 1 : Un exemple d'instanciation générique du méta-modèle HyperTopic

### 3.2 Les composants d'HyperTopic

Le méta-modèle HyperTopic comprend cinq types de composants, les « points de vue », les « thèmes », les « entités », les « attributs » et les « ressources » et cinq types de relations, la relation entité/thème « traite de » (r1), la relation attribut/entité « est un attribut de » (r2), la relation thème/sous-thème « est un sous-type de thème » (r3), la relation transversale thème/thème « est associé à » (r4) et la relation ressource/entité « est un constituant » (r5). Un sixième type de relation a été utilisé dans l'application, la relation entité/entité « est un sous-type d'entité » dont l'utilité est sans doute moins générale que celle des précédentes.

#### 3.2.1 Entités, attributs, ressources

L'introduction des entités dans HyperTopic correspond à la divergence la plus notable avec l'approche Topic Maps même si, par l'application de fonctions de transformation, il est possible d'exprimer l'entité et ses attributs sous la forme d'associations entre thèmes. En effet, la norme TM possède deux principaux composants, les thèmes et les ressources qui leur sont directement associées. Or, dans la plupart des applications de recherche d'information, la recherche s'applique d'abord à des « objets » possédant une structure générique qui agrège un certain nombre de descripteurs permettant leur caractérisation « primaire » et auquel sont associés une ou plusieurs ressources matérielles porteuses de l'information cible.

C'est le cas, par exemple, dans une application documentaire qui permet de stocker les références

d'ouvrages. L'entité « ouvrage » agrège des informations « primaires » (qui correspondent dans l'approche KBM aux attributs standards) telles que l'auteur, la date de parution, le nombre de pages, la localisation dans les rayons... Dans les applications du WS, les ressources ne sont pas dans des armoires « en dur » mais sont accessibles par des liens URL. Néanmoins, dans les applications qui sont la cible du modèle HyperTopic, ces ressources URL sont rattachées à des entités plus abstraites qui gagnent à être explicitement représentées dans le modèle. Dans une application documentaire un ouvrage possède différentes « instanciations physiques » sous la forme de ses différents exemplaires qui peuvent avoir été édités en un ou plusieurs tomes, par exemple. Sur le WEB, les ressources peuvent être rattachées à des entités logiques : un projet (comme dans le cas de notre application qui gère les URL des livrables), un site WEB, un document « logique » (qui peut être instancié au travers de plusieurs ressources de format distinct), etc.

Dans HyperTopic, les ressources sont donc rattachées aux entités et structurées comme dans la norme TM [TMO 01] sous la forme d'éléments de type « occurrence » et « ressources ». Elles correspondent à des documents accessibles sur le Web dont on ne conserve que l'URI. Les informations « primaires » ou « standards » sont représentées sous forme d'attributs caractérisant les entités. Dans la norme ISO des TM [BIE 99] la notion d'attribut était présente. Elle n'a pas été retenue dans le standard XTM [TMO 01]. HyperTopic ré-introduit cette notion en la restreignant aux « entités ».

### 3.2.2 Les thèmes et les points de vue

#### *Notion de thème*

Proche en cela de la norme des TM, HyperTopic propose d'organiser l'information en permettant d'accéder aux entités à travers un réseau de « thèmes » ou de « sujets ». Selon une approche voisine de celle des réseaux sémantiques, l'interprétation précise de la sémantique du thème dépend pour beaucoup des thèmes environnants. Elle dépend également du point de vue auquel le thème appartient, le « point de vue » étant un composant spécifique d'HyperTopic.

Cette approche de la sémantique d'un thème se différencie d'approches plus structurées dans lesquelles un « concept » est défini soit en intension (par les conditions nécessaires et suffisantes associées à ses propriétés qui permettent de décider si une instance relève bien de ce concept), soit en extension (par l'ensemble de ses instances). HyperTopic visant avant tout des modèles supportant les activités de navigation de l'utilisateur dans l'information, l'approche de type réseau sémantique nous apparaît comme suffisante et plus adaptée aux critères du Web Cognitivement Sémantique.

#### *Les points de vue*

En se différenciant à nouveau de la norme des TM mais aussi des représentations usuelles des réseaux sémantiques, HyperTopic promeut une organisation des thèmes sous la forme de hiérarchies de points de vue. Les points de vue traduisent une vision fonctionnelle de l'entité qui sera pertinente pour les activités ultérieures de recherche d'information, pouvant être dans certains cas, d'authentiques activités de résolution de problème. Le nom du point de vue est donné par le thème placé au sommet de l'arborescence.

Dans l'approche KBM, les points de vue, qui correspondent aux « attributs heuristiques », sont mis en évidence en collaboration avec des experts du domaine qui sont en mesure de suggérer une structuration des thématiques pertinentes par rapport aux finalités des différents clients/utilisateurs de la base. La définition des points de vue et le regroupement ou la recherche de thèmes qui leur sont associés, n'est donc pas une activité neutre et « objective » comme peut l'être la définition des données associées aux attributs des entités. La construction des points de vue est justifiée de façon heuristique par rapport aux activités de recherche d'information et correspond le plus souvent à une approche fonctionnelle de l'entité répondant aux besoins des utilisateurs.

Les thèmes regroupés en point de vue sont organisés de façon hiérarchique à travers une relation « sous-type de ». Cette organisation hiérarchique

permet dans certains cas de proposer des ontologies du domaine (ontologies construites selon des critères heuristiques). Cette organisation hiérarchique a pour principale finalité de faciliter la navigation à l'intérieur d'un point de vue donné en utilisant la relation de généralisation portée par la relation « sous-type de ». En naviguant dans l'arborescence l'utilisateur se déplace des thèmes les plus généraux aux plus spécifiques.

Une question méthodologique importante est celle du caractère exclusif ou non du rattachement de l'entité à des thèmes « frères » (si une entité est rattachée à un thème peut-elle être aussi rattachée à son frère dans un point de vue donné). Le débat n'est pas tranché aujourd'hui. Le rattachement d'une entité à un thème via la relation « traite-de » peut signifier que l'entité est caractérisée exclusivement par ce thème ou que l'entité « traite de ce thème parmi d'autres voisins », ce qui signifie qu'il n'y a pas nécessairement de rattachement exclusif à l'un des thèmes d'un point de vue.

Ces deux interprétations ont bien sûr des conséquences sur les activités de navigation de l'utilisateur au sein des hiérarchies. Dans le cas de l'exclusivité, la navigation correspond au parcours dans un arbre de décision, la sélection d'une branche conduisant à exclure certaines entités au profit d'autres. Dans l'autre cas, la navigation correspond à la recherche « de traits sémantiques » composants l'entité. La sélection d'une branche signifie que l'on souhaite « rajouter des traits sémantiques » à l'entité cible, sans que cet ajout soit nécessairement contradictoire avec l'ajout de traits distinctifs possédés par les frères du thème.

Selon cette dernière interprétation, si à l'intérieur d'une hiérarchie de thèmes (correspondant à un point de vue) la présence de différents fils d'un thème correspond bien à la possession de traits sémantiques exclusifs par ces fils, la possession par l'entité de ces traits sémantiques exclusifs n'est pas nécessairement contradictoire en tant que caractérisation des propriétés effectives de l'entité. Par exemple, si dans la recherche d'une automobile, les traits sémantiques fonctionnel « tout-terrain » et « routière » s'opposent bien sémantiquement, ils se rencontrent bien « matériellement » de façon conjointe dans certains modèles de véhicule<sup>1</sup>.

---

<sup>1</sup> Différentes interprétations de cette apparente contradiction pourraient être fournies : si une voiture ne peut pas exceller à la fois dans deux caractéristiques en partie exclusive elle peut représenter un compromis satisfaisant, ces deux caractéristiques pourraient exprimer des spécifications pour une entité à concevoir, l'entité peut présenter deux caractéristiques opposées mais pas au même moment, certaines caractéristiques peuvent s'appliquer à certaines parties de l'entité mais pas à d'autres, etc.

La figure 2 fournit un exemple partiel de l'application du modèle HyperTopic à l'application KBM dans le domaine des projets Système d'Information. Seuls trois points de vue sont

représentés sur la dizaine effectivement utilisés qui organisent près de deux mille thèmes.

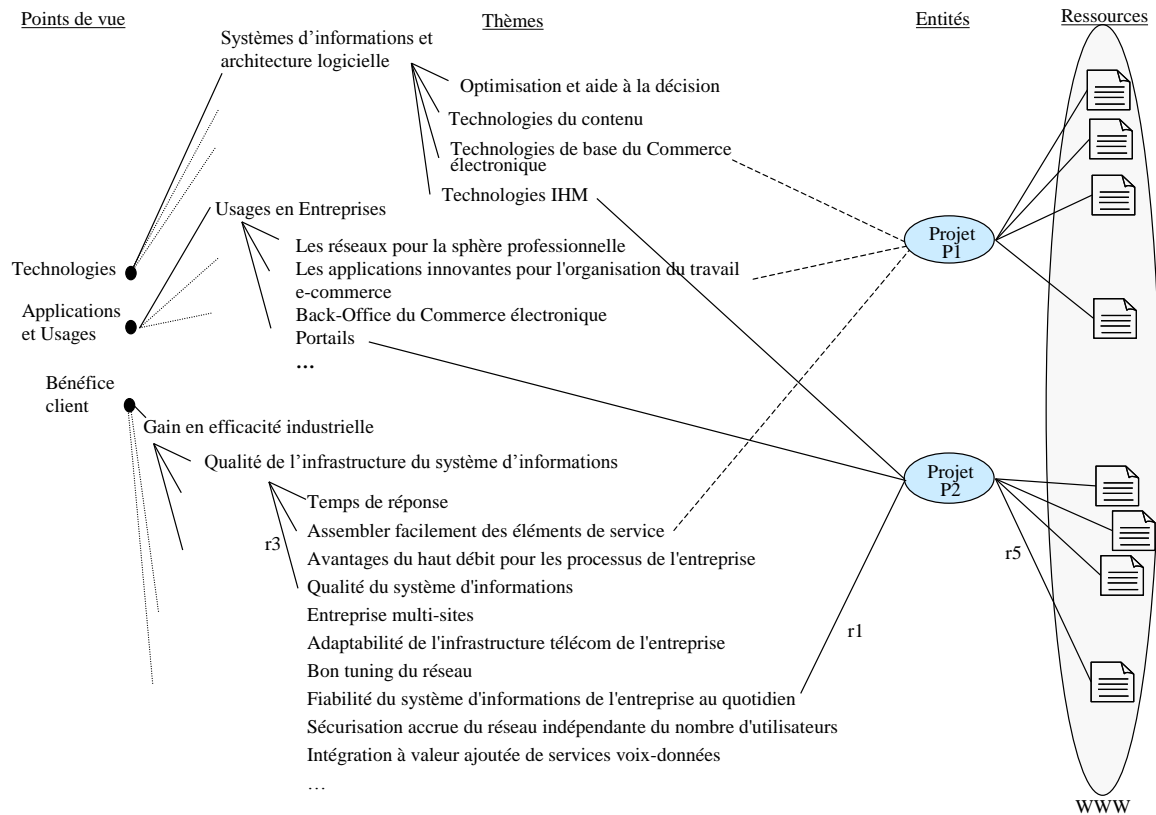


Figure 2 : Un extrait de la représentation HyperTopic de la KBM « projets du SI »

#### 4. Conclusion

Le méta-modèle HyperTopic a donné lieu à une opérationnalisation dans un prototype informatique. Celui-ci peut être considéré comme un générateur de KBM, une classe d'applications relevant du Web Cognitivement Sémantique, puisque visant à faciliter tant les activités de structuration de l'information des fournisseurs/concepteurs que les activités de navigation des clients/utilisateurs. Ce prototype va nous permettre de poursuivre nos recherches dans les deux directions.

En ce qui concerne la première, nous souhaitons approfondir la démarche d'ingénierie des connaissances proposée dans l'approche multipoints de vue. Il nous faut pour cela clarifier encore la nature des relations entre les thèmes et entre thèmes et entités. La nature de la relation d'association transverse entre les thèmes, par exemple, que nous n'avons pas discutée ici (r4), mérite d'être précisée. Cet approfondissement pourra ensuite se traduire par l'ajout de nouvelles contraintes dans le méta-modèle HyperTopic, qui permettront d'améliorer l'assistance à la structuration des connaissances.

En ce qui concerne la seconde, nous cherchons actuellement à mieux caractériser les inférences effectuées par les utilisateurs dans leurs activités de navigation. Ceci implique d'une part, de bien caractériser la nature des problèmes qu'ils rencontrent et, d'autre part, de caractériser le type d'aide fourni par le modèle, via l'interface, dans l'exploration progressive du réseau sémantique.

#### Références

- [BIE 99] Biezunski M, Martin Bryan, Steven R. Newcomb, ISO/IEC FCD 13250:1999 - Topic Maps, 19 April 1999.
- [CAU 02] Caussanel J., Cahier J.-P., Zacklad M., Charlet J. "Les TopicMaps sont-ils un bon candidat pour l'ingénierie du Web Sémantique ?", *Actes conférence IC'2002*, Rouen, 28-30 Mai 2002.
- [CAH 02] Cahier J.-P., Zacklad M., "Towards a Knowledge-Based Marketplace model (KBM) for cooperation between agents", *Actes conférence COOP'2002*, St Raphael, 4 - 7 June 2002, IOS Press (à paraître).
- [TMO 01] XML Topic Maps (XTM) 1.0, TopicMaps.Org Specification, 2001. <http://www.topicmaps.org/xtm/1.0>

