# Construction, Évolution et Visualisation de Topic Maps contextualisées

Lydia Nadia Khelifa

▶ **To cite this version:**

Lydia Nadia Khelifa. Construction, Évolution et Visualisation de Topic Maps contextualisées. Recherche d'information [cs.IR]. Conservatoire national des arts et metiers - CNAM, 2014. Français. NNT : 2014CNAM0983 . tel-01284488

**HAL Id: tel-01284488**
**https://theses.hal.science/tel-01284488**

Submitted on 7 Mar 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

le cnam

*ÉCOLE DOCTORALE INFORMATIQUE, TÉLÉCOMMUNICATIONS ET ÉLECTRONIQUE (PARIS)*

**ÉQUIPES ISID – LABORATOIRE CEDRIC**

# THÈSE présentée par :


## Lydia Nadia KHELIFA


soutenue le : 18 décembre 2014

Pour obtenir le grade de : **Docteur du Conservatoire National des Arts et Métiers**

Discipline/ Spécialité : Informatique

# Construction, Évolution et Visualisation de Topic Maps contextualisées

**THÈSE dirigée par :**

| | |
|---|---|
| **PR. Jacky Akoka** | Professeur, Cnam |
| **DR. Nadira Lammari** | Maître de Conférence-HDR, Cnam |

**RAPPORTEURS :**

| | |
|---|---|
| **PR. Sylvie Desprès** | Professeur, Université Paris 13, France |
| **PR. Jean-Marie Pinon** | Professeur émérite, INSA de Lyon, France |

---

**JURY :**

| | |
|---|---|
| **PR. Jacky Akoka** | Professeur, Cnam, France |
| **PR. Maurice Aymard** | Professeur, FMSH, France |
| **PR. Sylvie Desprès** | Professeur, Université Paris 13, France |
| **DR. Nadira Lammari** | Maître de Conférence-HDR, Cnam, France |
| **PR. Elisabeth Métais** | Professeur, Cnam, France |
| **PR. Christophe Nicolle** | Professeur, Université de Bourgogne, France |
| **PR. Jean-Marie Pinon** | Professeur Émérite, INSA de Lyon, France |

# Remerciements

Je remercie mes directeurs de thèse, Dr. Nadira LAMMARI et Pr. Jacky AKOKA pour m'avoir soutenu et conseillé tout au long de cette thèse et jusqu'au dernier moment. Je les remercie de m'avoir offert l'opportunité d'intégrer une équipe de recherche et de m'avoir initié au métier d'enseignant chercheur. Mes remerciements s'adressent plus particulièrement à Nadira, je la remercie pour sa présence, sa disponibilité, ses encouragements et ses conseils qui ont de loin dépassé le cadre de la thèse. C'est grâce à ses conseils et orientations que j'ai pu réaliser ce travail.

Je remercie sincèrement Pr. Jean-Marie PINON et Pr. Sylvie DESPRES d'avoir porté autant d'intérêt à ce travail en acceptant d'être les rapporteurs de ma thèse. Je remercie également Pr. Maurice AYMARD, Pr. Christophe NICOLLE et Pr. Elisabeth METAIS d'avoir eu l'amabilité de faire partie de mon jury de thèse.

Je remercie la Fondation Maison des Sciences de l'homme (FSMH) d'avoir financé ma thèse, et Pr. AYMARD pour m'avoir aider à mieux comprendre le domaine des sciences humaines et sociales. Je remercie également, Mr. Hammou FADILI pour sa collaboration.

Je tiens à remercier Pr. Habiba DRIAS qui m'a encouragé à entreprendre cette thèse.

Ensuite, mes remerciements s'adressent à tous les membres de l'équipe ISID pour leur soutien et leur encouragement pendant ma thèse et pour les sympathiques repas d'équipes organisés régulièrement. Particulièrement, je remercie mes amis les doctorants avec qui j'ai partagé les moments les plus durs et les plus beaux de ma

thèse : Amina, Anh, Feten, Houda, Nabil, Ryadh, Sarah, Odette, Nora et Quentin ; et les ingénieurs Pascal et Rodney. Un très grand merci est à mon amie Zeineb, qui a été présente pour moi, je tiens vraiment à la remercier pour sa gentillesse, sa sincérité et son amitié. Je remercie également mes amies de longue date Nadia, Ismahane, Lamia, Djalila et Amel qui m'ont toujours soutenue.

Mes pensées et mes remerciements les plus sincères, et sans qui je n'aurai jamais pu aller jusqu'au bout de cette thèse, à mes chers parents, mon mari, mes sœurs et mon frère.

# Table des matières

# Chapitre 1

# Introduction générale

Les technologies du Web ont produit un contexte qui a changé le rapport des organisations à l'information. Ce changement s'est matérialisé par un besoin d'organisation, de stockage et d'accessibilité à des ressources informationnelles qui sont le plus souvent hétérogènes, disponibles dans des contenus publics ou privés.

L'organisation des ressources informationnelles nécessite, la plupart du temps, une modélisation des connaissances les contenant. Ces connaissances sont, très souvent, en relation avec le métier de l'organisation. Elles constituent les ressources sémantiques du Système d'Information de l'organisation. Ces dernières peuvent être classées dans l'une des catégories suivantes : (a) celles servant à modéliser le ou les domaines couverts par les ressources informationnelles et (b) celles servant à la description du contenu des ressources informationnelles.

La première catégorie regroupe pour l'essentiel les dictionnaires spécialisés, les classifications, les taxonomies, les thésaurus, et les ontologies de domaine. Certaines d'entre elles sont régies par des normes ou standards nationaux et/ou internationaux. Un dictionnaire spécialisé réuni, dans un ordre donné, des termes couramment utilisés dans un domaine avec leur signification. Les classifications, les taxonomies ainsi que les thésaurus trouvent leur origine dans l'ingénierie documentaire. Ceux sont des vocabulaires contrôlés fondés sur une structuration hiérarchisée des concepts d'un

domaine de spécialité. Les thésaurus disposent, en plus des concepts et des liens hiérarchiques entre concepts, d'un ensemble prédéfini de relations sémantiques entre les concepts. Les ontologies de domaine sont des formes élaborées de vocabulaires contrôlés. Elles sont issues de l'évolution du web vers le web sémantique. Elles représentent les concepts du domaine sous forme d'un réseau sémantique et dispose de concepts favorisant le raisonnement et donc la déduction de connaissances complémentaires.

Parmi les ressources sémantiques de la seconde catégorie, on peut citer les métadonnées descriptives, les Topic Maps et les folksonomies. Une métadonnée descriptive est une donnée que l'on assigne à une ressource informationnelle. Cette donnée correspond à une valeur d'une propriété composant, la plupart du temps, un vocabulaire de description de ressources prédéfini. Une folksonomie est un agglomérat de mots-clés librement attribués à une ressource informationnelle par une communauté d'utilisateurs. Une Topic Map est un réseau sémantique regroupant un ensemble de sujets permettant de décrire un ensemble de ressources informationnelles selon un ou plusieurs points de vue. Les liens existants entre les sujets et entre un sujet et les ressources qu'il décrit, favorisent la navigation au sein du contenu.

La prolifération de ressources informationnelles devenant de plus en plus multilingues, pluridisciplinaires et disposant d'une couverture spatio-temporelle de plus en plus large ou parfois même restreinte, soulève le problème de définition de ressources sémantiques adaptées. En effet, les termes et concepts peuvent :

– subir des transformations de sens dues à des causes diverses telles que les innovations technologiques et scientifiques ou encore le développement de la vie sociale, économique et culturelle [Dury, 2013] (c'est le cas, par exemple, des termes « chômer [1] » et « usine [2] »),

1. « Chômer » qui initialement voulait dire « ne pas travailler pendant la chaleur » a pris son sens actuel avec le développement de la société capitaliste et avec l'apparition des « sans travail ».

2. Selon [Dury, 2013], le terme « usine » avait pour sens, au 18ème siècle, un établissement où l'on travaillait le fer à l'aide de machines. Au 19ème siècle, le mot prend son sens actuel c'est-à-dire « tout établissement pourvu de machines modernes ».

- être empruntés à d'autres disciplines connexes ou pas [Dury, 2000] (c'est le cas, par exemple des termes « biosphère[3] » et « biocénose[4] »),
- avoir un sens variant selon le lieu géographique d'usage (c'est le cas du terme « entrepreneur »),
- accumuler des sens différents au fil de leur évolution et disparaître (c'est le cas par exemple du terme « manufacture[5] »).

Cette variabilité du sens des termes nous oblige, pour des raisons d'efficacité dans la recherche d'information et de précision dans la traduction, à privilégier des ressources sémantiques modélisant ce changement de sens. C'est dans ce cadre bien précis que s'inscrit la problématique de cette thèse. Elle a découlé du programme FSP-Maghreb. Ce programme a été initié par le FMSH (Fondation Maison des Sciences de l'Homme). Il vise à développer les échanges entre chercheurs (en Science Humaine et Sociales) maghrébins et leurs partenaires français et à mettre en commun un ensemble de savoirs sur les deux cultures et les deux sociétés. De cet objectif ont découlé deux besoins. Le premier est celui de la conception d'une structure pour un wiktionnaire[6] multilingue et multiculturel dédié aux SHS ainsi que la conception et le développement d'une application pour son peuplement. Le second besoin est la

3. Selon [Dury, 2000], le terme « biosphère » a d'abord été utilisé en biologie pour désigner un atome globuleux d'une existence hypothétique et qui serait la base unique de tous les corps organisés. Il a ensuite été utilisé au moyen d'un transfert sémantique dans le domaine de la biogéographie pour désigner l'une des couches géochimiques de la sphère terrestre, constituée par la masse organique des êtres vivants. Enfin, au cours d'un second transfert, l'écologie a repris biosphère à son compte et l'utilise aujourd'hui pour décrire la région de la planète qui renferme l'ensemble des êtres vivants et dans laquelle la vie est possible en permanence.

4. Selon [Dury, 2006], le terme « biocénose » est un terme apparu en zoologie (crée en 1877 par le zoologiste allemand Karl Möbius) pour ensuite passer dans le domaine de la biologie végétale (début du 20ème siècle) et être finalement adopté par l'écologie. Aujourd'hui, il désigne l'ensemble des êtres vivants (micro-organismes , plantes et animaux) qui peuplent un biotope.

5. Selon [Dury, 2013], le terme « manufacture » apparaît au 16ème siècle. Au 17ème siècle il commence à désigner un établissement industriel. De nos jours, il sort de l'usage.

6. Dictionnaire électronique implémenté avec la technologie du wiki sémantique

mise en œuvre d'une structure adéquate pour l'organisation d'un contenu SHS ainsi qu'une mise en œuvre d'une approche pour son peuplement.

Les SHS rassemblant plusieurs champs disciplinaires hétérogènes, le problème de la variation des sens des termes dans les contenus était à prendre en compte. Ceci nous a amené à définir la problématique de la thèse. Il s'agit de prendre en compte la variation du sens des termes et des concepts dans les ressources sémantiques décrivant un domaine couvrant plusieurs disciplines ainsi que dans des ressources sémantiques permettant l'organisation de contenus pluridisciplinaires et multilingues. Cependant, compte tenu du temps imparti et de la diversité des modèles de ressources sémantiques disponibles, nous nous sommes focalisés dans cette thèse sur le modèle du dictionnaire, conformément au cahier des charges du projet, et sur le modèle de Topic Maps pour les avantages qu'il offre en terme de recherche d'information.

## 1.1 Résumé des contributions de la thèse

Les contributions de la thèse peuvent être résumées en les points suivants :

1. La conception d'une structure générique d'un dictionnaire électronique multilingue et pluridisciplinaire conforme à la norme ISO 1951 [ISO1951, 2007]. Cette structure a été personnalisée pour les SHS mais est utilisable pour tous les autres domaines.
2. La conception et la mise en oeuvre d'une application wiki pour le peuplement, la mise à jour et la consultation de ce dictionnaire.
3. La transposition de cette application dans un contexte distribué afin de permettre à des chercheurs de l'utiliser quel que soit leurs lieux géographiques.
4. Une nouvelle extension du modèle standard de Topic Maps, nommée Modèle de Topic Maps contextualisées, permettant d'indexer de façon précise des ressources se trouvant dans des contenus à la fois multilingues et pluridisciplinaires caractérisées par la présence de termes ou concepts dont le sens peut

varier d'une ressource à une autre. Notre proposition est motivée par le fait que ni le modèle standard [ISO/IEC, 2008], ni les extensions proposées dans la littérature, ne prennent en charge la variation du sens des termes et concepts.

5. Une famille d'approches, dirigées par les modèles, permettant de construire une Topic Map conforme à l'extension de modèle proposé. Cette famille regroupe trois approches de construction : construction à partir de contenus hétérogènes, par fusion de Topic Maps contextualisées et par ré-ingénierie. Chacune d'elle est un agencement de composants méthodes.

6. Une famille d'approches permettant de faire évoluer une Topic Map contextualisée. Cette famille rassemble trois approches d'enrichissement dont l'enrichissement par intégration de ressources. Comme pour les approches citées ci-dessus, ces approches se construisent par assemblage de composants méthodes et sont toutes dirigées par les modèles.

7. La proposition de deux mécanismes, l'un délivrant une vue personnalisée d'une Topic Map, l'autre proposant une vue synthétique d'une Topic Map ou d'une vue personnalisée. Le premier mécanisme permet à un utilisateur d'effectuer une recherche ciblée. Le second lui offre la possibilité d'obtenir dans un premier temps une vue globale de la Topic Map ou de la vue personnalisée, puis une vue de plus en plus détaillée.

8. Une plateforme regroupant tous les mécanismes de visualisation proposés ainsi que la plupart des approches conçues. Le développement de cette plateforme suit les principes de MDA (Model Driven Architecture).

## 1.2 Organisation du document

La suite de ce document est organisée en deux grandes parties.

La première partie est structurée en deux chapitres. Le premier chapitre (chapitre 2) rappelle la structure des différentes ressources sémantiques les plus connues et

utilisées. Il présente notre analyse de ces modèles qui nous a permis de les positionner par rapport à la problématique de la thèse. Le second chapitre (chapitre 3) présente et compare les modèles de Topic Maps existants. Il dresse aussi un état des lieux de la recherche dans le domaine de la construction et de l'évolution de Topic Maps. Les différentes approches sont comparées en se basant sur des critères d'évaluation que nous avons définis. Ces comparaisons nous ont permis de dégager les limites des modèles et approches existants et d'introduire nos contributions.

La seconde partie est dédiée à la description de nos contributions. Elle est composée de quatre chapitres. Le chapitre 4 réunit nos résultats de recherche sur la conception de la structure du wiktionnaire des SHS ainsi que sur la conception et la mise en œuvre de l'application permettant de le manipuler. La description du modèle de Topic Maps proposé, en l'occurrence le modèle de Topic Maps contextualisées ainsi que la description de la famille d'approches pour la construction et l'évolution des Topics Maps contextualisées, font l'objet du chapitre 5. Le chapitre 6 concerne les aspects visualisation de Topic Maps contextualisées. Il présente les deux mécanismes cités dans nos contributions. Le chapitre 7 décrit le prototype relatif à la plateforme de construction, d'évolution et de visualisation de Topic Maps contextualisées.

La conclusion constitue le chapitre 8 de cette thèse. Elle résume nos contributions et propose un ensemble de perspectives.

# Chapitre 2

# Panorama des ressources sémantiques

## 2.1 Introduction

La société d'aujourd'hui produit et consomme une énorme quantité d'informations. Le besoin de partager des contenus informationnels pertinents a connu, de ce fait, une croissance rapide. La satisfaction de ce besoin passe inévitablement par une organisation des contenus reposant sur des ressources sémantiques, qu'il faut non seulement créer, mais aussi gérer.

Les travaux de recherche relatifs à la création et à la gestion des ressources sémantiques ne datent pas d'aujourd'hui. Ils ont fait partie, à l'origine, de l'ingénierie documentaire avant de prendre une grande place dans l'ingénierie des connaissances. Ils ont concerné tout un panel de ressources sémantiques. Certains ont trait aux ressources sémantiques en tant que modèles de représentation des connaissances d'un domaine donné. D'autres se focalisent plutôt sur les ressources sémantiques en tant que modèles de description de contenus des ressources informationnelles. Les deux prochaines parties de ce chapitre sont consacrées respectivement à ces deux catégories de ressources. Elles dressent un panorama non exhaustif des ressources sémantiques

les plus connues. La synthèse générale de ce chapitre sert de justification des modèles de ressources sémantiques étudiés dans cette thèse.

## 2.2 Ressources sémantiques pour la description de domaines

Les ressources sémantiques pour la description de domaines tentent de modéliser la structure sémantique sous-jacente d'un ou plusieurs domaines. Elles disposent de concepts ou de termes associés à un domaine et imposent une structure permettant de lier les concepts ou termes via des relations sémantiques. Il y a plusieurs façons de catégoriser ce type de ressources sémantiques. Nous avons opté pour une catégorisation orientée structure. Selon ce critère, nous les regroupons en quatre grandes catégories : les classifications et les taxonomies, les thésaurus, les ontologies de domaine et enfin les dictionnaires spécialisés. Chacune de ces catégories est décrite dans les paragraphes qui suivent.

### 2.2.1 Les classifications et les taxonomies

Les classifications structurent hiérarchiquement les connaissances d'un ou plusieurs domaines [Weller, 2010]. Pour capturer le ou les domaines concernés, on s'efforce dans une classification à respecter deux principes fondamentaux : le regroupement et le classement [Hudon and Hadi, 2011]. Le regroupement consiste à rassembler les connaissances qui possèdent au moins une caractéristique commune qu'on appelle parfois la caractéristique de division. Le classement consiste en la mise en séquence des groupes ainsi constitués. On aboutit ainsi à des classes hiérarchisées codées le plus souvent. Certaines codifications reflètent la position de la classe dans la hiérarchie. À titre d'exemple, dans la CIB (Classification International des Brevets [1]), la classe « Fermetures à lacets » est codée A43C 1/00. Elle fait partie de la

1. http ://www.wipo.int/classifications/ipc/fr/

classe « Attaches ou accessoires pour chaussures » de code A43C.

Il existe plusieurs formes de classifications dont les classifications décimales, les classifications à facettes et les taxonomies.

Les classifications décimales sont fondées sur le principe que toute classe est divisée en dix sous-classes. Une des plus connue est la CDU (Classification Décimale Universelle [2]). Cette dernière regroupe plusieurs domaines dont l'informatique et les sciences sociales et est publiée dans au moins de 40 langues.

Les classifications à facettes regroupent, en réalité, plusieurs classifications, nommées facettes et cela au sein d'un domaine [Hudon and Hadi, 2011]. Par conséquent, elles fournissent plusieurs systèmes hiérarchiques au sein d'un même domaine. Notons cependant que les facettes n'ont pas besoin d'être ordonnées, ni d'être de même type, mais elles doivent être clairement définies et doivent être mutuellement exclusives [Herring, 2007].

Les taxonomies peuvent être vues comme des formes de classifications simples dans le sens où elles n'obligent pas de disposer de codes pour les termes. Cependant, elles sont porteuses de plus de sémantique [Lambe, 2014] car elles fournissent un vocabulaire contrôlé et les liens entre termes sont des liens hiérarchiques du type « fait partie de », « est sous-catégorie de », etc.

### 2.2.2 Le thésaurus

Contrairement à la plupart des schémas de classification cités ci-dessus, les thésaurus adhérent obligatoirement à la norme internationale ISO 25964-1 :2011 [3] [Afnor, 2013]. Cette dernière fournit des directives et des recommandations pour la conception et la gestion des thésaurus monolingues et multilingues tout en assurant leur interopérabilité avec d'autres vocabulaires. Elle offre aussi une structure com-

2. http ://www.udcc.org/

3. Cette norme annule les deux anciennes normes internationales ISO 2786 :1986 (thésaurus monolingue) et ISO 5964 :1985 (thésaurus multilingue), ainsi que les normes françaises NF Z47-100 :1981 (thésaurus monolingue) et NF Z47-101 :1990 (thésaurus multilingue.)

patible avec le format SKOS [4]. L'objectif de SKOS est de présenter les données sous un format qui convient à une utilisation dans le cadre du web sémantique.

Les thésaurus structurent un ou plusieurs domaines à l'aide d'une collection de termes et de relations entre ces termes. La figure 2.1 fournit un extrait de son méta-modèle décrit dans [Afnor, 2013].

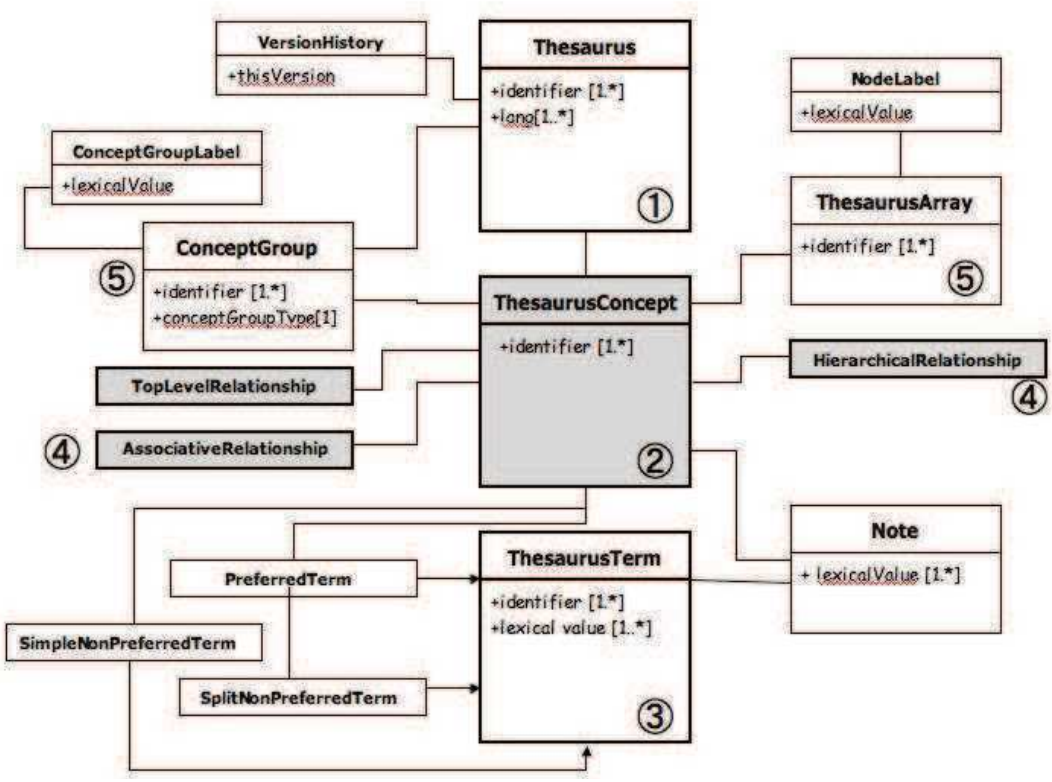


FIGURE 2.1 – Un extrait du méta-modèle du thésaurus [Afnor, 2013]

Un terme dans un thésaurus est un label associé à un concept du domaine. Il fait partie d'un vocabulaire contrôlé et peut être constitué d'un ou de plusieurs mots. À un concept, sont associés un ou plusieurs termes de langues différentes. Ces termes peuvent être regroupés par classification ou par facette. Un terme d'une langue peut

4. SKOS (Simple Knowledge Organization System) est un système d'organisation des connaissances recommandé par le W3C depuis le 18 août 2009. Il est fondé sur RDF.

disposer de propriétés telle que la propriété « Note explicative » (NE) (« Scope Note » ou SN en anglais) utilisée pour clarifier le périmètre sémantique du terme.

La norme met aussi à la disposition d'un concepteur de thésaurus une variété de relations sémantiques permettant de lier des concepts entre eux, et par conséquent, les termes qui les représentent. Ces relations sont l'équivalence intralinguistique et interlinguistique, l'équivalence composée, la hiérarchie et l'association [Hudon and Hadi, 2011].

La relation d'équivalence est symbolisée par le terme « Employé Pour » (EM) ou réciproquement « Employer » (EM). L'équivalence inter-linguistique est établie entre termes de langues naturelles distinctes mais représentant le même concept. Elle peut être totale ou partielle. L'équivalence intralinguistique sert à lier des termes de même langue associés à un même concept. L'équivalence composée, notée « EM+ » est utilisée pour permettre la construction de terme (donc de concept) à partir de combinaisons de termes (tel que « logiciel de gestion de thésaurus » qui se construit par combinaison de « logiciel de gestion » et « thésaurus » [Will, 2012]).

La relation hiérarchique permet de clarifier les relations entre les concepts en termes de supériorité ou de subordination. Elle est symbolisée par le terme « TermeGénérique » (TG) ou réciproquement par le terme « TermeSpécifique » (TS). Contrairement aux normes précédentes, il est possible dans la norme actuelle de distinguer :

- la relation hiérarchique dite « générique » pour exprimer un lien sémantique « Est un » (utilisation dans ce cas du sigle TGG/TSG),
- de la relation hiérarchique d'instance pour préciser que le lien en question est un lien sémantique de type « instance de » (utilisation dans ce cas du sigle (TGI/TSI)),
- de la relation hiérarchique partitive pour mentionner un lien sémantique « fait partie de » (utilisation dans ce cas du sigle TGP/TSP).

De plus, la norme actuelle n'interdit pas l'existence de relations poly-hiérarchiques

(un terme disposant de plusieurs relations hiérarchiques).

La relation d'association symbolisée par « Terme Associé » est une relation autre que hiérarchique ou d'équivalence. Parmi les relations, fréquemment utilisées, citons la relation de causalité, de composition, etc. Comme pour la relation hiérarchique, il est possible de spécifier la nature de la relation (cause/effet, processus/produit, personne/discipline, etc.)

### 2.2.3 Les ontologies de domaine

Les ontologies dont l'étymologie nous renvoie au mot grec ontologia, qui signifie « parler » (logia) de « l'être » (onto), est une branche de la philosophie qui décrit la science de l'être. Le philosophe grec Platon (427-347 AJC) était l'un des premiers philosophes à s'intéresser à la représentation du monde et à l'abstraction des entités dont on parle. Cette représentation est l'idée fondamentale de l'ontologie. Quant à Aristote (384-322 AJC), il a introduit la notion de concept et de taxonomie entre ces concepts. Il a également introduit la notion de sous-concept/super-concept, pour distinguer les genres et les classifier formellement. C'est sur ce principe que se base la notion moderne de concept d'ontologie et de l'héritage entre concepts. Aristote a également introduit un certain nombre d'inférences appelées « syllogismes [5] » qui ont été utilisées dans la logique moderne des systèmes de raisonnement [Sowa, 2000].

En informatique, l'ontologie a été définie par Gruber comme étant « une spécification explicite et formelle d'une conceptualisation partagée » [Gruber, 1993]. [Studer et al., 1998] a précisé les termes qui constituent cette définition, à savoir :

- « explicite » qui veut dire que les concepts de l'ontologie sont explicitement définis,
- « formelle » pour mentionner que l'ontologie est représentée dans un formalisme exploitable par les machines,
- « conceptualisation » pour référer à un modèle d'abstraction

5. Syllogisme. http ://fr.wikipedia.org/wiki/Syllogisme (Consulté le 18/09/2012).

- et « partagée » pour préciser qu'une ontologie n'est pas propre à un seul individu mais validée par un groupe.

Il existe plusieurs classifications des ontologies [Psyché et al., 2003]. Nous nous intéressons dans le cadre de cette thèse à la classification des ontologies en tant qu'objet de conceptualisation. Dans ce type de classification, nous retrouvons les catégories d'ontologies suivantes :

- Les ontologies pour la représentation des connaissances qui rassemblent des primitives de représentation d'ontologies (exemple l'ontologie de FRAME qui intègre les primitives de représentation des langages à base de frames : classes, instances, facettes, propriétés/slots, relations, restrictions, valeurs permises, etc. ),
- Les ontologies de haut niveau qui visent à étudier les catégories des choses qui existent dans le monde (exemple l'ontologie « Upper Cyc » qui contient les concepts les plus généraux sur les connaissances humaines),
- Les ontologies génériques qui s'efforcent de décrire des concepts génériques moins abstraits que ceux décrits par des ontologies de haut niveau (exemple l'ontologie Wordnet),
- Les ontologies de domaine qui rassemblent les concepts du domaine (exemple l'ontologie du domaine de l'énergie éolienne [Küçü and Arslan, 2014]),
- Les ontologies de tâche qui se focalisent sur la description d'une activité ou tâche générique (exemple l'ontologie décrite dans [Lammari et al., 2003]),
- Et enfin les ontologies d'applications qui sont les plus spécifiques et qui regroupent le plus souvent des concepts correspondants aux rôles joués par les entités du domaine tout en exécutant une certaine activité.

Nous nous intéressons dans le cadre de cette thèse aux ontologies de domaine. Ces dernières sont avant tout des réseaux sémantiques permettant de modéliser un domaine d'étude. Les connaissances du domaine sont explicitées via les éléments suivants : les concepts, les relations, les axiomes et les instances.

Les relations traduisent les associations entre concepts. La relation de généralisation/spécialisation constitue le pilier de la stucture. En d'autres termes, on ne peut pas parler d'ontologies sans parler de liens hiérarchiques entre concepts. D'autres types de relations peuvent être ajoutés pour compléter la description du domaine. Une discussion approfondie des types de liens est présentée dans [Storey et al., 1994].

Les axiomes sont des assertions sensées fournir une aide dans le raisonnement.

Les instances constituent la définition extensionnelle de l'ontologie.

### 2.2.4 Les dictionnaires spécialisés

Les dictionnaires spécialisés sont avant tout des dictionnaires, c'est-à-dire un recueil de lemmes d'une langue réunis selon une nomenclature présentée généralement par ordre alphabétique. Ils fournissent pour chaque mot un certain nombre d'informations relatives à son sens et à son emploi [CNRTL, 2012]. Ils sont dits spécialisés parce qu'ils traitent soit d'un domaine de la connaissance (telles que la philosophie, la biologie, la botanique, l'informatique, etc) soit d'un domaine du mot (telles que l'étymologie, l'homonymie, etc.). Dans ce cas, ils sont dits être des dictionnaires linguistiques. Dans le cadre de cette thèse, nous nous focalisons sur les dictionnaires qui traitent d'un domaine de la connaissance.

Les dictionnaires, qu'ils soient spécialisés ou généraux, obéissent à la norme ISO 1951 [ISO1951, 2007]. Cette dernière permet de représenter des entrées de n'importe quel type de dictionnaire, qu'il soit général ou spécialisé, qu'il soit sous forme électronique ou papier, ou encore monolingue ou multilingues. Elle propose une structure générique et formelle indépendante des supports de publication.

Selon cette norme, un dictionnaire recense les lemmes d'une langue source. Chaque lemme constitue une entrée du dictionnaire à laquelle on rattache la définition du lemme et éventuellement d'autres mots clés ou éléments de données tels que sa forme abrégée, sa prononciation, ses variantes orthographiques, son ou ses antonymes, son ou ses synonymes, son étymologie, la liste des lemmes associés, son usage

géographique, ses traductions dans les différentes langues cibles, etc. Plusieurs entrées peuvent être regroupées dans une classe unique afin de rassembler les mots-clés connexes.

### 2.2.5 Synthèse

Les différentes ressources sémantiques exposées ci-dessus ont pour point commun la description des connaissances de domaines.

Bien que ces types de ressources soient utilisés par les méthodes de recherche d'information (traditionnelles ou par croisement de langues), ils se distinguent sur plusieurs points de vue.

Les dictionnaires, contrairement aux autres types de ressources, ont pour première vocation la définition des sens des mots. Les classifications, taxonomies, thésaurus et ontologies rendent plutôt compte de l'organisation du domaine. Ils inventorient les termes ou concepts d'un domaine tout en les hiérarchisant.

La sémantique du lien hiérarchique, pour ce qui est des thésaurus, n'est pas directement exploitable par des systèmes automatisés dans la mesure où dans une chaîne de lien hiérarchique, la transitivité risque de ne pas être maintenue lorsque cette chaîne regroupe plusieurs types de liens hiérarchiques (« est générique de » et/ou « fait partie de » et/ou « instance de »). Ceci n'est pas le cas des ontologies où un lien n'a qu'une sémantique à la fois et des autres types de ressources (classification et taxonomies) qui ne disposent que du lien hiérarchique « Est un ».

En termes de richesse de la structure pour la description du domaine, de complexité, de degré de formalisme, et d'interopérabilité nous pouvons classer les ontologies en première position suivies des thésaurus puis des taxonomies et des classifications (voir tableau 2.2). De plus, les ontologies, de par le mécanisme de raisonnement qu'elles offrent, permettent d'inférer des connaissances, possibilités que n'offrent pas les autres types de ressources sémantiques.

Pour ce qui est de la variation intralinguistique du sens des termes qui est au

coeur de la problématique traitée dans cette thèse, nous constatons que seuls les dictionnaires spécialisés dans leurs descriptions des entrées mentionnent parfois les différents sens des mots. Cependant, cette description demeure en langage naturel et ne permet pas, par conséquent, une exploitation automatique et directe de cette connaissance.

| Critères / Ressources sémantiques | Richesse | Complexité | Degré de formalisme | Intéropérabilité |
|---|---|---|---|---|
| Classification | • | • | • | • |
| Taxonomie | •• | •• | •• | •• |
| Thésaurus | ••• | ••• | ••• | ••• |
| Ontologie | •••• | •••• | •••• | •••• |

FIGURE 2.2 – Tableau comparatif des ressources sémantiques modélisant des domaines

## 2.3 Ressources sémantiques pour la description et l'organisation de contenus

Rechercher une information spécifique dans un contenu large en l'absence d'information sur les ressources informationnelles constituant ce contenu, est une tâche quasiment impossible. Pour satisfaire ce besoin en information, des ressources sémantiques permettant de nous renseigner sur le contenu des ressources informationnelles et parfois même de les organiser, peuvent être constituées. Parmi ces ressources, nous pouvons citer les méta-données, les folksonomies et les Topic Maps. Les sections suivantes présentent chacune d'elles.

### 2.3.1 Les méta-données

Dans le cadre de la gestion de contenu, les méta-données se définissent comme étant des informations concernant des ressources [Garshol, 2004]. Plus précisement, il s'agit d'ensembles de données structurées pour décrire, expliquer, localiser des ressources et en faciliter la recherche, l'usage et la gestion [NISO, 2004].

Dans [Morel-Pair, 2007] sont recensées plusieurs catégories de méta-données dont :

- Les méta-données déscriptives du contenu (titre, résumé, mots-clés, langue, couverture temporelle et spatiale, etc.) dont l'objectif est l'amélioration de la recherche d'information et la découverte de ressources,
- Les méta-données administratives (propriété intellectuelle, responsabilité, sources utilisées, etc.) et techniques (format, taille, logiciels de consultation, largeur, hauteur, nombre de bits, compression pour les documents d'un format image, etc.) dont l'objectif est la gestion des ressources,
- Les méta-données de gestion des archives (elles sont en cours de normalisation par la commission ISO/TC46/SC11),
- Les meta-données de personnalisation pour la gestion des droits d'accès (droits d'auteur) et d'usage (droits d'impression, de reproduction, de modification, etc.).

Ces méta-données sont, selon le cas, soit présentes dans la ressource elle-même, soit dans un fichier externe qui lui est associé ; la seconde possibilité est la plus pertinente pour la recherche d'information.

Dans un souci d'interopérabilité, plusieurs normes de méta-données ont été élaborées dont la norme ISO 15836 :2009 qui regroupe l'ensemble des éléments de méta-données du Dublin Core. Ces méta-données s'appliquent à tout type de ressources et ont pour objectif l'amélioration de la recherche d'information. D'autres normes plus spécifiques à des domaines d'activité existent. Nous pouvons citer la norme ISO 19115-2 :2009 pour la description de documents géographiques et la norme ISO/CEI 19788-1 pour la description de ressources pédagogiques. Enfin, notons l'existence d'autres types de méta-données mais cette fois ci non normalisées, pour décrire le contexte tel qu'il est défini dans les applications sensibles au contexte [Svensson, 2009].

### 2.3.2 Les folksonomies

Les folksonomies[6] sont des collections d'annotations (appelées « Tags ») créées de manière collaborative par plusieurs utilisateurs [Sergieh et al., 2013]. Ces annotations servent à décrire des ressources informationnelles. Elles sont librement attribuées par des utilisateurs d'une communauté ; d'où le terme « social tagging » habituellement associé aux folksonomies.

Deux types de folksonomies sont à considérer : les folksonomies larges et restreintes [Weller et al., 2010]. Dans le cas de folksonomies larges, une ressource peut être annotée de tags provenant d'utilisateurs différents. Dans le second cas, chaque tag n'est enregistré pour un document qu'une seule fois.

Notons enfin que les annotations générées ne font pas partie d'un vocabulaire contrôlé et ne sont régies par aucune norme qu'elle soit nationale ou internationale.

### 2.3.3 Les Topic Maps

L'idée des Topic Maps est empruntée aux indexes se trouvant en fin de livre dans lesquels sont mentionnés un certain nombre de termes pouvant être recherchés et les numéros de pages où ils se trouvent.

Actuellement, les Topic Maps sont un standard ISO 13250 et servent à décrire et organiser des contenus. Ce sont des réseaux sémantiques placés au dessus des contenus et construits moyennant l'exploitation de trois éléments d'information fondamentaux : le topic ou thème, l'association entre thèmes et l'occurrence.

Le thème représente un sujet que le créateur de la Topic Map souhaite représenter et pour lequel des ressources sont disponibles pour fournir de la connaissance.

L'association représente un lien sémantique bidirectionnel entre deux topics. Elle est spécifiée par le créateur de la Topic Map selon les besoins en information et selon l'application à laquelle elle est destinée. Elle permet de naviguer dans la Topic Map

6. Le terme anglais « folksonomy » est un terme créé par Thomas Vander Wal en combinant le terme « taxonomy » avec « folk » [Steele, 2009].

et rend possible l'interconnexion des ressources via les topics qui les référencent.

L'occurrence permet de lier un topic à une ressource informationnelle. Ainsi tout topic a une ou plusieurs occurrences qui regroupent toutes les informations permettant d'accéder aux différentes ressources concernées par ce topic.

La norme ISO offre, pour ce qui est des Topic Maps, deux mécanismes intéressants : le mecanisme de « scoping » et le mécanisme de « réification ». Le mécanisme de « scoping » permet de mentionner le contexte pour lequel un thème est lié à une ressource. Le second mécanisme offre la possibilité de construire des sujets à partir d'associations entre topics.

Dans la mesure où une Topic Map organise tout un contenu, son organisation des ressources informationnelles est orientée utilisation. Elle est décrite dans un document séparé.

### 2.3.4 Synthèse

La fourniture d'information sur un contenu d'une ressource informationnelle peut être réalisée, comme on l'a vu dans les paragraphes ci-dessus, via les méta-données, les folksonomies ou les Topic Maps. Ces trois ressources sémantiques ont toutes pour vocation la description de contenus.

De par leurs structures, les Topic Maps et les folksonomies permettent, en supplément, de retrouver pour chaque thème ou annotation, les documents y afférants. Ajouté à cela, les Topic Maps offrent aussi la possibilité de naviguer au sein des contenus et cela grâce aux liens d'association et d'occurrence. Les folksonomies, quant à elles, offrent aussi la possibilité d'identifier des communautés d'intérêt dés lors qu'un tag eu été enregistré par plusieurs utilisateurs.

Du fait de la participation des utilisateurs dans la description des ressources, les folksonomies incluent une dimension sociale qui n'est pas prise en charge par les autres ressources sémantiques.

Les Topic Maps, contrairement aux méta-données, permettent une description

des contenus orientée utilisation du fait de la flexibilité qu'elles offrent dans la détermination des topics. D'ailleurs, il est possible de concevoir plusieurs Topic Maps pour un même contenu. Il est aussi possible de construire une Topic Map dans laquelle les topics et/ou les liens d'associations font partie d'éléments constituant une ressource sémantique tels que les ontologies et les thésaurus.

Pour des contenus volumineux et en perpétuelle évolution, il est évident qu'il n'est pas envisageable de construire manuellement une Topic Map. Des approches de construction et de gestion des évolutions de Topic Maps s'imposent. En revanche, l'annotation sociale peut s'avérer une approche intéressante. Cependant, elle génère un certain nombre de problèmes qui rendent l'indexation des contenus complexe. Parmi ces problèmes on peut citer : l'impossibilité de distinguer les polysémes et les homonymes, la possibilité de combiner plusieurs langues pour un même contenu sans avoir la possibilité de les reconnaître, etc. Un inventaire des limites des folksonomies est fourni dans [Weller, 2010].

Enfin, notons qu'aucune des ressources sémantiques énoncées ci-dessus, ne dispose dans sa configuration actuelle d'artifices facilitant la prise en charge de la variation du sens des termes et concepts.

## 2.4 Conclusion

Ce chapitre a permis de faire un tour d'horizon sur les ressources sémantiques contribuant à la description et à l'organisation de contenus. Nous nous sommes focalisés dans un premier temps sur les ressources permettant de décrire des domaines, puis dans un second temps, sur les ressources permettant de décrire des contenus. Des synthèses associées à chacune des deux parties, il ressort que la modélisation du changement du sens n'est pas supportée par les modèles de ressources sémantiques actuels. Par conséquent, le choix du type de ressource sémantique décrivant plusieurs domaines ainsi que celles permettant de décrire et d'organiser des contenus pluridisciplinaires et multilingues va plutôt reposer sur d'autres critères. Une fois le

choix effectué, il va falloir réfléchir à une possibilité d'extension de ces ressources afin de prendre en charge la variabilité du sens des termes dans des contenus. C'est la démarche entreprise dans le cadre de cette thèse.

Pour ce qui est du modèle de ressources sémantiques décrivant plusieurs domaines, notre choix se serait porté sur les ontologies de domaine pour leur degré de formalisme, leur degré d'expressivité et leur capacité de raisonnement. Cependant, les contraintes associées au programme FSP Maghreb d'où est née la problématique de cette thèse, nous ont imposé comme ressource sémantique un wiktionnaire spécialisé (dans notre cas pour les SHS) qui n'est autre qu'un dictionnaire électronique sous un format wiki. La conception de sa structure et sa mise œuvre ainsi que l'application permettant de la créer, la manipuler et la gérer font partie du chapitre 4 de cette thèse.

En ce qui concerne la description et l'organisation de contenus multilingues et pluridisciplinaires, notre choix s'est porté sur les Topic Maps pour les avantages qu'elles offrent en termes de navigation au sein des contenus et de recherche et de partage de connaissances. Nous montrons dans le chapitre suivant les limites des modèles de Topic Maps existants ainsi que celles des approches permettent leur construction et leur évolution. Ceci permettra d'introduire notre modèle et nos approches dont la présentation détaillée fait l'objet du chapitre 5 de cette thèse.

# Chapitre 3

# Modèles et approches de construction et d'évolution de Topic Maps : Un état de l'art

Les Topic Maps, souvent appelées « GPS de l'univers de l'information [1] », dérivent du concept théorique des réseaux sémantiques. Elles ont pour but d'organiser des contenus informationnels. Grâce aux réseaux de liens sémantiques entre les sujets qu'elles représentent, elles permettent une navigation facile et sélective améliorant ainsi la recherche de l'information dans les contenus.

Les Topic Maps, au même titre que les autres ressources sémantiques, ont été au centre des préoccupations de travaux de chercheurs intéressés par la gestion documentaire ou encore par la gestion des connaissances. La littérature contient une panoplie de travaux sur la construction, l'évolution, la visualisation de Topic Maps ainsi que sur leur utilisation dans les organisations.

Le but de ce chapitre est de dresser un état de l'art critique des travaux les plus récents dans ce domaine. Après quelques généralités constituées d'une description

1. Ontopia. http ://www.ontopia.net/topicmaps/materials/tao.html (consulté le 13/7/2012)

des concepts de base des Topic Maps, de leur modèle standard et d'une énumération des langages de description et de manipulation associés, nous fournissons dans la même section quelques exemples de domaines d'utilisation. Nous présentons aussi les différentes architectures de mise en œuvre existantes. La section 2 est consacrée à un état de l'art sur les modèles étendus de Topic Maps proposés dans la littérature. Les deux sections suivantes dressent un inventaire des travaux les plus récents dédiés à la construction et à la gestion des évolutions des Topic Maps. L'avant dernière section donne un aperçu des outils d'édition existants. La conclusion synthétisera le contenu de ce chapitre et introduira nos contributions.

## 3.1 Concepts, modèle de base et langages associés aux Topic Maps

Les Topic Maps sont un standard ISO/IEC 13250[2] dont les premières publications datent de l'an 2000. Ce standard regroupe plusieurs parties dont la partie relative à la description des concepts de base, celle relative au modèle de données, celle relative au langage de description (syntaxe) et enfin celle relative au langage d'interrogation.

Les sous sections suivantes présentent, de façon succincte, ces différents constituants.

### 3.1.1 Les concepts de base des Topic Maps

Le paradigme Topic Map défend le fait que le moyen le plus efficace pour organiser des ressources informationnelles est la notion de thème ou sujet (Topic) [ISO/IEC, 2003]. Ceci fait du concept « Topic », le concept central des Topic Maps. Ainsi, une Topic Map est une collection de topics (thèmes). Un topic peut être lié :

– à un ou plusieurs topics via un lien sémantique appelé « association »,

2. Topic Maps. http ://www.topicmaps.org/standards

– et à un ensemble de ressources traitant le même sujet via des liens d'occurrence.

Les concepts de Topic, d'association et d'occurrence constituent ainsi les concepts de base des Topic Maps.

On peut assigner à un topic un ou plusieurs noms et préciser les différents rôles qu'il joue dans les différentes associations auxquelles il participe. Les associations assurent la navigation dans une Topic Map.

Les liens d'occurrence rattachés à un topic permettent d'accéder aux ressources informationnelles traitant du sujet représenté par ce topic. Ces ressources peuvent être de n'importe quel type (multimédias, documents non structurés ou semi structurés, bases de données, etc.) et de n'importe quel format. La figure 3.1 présente

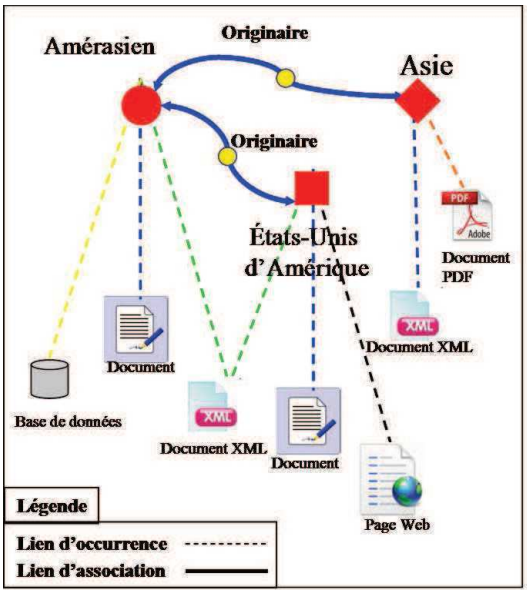


FIGURE 3.1 – Exemple de Topic Maps

un exemple de Topic Map constituée de trois topics : « Amérasien », « Asie » et « États-Unis d'Amérique ». Cette Topic Map comprend deux occurrences de l'association « Originaire » : l'une entre le topic « Amérasien » et le topic « Asie », et l'autre entre le topic « Amérasien » et le topic « États-Unis d'Amérique ». Les liens d'occurrence sont au nombre de huit dont trois partant du topic « Amérasien » vers

une base de données, un document textuel et un document XML.

### 3.1.2 Le modèle de données des Topic Maps

Le modèle de données des Topic Maps (TMDM pour Topic Map Data Model) est le cœur de la famille des standards de Topic Maps [ISO/IEC, 2008]. Il fournit une description sémantique du méta-modèle en langage naturel. Il précise, par ailleurs, les règles de fusion des Topic Maps et de leurs éléments. La figure 3.2 présente un extrait du TMDM [ISO/IEC, 2008].

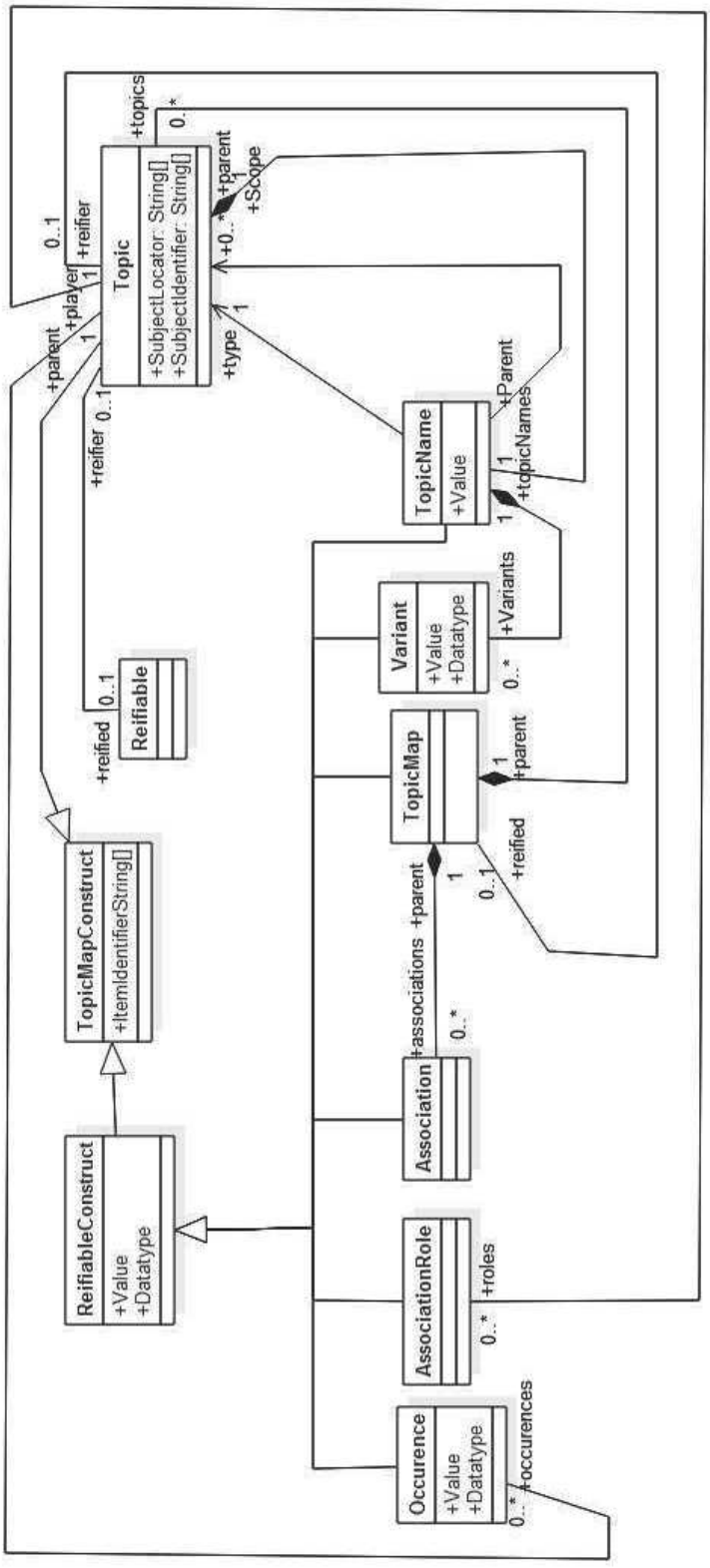


FIGURE 3.2 – Extrait du modèle de données de Topic Maps [ISO/IEC, 2008]

Dans cette figure 3.2, les classes « TopicName » et « Variant » rassemblent respectivement les noms de base des topics et les variantes des noms de topics. En effet, un topic pouvant avoir plusieurs noms, un de ses noms sera considéré comme nom de base. Les autres noms seront classés comme des variantes du nom de base.

Un topic est identifiable par deux propriétés : le « SubjectLocator » et le « SubjectIdentifier ». La première propriété est utilisée pour faire référence à la ressource qui constitue le thème du topic. À titre d'exemple, si la chaîne de caractères http ://www.cnam.fr est un SubjectLocator alors ce topic représentera le site du Cnam et non l'institution Cnam.

La seconde propriété n'est autre que l'Identificateur Uniforme de Ressources (URI, Uniform Ressource identifier) ou l'Identificateur Internationalisé de Ressources (IRI, Internationalized Resource Identifier) qui sont des séquences compactes de caractères identifiant une ressource physique ou abstraite. Ces identificateurs obéissent à la syntaxe normalisée RFC 3986 [Berners-Lee, 2005b] et RFC 3987 [Berners-Lee, 2005a] soutenue par le W3C dans [W3C, 2005].

Les classes « TopicMapConstruct » et « Reifiable » sont des classes abstraites. La première rassemble tous les composants d'une Topic Map. La seconde rassemble les composants réifiables, c'est-à-dire les composants sur lesquels le mécanisme de reification est applicable. En effet, le standard TMDM donne la possibilité, par le biais d'un mécanisme de réification similaire à celui existant dans le modèle RDF, de construire un thème d'un topic à partir d'autres composants de cette Topic Map tels qu'un lien d'occurrence ou une association. Par conséquent, tous les composants faisant partie de la classe « Reifiable » sont ceux sur lesquels peut s'appliquer la réification. La figure 3.3 illustre un exemple de réification. Dans cette figure, un nouveau topic nommé « Amérasien originaire des États-Unis d'Amérique » a été créé à partir de l'association « originaire » existante entre le topic de nom « Amérasien » et le topic « États-Unis d'Amérique ». Ce nouveau topic est de nouveau rattaché à des ressources informationnelles via le lien d'occurrence.

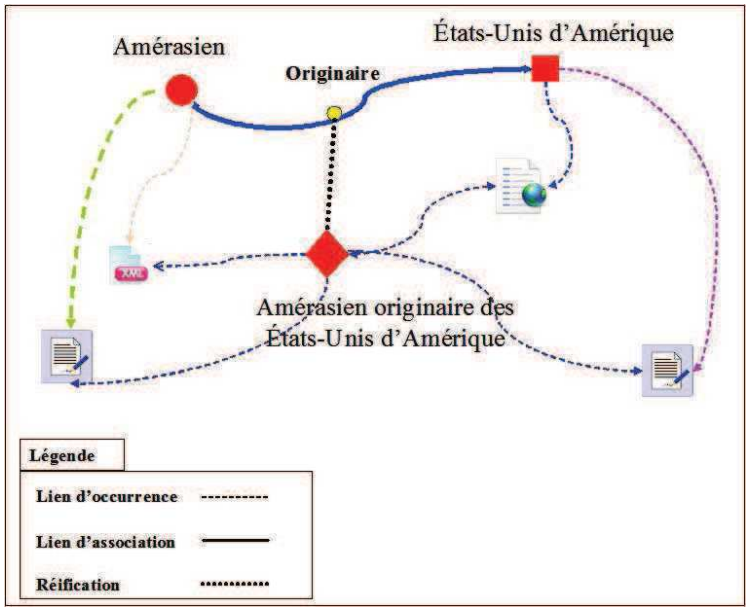


FIGURE 3.3 – Exemple de réification d'une association entre deux topics

En plus du mécanisme de réification, le standard TMDM offre un second mécanisme, appelé « scope », qui permet de fournir un niveau de conceptualisation de telle sorte que la Topic Map puisse correspondre à plusieurs visions utilisateurs. Il peut être vu comme un mécanisme de filtrage fondé sur les propriétés des topics. Il met en place des contextes à travers lesquels des noms et des occurrences de topics et des liens d'associations sont assignés à des topics. Une application de ce mécanisme pourrait être la prise en compte du multilinguisme dans les contenus. Dans ce cas, les noms de topics d'une langue pourraient être regroupés, si cela s'avère nécessaire, dans un même contexte donc dans un même scope. Ce mécanisme se traduit dans le modèle de la figure 3.2 par l'association « scope » entre la classe « Topic » et la classe « TopicName ».

Pour conclure cette sous section, notons que le TMDM a été inclus, par l'OMG, dans l'ODM (Ontology Meta-Model Definition) [OMG, 2009] afin de favoriser l'applicabilité des concepts du MDA (Model Driven Architecture) à l'ingénierie des Topic Maps et par conséquent, de permettre l'interopérabilité entre UML, RDF, OWL

CL et les Topic Maps [Hart and Emery, 2004]. Un méta-modèle standard TM-UML équivalent a été élaboré en conséquence.

### 3.1.3 Les langages de description et de manipulation de Topic Maps

Le standard ISO/IEC 13250 inclut, dans l'une de ses parties, un langage de spécification pour décrire une Topic Map à l'aide des concepts et des mécanismes présents dans le TMDM. Il s'agit de XTM (XML Topic Map). Sa syntaxe est textuelle et est fondée sur XML. Sa version 2.0 constitue la dernière version et a été éditée en 2006. Il constitue le format d'échange et de partage le plus utilisé pour les Topic Maps. Cependant, l'utilisation de standards voisins, tels que RDF ou OWL, sont aussi envisageables.

Pour ce qui est des langages d'interrogation, une spécification standardisée d'un langage dédié aux Topic Maps, dont la dernière version date de 2008, est en cours d'élaboration sous l'appellation ISO 18048. Il s'agit du langage TMQL (Topic Map query language). Une des raisons pour lesquelles les travaux sur TMQL sont inachevés, est son incapacité à exprimer des requêtes complexes. L'entreprise Ontopia[3] propose aussi le langage TOLOG. Ce dernier est aussi dédié aux Topic Maps. Il s'inspire des langages PROLOG et DATALOG. Il présente aussi des insuffisances telle que l'impossibilité de poser des requêtes de modification. Fort heureusement, d'autres langages non dédiés aux Topic Maps, tel que SPARQL, sont disponibles et peuvent permettre d'interroger et de modifier une Topic Map.

### 3.1.4 Quelques domaines d'application des Topic Maps

La technologie Topic Maps possède un large spectre d'utilisation de par les avantages qu'elle offre en termes d'organisation de contenus et de recherche d'informations. Le tableau ci-dessous (voir figure 3.4) recense quelques références bibliographiques parmi les plus récentes et donne un aperçu non exhaustif de quelques do-

3. Ontopia. http ://www.ontopia.net/

maines d'application des Topic Maps.

| Domaine d'application | Références bibliographiques |
|---|---|
| Education | [Burita, 2014] |
| Recommandation | [Garrido and Ilarri, 2014] |
| Conception assistée par ordinateur | [Goldhammer, 2012] |
| Médecine | [Gomoi et al, 2012], [Damen, 2012], [Dragu et al., 2011] |
| Ingénierie de moulage | [Jung et al., 2008] |
| Santé et sécurité au travail Ecologie | [Arndt et al., 2011] |
| Héritage culturel | [Maicher et al., 2009] |

FIGURE 3.4 – Domaines d'application des Topic Maps

Dans [Burita, 2014], une Topic Map couvrant le domaine des sciences militaires ainsi que les domaines de la médecine et de l'ingénierie a été élaborée. L'objectif est de favoriser l'échange et le partage de connaissances entre enseignants et étudiants de plusieurs universités européennes.

[Garrido and Ilarri, 2014] ont construit une Topic Map qu'ils utilisent, dans leur système de recommandation proposé, pour le filtrage et la suggestion d'items aux utilisateurs.

Les auteurs de [Goldhammer, 2012] ont conçu une Topic Map permettant de décrire des fichiers CAD (Computer-aided design) de conception assistée par ordinateur. Pour ce faire, ils ont utilisé l'outil Omnigator d'Ontopia pour construire manuellement la Topic Map.

Les travaux de [Gomoi et al., 2012], [Damen and Van den Bulcke, 2012] et [Dragu et al., 2011] concernent la construction de Topic Maps pour le domaine médical. [Gomoi et al., 2012] et [Dragu et al., 2011] ont conçu une Topic Map comme support à la décision clinique. Cette Topic Map contient des enregistrements médicaux virtuels (VMR ou Virtual Medical Record). La problématique étudiée dans [Damen and Van den Bulcke, 2012] est dans le domaine de la recherche clinique. Les auteurs de cette publication ont exploité la technologie des Topic Maps dans le

cadre de l'automatisation des choix des candidats aux essais cliniques. La Topic Map conçue permet, dans le cadre de ce travail de recherche, le partage de concepts liés au domaine de la recherche clinique.

[Jung et al., 2008] ont proposé un processus de création automatique de Topic Maps à partir de rapports de projets d'ingénierie de moulage dans un hub de collaboration. Leur méthode de construction de la Topic Map est extraite de leurs expériences précédentes de projets.

La Topic Map construite dans [Wang and Guo, 2007], est dédiée au domaine de la sécurité de l'aviation. Elle a été construite à partir d'un corpus de documents d'enquêtes associés à la sécurité de l'aviation.

[Arndt et al., 2011] ont décrit leur expérience dans la construction de deux Topic Maps : une dans le domaine de la santé et la sécurité au travail et une seconde dans le domaine de l'écologie et plus précisément dans le domaine de la gestion des influences écologique.

[Maicher et al., 2009] ont présenté plusieurs exemples d'application de Topic Maps, dans plusieurs projets de pays différents (Italie, Finland, Royaune-unis, Allemagne et Corée du sud), dans le domaine de l'héritage culturel notamment pour le recensement du patrimoine et pour les musée virtuels.

### 3.1.5 Les architectures de Topic Maps

En règle générale, deux types d'architectures sont à considérer : distribuées et centralisées. Une architecture distribuée est une architecture fondée sur un réseau sur lequel repose une collection de ressources hétérogènes, distribuées et connectées. Ce réseau peut être un réseau pair-à-pair (P2P) ou une grille de calcul (Grid). L'architecture centralisée, quant à elle, adhère à une politique de regroupement des ressources de traitement et de stockage en un seul point.

L'architecture centralisée a été choisie majoritairement et implicitement dans la plupart des travaux de recherche sur les Topic Maps. Par conséquent, la plupart

des contributions supposent l'existence, la création ou l'évolution de Topic Maps centralisées.

Nous avons recensé très peu de contributions allant dans le sens d'une architecture distribuée. Parmi elles citons, *TMShare* [Ahmed, 2003], *Shark*(SHARed Knowledge) framework [Schwotzer, 2008], *Kpeer* (Knowledge peer) [Sigel, 2004] et *TMGrid* [Korthaus and Hildenbrand, 2003], [Geisser et al., 2008], [Korthaus et al., 2006] et [Korthaus and Schader, 2006].

Dans ([Korthaus and Hildenbrand, 2003], [Geisser et al., 2008], [Korthaus et al., 2006] et [Korthaus and Schader, 2006]), une proposition de Topic Maps sur une grille de calcul (Topic Maps Grid) est présentée (voir figure 3.5).

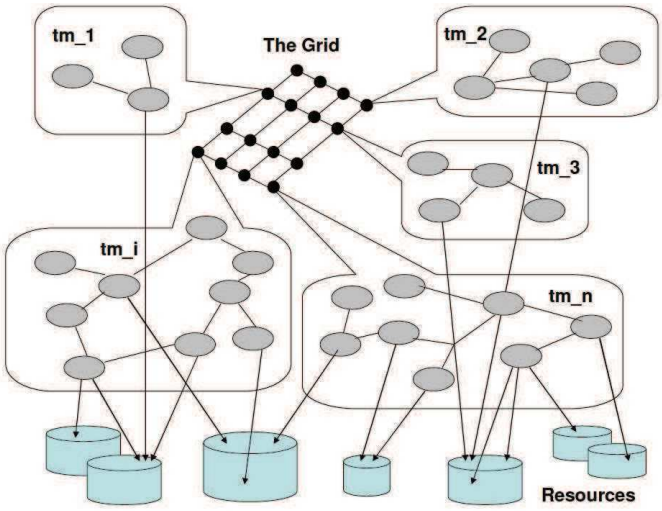


FIGURE 3.5 – Topic Map Grid [Korthaus et al., 2006]

Comme illustré dans la figure 3.5, la grille se compose d'un ensemble de nœuds. Chaque nœud fournit une base de connaissances sous forme d'une ou de plusieurs Topic Maps. Ces Topic Maps sont exposées via une interface. Chaque nœud peut, optionnellement, héberger des contenus informationnels locaux.

Dans ce type d'architecture, on donne l'impression à l'utilisateur de travailler sur

une grande Topic Map. Chaque membre utilisateur de la grille exécute un service qui fournit un accès aux objets de la Topic Map, aux fonctionnalités de fusion et à une interface qui retourne des topics pertinents. Le développement d'un wiki sémantique comme front-end à l'infrastructure de la Topic Map Grid a été décrit dans [Geisser et al., 2008]. Appelé ToMaWiki, ce wiki sémantique fournit des composants de navigation, fondés sur les graphes, servant à afficher les fragments de la Topic Maps globale.

[Ahmed, 2003] propose une Topic Map construite sur une architecture pair-à-pair (P2P) décrite dans ce papier. Les pairs, nommés TMshare, échangent des fragments de Topic Maps exprimés en XTM. Les pairs interrogés renvoient les fragments de Topic Maps demandés. Ces fragments sont, par la suite, fusionnés au niveau du pair client.

Dans [Schwotzer, 2008], un projet open source, nommé Shark (SHARed Knowledge), de mise en œuvre de Topic Maps sous une architecture pair-à-pair, est présenté. Dans ce projet sont fournis un protocole pair-à-pair sans état (stateless), appelé KEP (Knowledge Exchange Protocol) et une API pour la construction d'applications P2P. Cette API, au même titre que TMQL, est utilisée pour l'extraction des fragments de Topic Maps. Dans ce projet a été aussi mis en œuvre le modèle ACP (autonomous context-aware peers) à des fins d'implémentation de pairs autonomes pouvant échanger des connaissances.

Enfin, dans [Sigel, 2004], est présenté un autre projet nommé KPeer(Knowledge peer), où une architecture fondée sur la notion de registre de sujets publiés a été utilisée pour lier les pairs.

## 3.2 État de l'art sur les modèles de Topic Maps étendus

Nous avons recensé plusieurs propositions d'extension du modèle standard de Topic Maps. Chacune de ces propositions est motivée par la prise en compte d'une problématique de recherche spécifique. Les paragraphes qui suivent décrivent chacun

de ces modèles.

### 3.2.1 Le modèle de Topic Map SocioTM

L'idée sous-jacente de cette proposition [Sasha and Rudan, 2008] est la prise en compte de profils sociaux d'utilisateurs ou de groupes d'utilisateurs (préférences, comportements, etc.) à des fins de personnalisation d'une Topic Map globale. Le but est la construction à la volée d'une Topic Map personnalisée à partir d'une Topic Map globale persistente. L'extension apportée au modèle standard est, par conséquent, l'ajout de métadonnées à chaque concept de base d'une Topic Map (topic, association et occurrence). Ces métadonnées correspondent à des mesures de pertinence. Elles sont initialisées et mises à jour de façon permanente grâce à un système apprenant qui se base sur les profils sociaux des utilisateurs et sur les actions qu'ils ont accomplies sur la Topic Map.

### 3.2.2 Le modèle de Topic Map Hypertopic

Le modèle Hypertopic ([Cahier et al., 2004], [Zaher et al., 2007], [Zaher et al., 2006], [Cahier et al., 2006], et [Zaher et al., 2007]) a été conçu pour pallier les limites du modèle standard dans la prise en compte des points de vue des acteurs ou groupes d'acteurs sensés utiliser un contenu donné. La première application de ce modèle a été destinée à un grand groupe international de télécommunications qui souhaitait organiser le catalogue décrivant leurs projets R&D [Cahier, 2005]. La documentation associée à ce catalogue a été thématisée selon sept points de vues correspondants aux grands métiers de l'entreprise.

Le modèle Hypertopic étend le modèle standard par l'intégration de trois concepts supplémentaires qui sont : les concepts d'entité, de point de vue et d'attribut standard (voir figure 3.6). Une entité sert à regrouper des ressources traitant d'une même thématique. Elle est de ce fait liée aussi bien à des ressources qu'à des topics. Elle dispose de propriétés la caractérisant. Plusieurs ressources peuvent être associées à

une entité et plusieurs entités peuvent être associées à une même ressource.

Le concept de point de vue permet la création de différents regroupements de thématiques selon des points de vue d'acteurs ou de groupes d'acteurs. Un thème est inclus dans un point de vue alors qu'un point de vue peut ne pas inclure plusieurs thèmes donc plusieurs topics.

Le concept d'attribut standard sert à la caractérisation des entités.

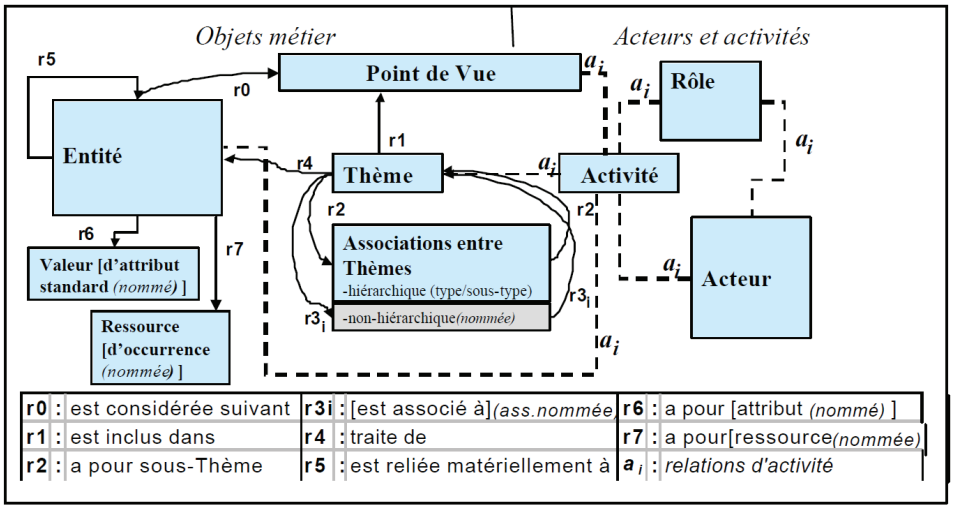


FIGURE 3.6 – Modèle Hypertopic [Cahier et al., 2004]

### 3.2.3 Le modèle ETM de Topic Maps

Les auteurs de ce modèle ([Lu and Feng, 2009],[Lu et al., 2008b], et [Lu et al., 2008a]) ont souhaité, via la structuration en couches ou niveaux d'une Topic Map, introduire un peu plus d'expressivité dans le modèle standard afin d'améliorer la navigation au sein des contenus.

L'extension du modèle standard a consisté à ajouter deux concepts : le concept de cluster et le concept d'éléments de connaissance (voir figure 3.7). Un cluster sert à regrouper des thématiques. Un élément de connaissance est la plus petite unité de connaissance associée à un domaine. Il sert à décrire plus finement les ressources

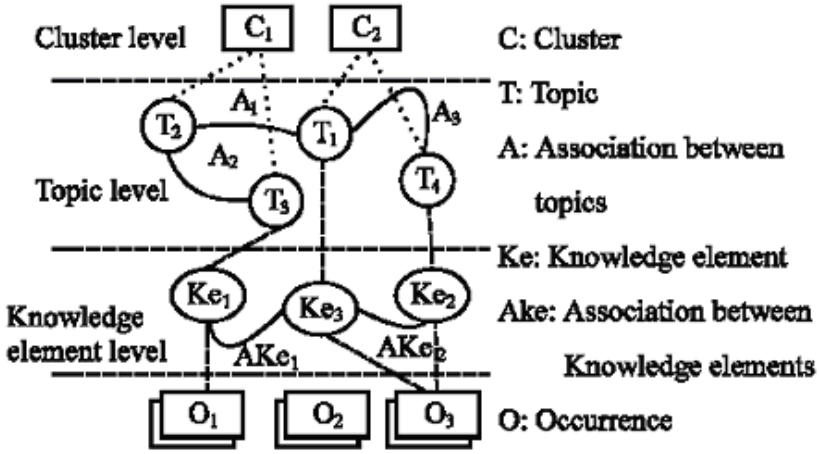


FIGURE 3.7 – Structure d'un ETM [Lu and Feng, 2009]

à travers les éléments de connaissance qu'elles renferment. Ainsi, un élément de connaissance est lié à un ou plusieurs thèmes via le lien « TopicElementAssoc ». Les éléments de connaissances sont liés entre eux via le lien « ElementAssoc » faisant partie d'un ensemble prédéfini. La figure 3.8 est un extrait de la structure d'une ETM construite dans le domaine des réseaux informatiques.

### 3.2.4 Le modèle de Topic Map de la méthode CITOM

Trois aspects ont motivé les auteurs à proposer une extension du modèle standard ([Ellouze et al., 2012], [Ellouze et al., 2008] et [Ellouze et al., 2010]). Le premier concerne la prise en charge du multilinguisme. Le second est l'introduction d'une autre forme de recherche d'information. Le troisième aspect est l'introduction du mécanisme d'élagage afin de pallier le phénomène de « grossissement » de la Topic Map.

Pour ce qui est du multilinguisme, en plus de l'exploitation du scope, les auteurs ont introduit, dans leur modèle étendu, le concept de facette (métadonnées associées aux occurrences d'une Topic Map) figurant dans les premières propositions de modèle

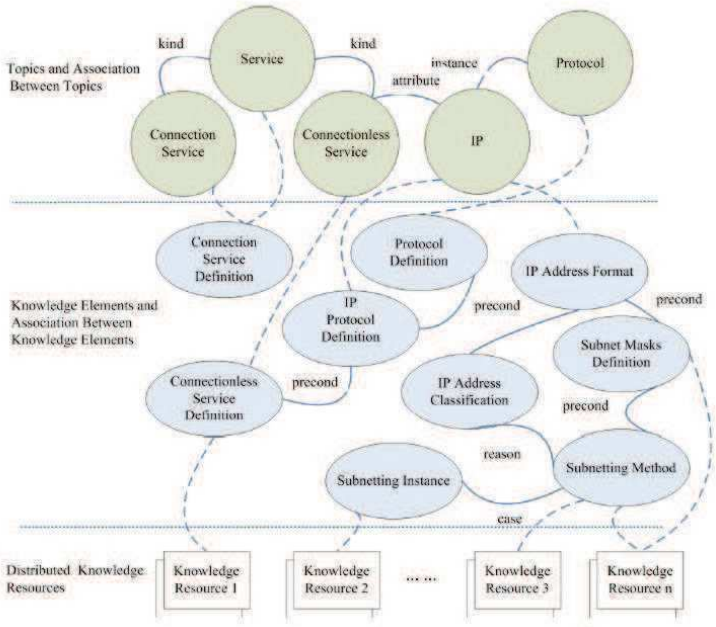


FIGURE 3.8 – Exemple d'une ETM [Lu et al., 2008b]

de Topic Maps mais qui n'a pas été inclus dans le TMDM. Les facettes dans ce modèle comportent la langue du document lié au topic. Elles permettent de filtrer les documents selon leur langue.

Le second aspect est matérialisé par l'ajout de topics de type « question potentielle » et de deux nouveaux concepts : le concept de segment thématique et le concept de lien d'usage. Les questions potentielles favorise une recherche orientée question. Le premier concept contribue à un meilleur ciblage de la connaissance recherchée par l'utilisateur au sein d'un document. Les topics ont ainsi des occurrences permettant d'atteindre un fragment du document textuel. Le second concept, de par sa structure (c'est-à-dire un lien entre les mots clés d'une question potentielle et les mots clés de sa réponse), offre la possibilité à utilisateur de choisir entre une navigation traditionnelle au sein des ressources et une recherche d'information via les questions.

Le troisième aspect est concrétisé par l'ajout de méta-données contribuant, pour certaines, à l'élagage de la Topic Map, et pour d'autres, à l'organisation en niveaux de la Topic Map. Ces métadonnées sont associées aux topics. Elles sont de deux types : métadonnées orientées usage et métadonnées orientées structure. Les méta-données orientées usage fournissent des renseignents sur l'importance des topics en terme de visites utilisateurs. Les métadonnées orientées structure fournissent une classification des topics en thèmes, termes et questions. Les thèmes et les questions font partie du niveau supérieur de la Topic Map. En revanche, les mots clés font partie de son niveau inférieur.

La figure 3.9 présente un extrait du modèle proposé dans CITOM.

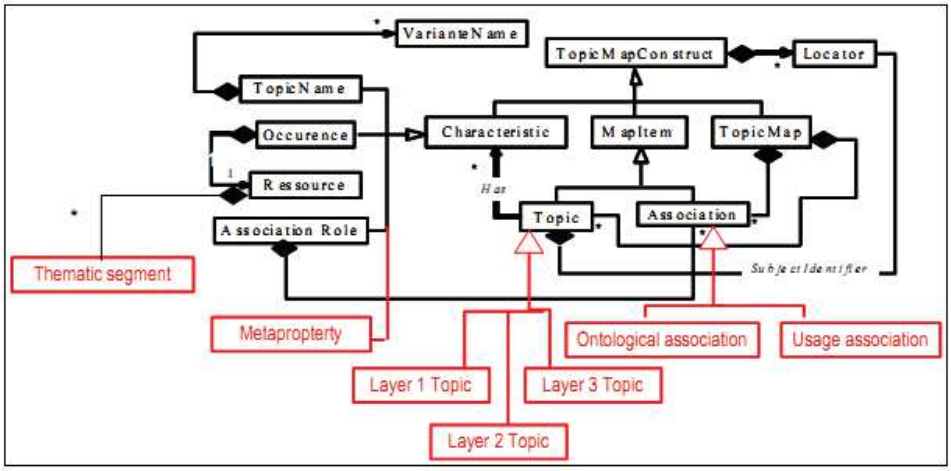


FIGURE 3.9 – Le modèle CITOM [Ellouze et al., 2012]

### 3.2.5 Synthèse sur les modèles étendus de Topic Maps

En règle générale, on peut associer aux Topic Maps deux fonctionnalités primordiales : l'organisation de contenu et la recherche d'information dans les contenus quels qu'ils soient. Le modèle standard, de par sa structure, permet de satisfaire ces deux fonctionnalités. Notre étude de l'état de l'art ci-avant nous a permis de dresser les constats suivants :

a) Via les extensions de modèles proposées, les auteurs ont tenté de répondre à la question suivante : pour un contenu donné, quelle organisation de Topic Map est à visualiser pour les utilisateurs afin de faciliter au mieux leur recherche d'information. L'aide améliorée à la recherche d'information s'est matérialisée, pour tous les modèles étendus proposés, excepté le modèle SocioTM, en permettant la construction de Topic Maps avec des niveaux d'indexation favorisant une exploration plus ciblée des contenus :
   - Le modèle HyperTopic propose d'organiser le contenu selon le point de vue utilisateur. A chaque portion de contenu, correspondante à un point de vue, il offre des vues partielles où chaque vue correspond à un thème.
   - Le modèle ETM suggère de partitionner les contenus en clusters. Chaque cluster correspond à un regroupement de thématiques. A une portion d'un contenu correspondant à un cluster, il associe plusieurs vues partielles où chaque vue correspond à une thématique.
   - Le modèle de la méthode CITOM, permet de créer des topics de type « thème » et « terme » pour pouvoir cibler une portion de contenu correspondant à une thématique puis une sous portion de contenu associé à un terme.

   Le modèle socioTM, quant à lui, constitue avant tout un support à la construction de vues partielles de Topic Maps. L'obtention d'une vue partielle d'une Topic Map permet, par conséquent, de visualiser une portion du contenu qu'elle organise. La navigation au sein de ce contenu repose sur la structure du modèle standard.

b) Seul le modèle de la méthode CITOM intègre des métadonnées permettant de procéder à un instant donné à l'élagage de la Topic Map. Cependant ces métadonnées ne concernent que les topics.

c) De même, seul le modèle de la méthode CITOM prend en charge, via sa structure, la recherche dans des contenus multilingues.

d) Enfin, aucun des modèles proposés, ne gère la variation du sens des termes ou des concepts.

## 3.3 Approches de construction de Topic Maps

Un contenu à organiser étant très souvent volumineux, la construction manuelle d'une Topic Map est une tâche non envisageable. Elle nécessite des guides méthodologiques et des outils d'aide. C'est ce qui explique l'abondance des travaux de recherche qui se sont concentrés sur cette problématique et qui ont proposé des approches. Les contributions les plus récentes sont décrites dans cette section. Un état de l'art présentant des travaux plus anciens est présenté dans [Ellouze et al., 2012].

Nous présentons une classification des approches les plus récentes en quatre catégories : celles « from scratch » (partant de rien), celles s'appuyant sur des ressources sémantiques, celles s'appuyant sur des ressources informationnelles et enfin celles combinant des ressources informationnelles et des ressources sémantiques. Étant donné le nombre de contributions récentes de la toisième catégorie, nous préférons les présenter, à part, dans la sous-section 3.3.1 ci-dessous.

En ce qui concerne les résultats de recherche relevant de la première catégorie nous avons recensé le travail original de [Bernotaityte et al., 2012]. Ce travail de recherche exploite les démarches de construction d'ontologie « from scratch » dans sa proposition de construction d'une Topic Map organisant des contenus d'un domaine en langue lituanienne. La Topic Map résultante réunit les concepts du domaine ainsi que les relations entre concepts. Les occurrences des concepts ne sont pas, bien sûr, renseignées. La première étape de la démarche est une étape de kick-off pour la capture des besoins. Elle est suivie par une étape de construction des constituants de l'ontologie de domaine sous forme de Topic Map. Dans cette étape un certain nombre de validations, similaires à celle existantes pour la construction des ontologies, sont utilisées. À notre connaissance il s'agit de la seule contribution de cette catégorie.

La deuxième catégorie de méthodes rassemble des approches exploitant des mé-

tadonnées ainsi que celles exploitant des thésaurus. La première sous-catégorie regroupe les travaux les plus anciens de cette catégorie tels que la contribution de [Lacher and Decker, 2001] et celle de [Eric Prud'hommeaux, 2002]. Ces deux approches proposent des règles de transformation de triplets RDF en éléments du TMDM. Dans ces deux approches la réification n'est pas exploitée et la mise en correspondance est purement syntaxique.

Parmi les travaux les plus récents de la seconde sous-catégorie, on peut citer la proposition de [Bold et al., 2010] où chaque terme du thésaurus est transformé en topic et chaque lien en association. Le thésaurus est d'abord transformé en un format interne spécifique avant d'être traduit en XTM en utilisant XSLT.

À notre connaissance, seule l'approche CITOM (Construction Incrémentale de Topic Map) de [Ellouze et al., 2012] fait partie de la dernière catégorie. Cette approche construit une Topic Map multilingue à partir d'un contenu multilingue constitué de documents textuels. Pour ce faire, elle exploite le contenu, c'est-à-dire les documents textuels multilingues, un thésaurus de domaine, des ontologies générales (par exemple Wordnet et WOLF) ainsi que des ensembles de questions associées aux documents. C'est une approche incrémentale. Elle construit de façon incrémentale une Topic Map $TM_i$ correspondante à un ensemble de documents D= {$d_1$, $d_2$, . . ., $d_i$, . . .} en fusionnant la Topic Map $TM_{i-1}$ correspondant à l'ensemble de document D-$d_i$ avec la Topic Map associé au document $d_i$. Chaque phase permettant de construire la Topic Map correspondant à un document $d_i$ utilise comme sources le document, un thésaurus de domaine, des ontologies générales (par exemple Wordnet et WOLF) ainsi qu'un ensemble de questions relatives à ce document et extraites de sources d'interrogation. L'ensemble des documents est intégré dans un référentiel sémantique dans lequel est indiqué pour chaque document la liste des segments thématiques qui le compose, et pour chaque segment d'un document, la liste des termes et concepts le représentant. À chacun des termes et concepts sont associés des degrés de pertinence.

La segmentation thématique est effectuée à l'aide du segmenteur thématique TextTiling. L'indexation exploite l'algorithme LSI. Ces deux mécanismes permettent d'indexer via la Topic Map des fragments de documents.

La Topic Map produite par CITOM fournit des topics dans plusieurs langues, tout en prenant en charge une des spécificités du multilinguisme, qui est l'absence éventuelle de termes sémantiquement équivalents d'une langue à une autre.

### 3.3.1 Les approches de construction de Topic Maps à partir de contenus

Les approches de cette catégorie se distinguent par la nature des ressources informationnelles constituant le contenu : des ressources non structurées tels que les documents textuels, ressources semi-structurées telles que les documents XML et enfin ressources structurées telles que les bases de données relationnelles.

La première sous-catégorie comporte, parmi les travaux les plus récents, les contributions suivantes : [Kásler et al., 2006], [Zheng et al., 2009], [Weber et al., 2010b] et [Garrido et al., 2013]. Dans [Kásler et al., 2006], la Topic Map est générée semi automatiquement à partir de documents textuels bilingues (hongrois et anglais). L'approche de construction s'appuie sur des techniques d'apprentissage et des techniques d'extraction d'informations. Elle se déroule en plusieurs étapes. La première étape consiste en la collecte de métadonnées à partir du contenu et leur stockage en XML. La deuxième étape est relative au processus d'analyse qui conduit à construire le squelette de la Topic Map. Ce processus fournit les types de topics et les mots clés. La troisième étape concerne le peuplement de la Topic Map. La dernière étape est la visualisation de la Topic Map générée.

[Zheng et al., 2009] ont décrit une approche de construction de Topic Maps étendues (ETM) [Lu and Feng, 2009] à partir de Topic Maps étendues locales (voir la description des ETM dans la section 3.2.3). Ces Topic Maps locales sont créées automatiquement à partir de documents textuels. L'approche proposée est un enchaîne-

ment de trois étapes. La première consiste en l'acquisition des éléments constituant la structure d'une ETM pour un document textuel. La seconde étape génère les ETMs associées aux documents. Elle est suivie de l'étape d'intégration permettant de générer l'ETM correspondant au contenu.

Dans [Weber et al., 2010b] une proposition de génération d'une Topic Map standard à partir de pages web est décrite. Cette génération de Topic Map standard se fait par application successive de deux types de transformations sur chacune des pages web. La première consiste à transformer les documents web en documents XML. Pour cela des services web, chargés d'extraire et de reconnâitre les données, sont utilisés. La seconde transformation permet de générer la Topic Map à partir de document XML. Une fois toutes les Topic Map générées, ces dernières sont fusionnées pour obtenir une Topic Map globale.

[Garrido et al., 2013] propose l'approche TM-Gen, un processus automatique qui génère à partir d'un contenu textuel monolingue, une Topic Map par document et qui par la suite procède à leur fusion en utilisant la méthode de fusion SIM citée dans [Lu et al., 2008a]. Pour la construction d'une Topic Map d'un document, TM-Gen effectue un traitement de texte permettant d'identifier les phrases clés candidates à être topics et effectue la simplification sémantique dans le cas où il existe des redondances, des incompatibilités ou des ambiguïtés dans les associations.

Dans la seconde sous-catégorie on a recensé, parmi les travaux de recherche les plus récents, les propositions de [Librelotto et al., 2004], de [Roy et al., 2013] et de [Eslami and Nazemi, 2013].

[Librelotto et al., 2004] proposent TM-builder, qui est une plateforme de génération de Topic Maps à partir de documents XML. La Topic Map générée est au format XML. TM Builder, tel qu'illustré dans la figure 3.10, comporte un extracteur de Topic Map XSTM fondé sur XML, et un processeur XSTM-P (XSTM-processor). Ce dernier génère des feuilles de style XSLT qui traitent les documents XML pour l'extraction des Topic Maps.

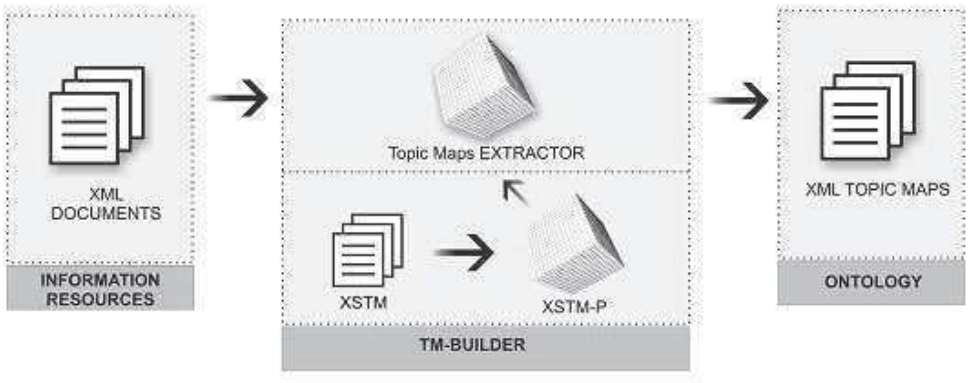


FIGURE 3.10 – Architecture de TM Builder [Librelotto et al., 2004]

[Roy et al., 2013] présentent une approche de transformation de schéma XML en Topic Map s'appuyant sur des règles de mise en correspondance d'éléments du XSD (XML schema) avec ceux du modèle de Topic Maps. La validation de leur processus de transformation a été effectuée sur la base de comparaisons des résultats de requêtes XQUERY sur les données XML avec ceux fournis par des requêtes Tolog sur la Topic Map correspondante.

L'approche de [Eslami and Nazemi, 2013] construit une Topic Map à partir d'un corpus de documents Wikipédia (XML), écrits en langue anglaise et ayant trait au domaine de l'informatique. Les auteurs de cette contribution ont, pour cela, proposé une approche composée de deux étapes. La première étape traite le corpus Wikipedia afin d'extraire pour chaque article Wikipedia son titre, les catégories assignées ainsi que les liens. Les textes des articles ne sont pas exploités. La seconde étape est une étape de construction de la Topic Map. Chaque concept (titre, catégorie et lien wikipedia) est représenté comme un topic. Les liens internes et les liens « see also » sont transformés en type d'association entre chaque article et ses catégories. Les liens entre catégories et sous-catégories sont de type « Part-of ». Un article est lié à une catégorie via le lien « assigned-to ». Enfin un titre est lié à un article via le lien « has a relation ».

La troisième catégorie comporte, parmi les travaux les plus récents, l'approche de [Eslami and Nazami, 2011], celle de [Weber et al., 2010a] ainsi que celle de [Jose-Garcia, 2012].

Dans [Eslami and Nazami, 2011], les constituants du schéma de la base de données relationnelle sont, avant tout, transformés en éléments du modèle de Topic Map. Chaque table et chaque colonne d'une table, hormis celle jouant le rôle de clé étrangère, devient un topic. Les clés étrangères se transforment en associations. Les données de la base de données constituent des ressources et sont liées aux topics de type « colonne » via des liens d'occurrences. La Topic Map résultat est visualisée via l'outil TM4L.

De la même façon que [Eslami and Nazami, 2011], les auteurs de [Jose-Garcia, 2012] définissent des règles de transformation du schéma de la base de données relationnelles en modèle de Topic Maps. Ils ont inclus les contraintes d'intégrité dans leur processus de transformation. Ces dernières sont traduites sous forme d'associations. Les validations syntaxiques et sémantiques de la Topic Map générée ont été effectuées en utilisant repectivement l'outil Ontopia Vizigator et des requêtes en TOLOG.

Pour permettre un accès par sujet aux données structurées, [Weber et al., 2010a] ont proposé de transformer les bases de données en Topic Maps. Pour cela, ils ont transformé les sources de données des bases de données sous forme de documents ASCII et HTML pour pouvoir, par la suite, les transformer en Topic Map à l'aide des services web. Ainsi, ils obtiennent des Topic Maps pour chacune des sources de données qui seront fusionnées, par la suite, pour donner lieu à une Topic Map globale. Dans le cadre de leur contribution, ils ont défini une ontologie de Topic Maps avec des types de topic, d'association, d'occurence et de rôle d'association.

### 3.3.2 Synthèse sur les approches de construction de Topic Maps

Les approches de construction de Topic Map présentées ci-dessus s'appuient soit sur le modèle standard soit sur un des modèles étendus présentés en section 3.2. Elles

peuvent être automatiques ou semi-automatiques. Il y a celles qui optent pour une construction préalable du schéma de la Topic Map puis procèdent au rattachement des ressources au schéma. Il y a celles qui choisissent de déduire le schéma sur la base du contenu. Dans le premier cas de figure, des topics peuvent être générés sans qu'ils soient rattachés à des ressources. L'avantage de telles approches est qu'elles assurent une meilleure stabilité de la Topic Map dans la mesure où tous les topics potentiels ont été extraits. Leur inconvénient est l'aboutissement, dans certaines situations, à des recherches infructueuses (le thème m'intéresse mais je ne peux pas recueillir des informations le concernant). Dans le second cas de figure, la Topic Map doit subir des enrichissements toutes les fois que le contenu s'enrichit de ressources informationnelles pour lesquelles des thématiques n'ont pas été répertoriées dans le schéma. Par conséquent, des opérations d'évolution de schéma de type enrichissement sont à prévoir. Cependant, tout thème mentionné dans la Topic Map est renseigné via les ressources informationnelles.

La plupart des approches qui construisent la Topic Map à partir du contenu optent pour un processus non incrémental de la Topic Map. Celles dont le processus est incrémental (construction itérative d'une Topic Map à partir de Topic Maps construites) sont réutilisables pour faire évoluer une Topic Map dont le contenu a subi des enrichissements.

Enfin, nous constatons que le problème de l'hétérogénéité des ressources informationnelles n'est pas totalement pris en charge par les approches existantes. En effet, aucune approche actuelle ne prend en charge l'hétérogénéité structurelle des contenus, c'est-à-dire des contenus combinant des ressources de types différents (documents structurés et/ou non structurés et/ou semi structurés et/ou documents multimédia).

De plus, pour des contenus multilingues, seule la méthode de Ellouze et al. permet de les organiser de telle sorte que toutes les ressources (toutes langues confondues) puissent être ciblées en une seule fois via un topic multilingue. Cependant, aucune

des méthodes proposées n'offre une recherche précise pour des contenus pluridisciplinaires où la variation du sens des termes et des concepts est inévitable.

Les tableaux ci-après résume nos propos.

| Critères / Approches | Source nécessaire | Type de contenu | Prise en compte de la variation de sens | Type d'approches | Degré d'automatisation | modèle sous-jacent | Prise en compte du multilinguisme |
|---|---|---|---|---|---|---|---|
| [Garrido and Harri, 2014] | Aucune | Documents textuels | Non | Extraction d'information<br>Technique de fusion | Semi-automatique | TM standard | Non |
| [Roy et al., 2013] | Aucune | Document XML | Non | Technique de transformation de schéma | Semi-automatique | TM standard | Non |
| [Eslami and Nazami, 2013] | Aucune | Document Wikipédia (XML) | Non | Technique de transformation de schéma | Automatique | TM standard | Non |
| [Gomoi et al., 2012] | Aucune | vMR (virtual Medical Records) | Non | Technique de transformation de schéma | Manuel | TM standard | Non |
| [Goldhammer, 2012] | Aucune | Documents CAD | Non | Non mentionné | Manuel | TM standard | Non |
| [Jose-Garcia, 2012] | Aucune | Base de données | Non | Technique de transformation de schéma | Automatique | TM standard | Non |
| [Bernotaityte et al., 2012] | Aucune | Aucun | Non | From scratch | Semi-automatique | TM standard | Non |
| [Ellouze et al., 2012] | Thésaurus<br>Ontologie | Documents textuels | Non | Technique Incrémentale<br>Traitement de texte<br>Technique de fusion | Semi-automatique | Modèle de CITOM | Oui |
| [Damen, 2012] | Aucune | Concepts | Non | Orientée schéma | Automatique | TM standard | Non |
| [Eslami et Nazami, 2011] | Aucune | Base de données | Non | Technique de transformation de schéma | Automatique | TM standard | Non |
| [Arndt et al., 2011] | Aucune | Documents textuels | Non | Non | Manuel | TM standard | Non |
| [Dragu et al., 2011] | Aucune | Documents textuels | Non | Technique de transformation de schéma | Manuel | TM standard | Non |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| [Weber et al., 2010a] [Weber et al., 2010b] | Ontologie de domaine | Base de données Documents textuels | Non | Technique de transformation Technique de fusion | Automatique | TM standard | Non |
| [Bold et al., 2010] | Aucune | Thésaurus | Non | Technique de transformation | Semi-automatique | TM standard | Non |
| Zheng et al., 2009] | Aucune | Documents textuels | Non | Intégration de schéma | Automatique | Modèle ETM | Non |
| [Jung et al., 2008] | Aucune | Documents textuels | Non | Non connu | Semi-automatique | TM standard | Non |
| [Lu et al. 2008a] | Aucune | Documents | Non | Technique de Fusion | Automatique | Modèle ETM | Non |
| [Zaher et al., 2007] | Aucune | Pages web | Non | Technique de transformation de schéma Technique de fusion | Semi-automatique | Modèle Hypertopic | Non |
| [Wang and Guo, 2007] | HowNet | Corpus documentaire | Non | Orientée contenu | Semi-automatique | TM standard | Non |
| [Kim et al., 2007] | Aucune | TM | Non | Technique de Fusion | Automatique | TM standard | Non |
| [Eric Prudfhommeaux, 2002] | Aucune | Méta-données RDF | Non | Technique de transformation de schéma | Automatique | TM standard | Non |
| [Grønmo, 2002] | Aucune | Base de données | Non | Technique de transformation de schéma Technique de fusion | Automatique | TM standard | Non |
| [Maicher and Witschel, 2004] | Aucune | TM Distribuées | Non | Technique de Fusion | Semi-automatique | TM standard | Non |
| [Kasler et al., 2006] | Aucune | Documents textuel bilingues | Non | Techniques dapprentissage Techniques d'extraction d'information | Semi-automatique | TM standard | Partiel (bilinguisme) |
| [Librelotto et al., 2004] | Aucune | Documents XML | Non | Technique d'extraction de feuille de style | Automatique | TM standard | Non |
| [Lacher and Decker, 2001] | Aucune | Méta-données RDF | Non | Technique de transformation de schéma | Automatique | TM standard | Non |

FIGURE 3.11 – Tableau comparatif des différentes approches de construction de Topic Maps

## 3.4 Approches d'évolution de Topic Maps

En règle générale, la gestion de l'évolution de ressources sémantiques est une exigence de base pour une utilisation durable de ces ressources. Cette règle s'applique aux Topic Maps. Cependant, si la gestion des évolutions d'ontologies de domaines a fait l'objet de plusieurs travaux de recherche [Khattak et al., 2009], [Zablith et al., 2013] et [Flouris et al., 2008] que cela soit sur les aspects processus, sur les approches de prise en charge des changements et sur la gestion des versions, l'évolution des Topic Maps demeure très peu traitée dans la littérature. De plus, la littérature en question, s'est beaucoup plus focalisée sur les aspects liés à la fusion des Topic Maps.

La fusion des Topic Maps a été traitée à travers la proposition d'approches et d'algorithmes de fusion. Pour ce qui est des approches de fusion, celles-ci ont été, le plus souvent, abordées via des approches de création de Topic Maps [Ellouze et al., 2012] ou encore de fusion de Topic Maps dans des contextes distribués [Zheng et al., 2009], [Lu et al., 2008a] et [Xue et al., 2010]. Toutes ces méthodes de fusion sont intégrées dans des méthodes de construction de Topic Maps citées dans la section précédente. Elles utilisent des modèles étendus.

Des algorithmes de fusion ont été proposés pour pallier les insuffisances du mécanisme de fusion proposé dans le standard TMDM [Lu and Feng, 2009], [Kim et al., 2007] et [Maicher and Witschel, 2004]. En effet, ce dernier ne considère que les fusions d'éléments de Topic Map équivalents et, par conséquent, ne prend pas en compte le fait que les Topic Maps ne partagent pas le même vocabulaire, ce qui peut être le cas dans des environnements distribués. Cette situation est aussi valable lorsque la fusion concerne deux Topic Maps ne partageant pas les mêmes langues. Ces algorithmes sont fondés sur des mécanismes de détection de conflits et sur des stratégies de résolution de conflits reposant sur des mesures de similarités entre éléments de Topic Maps telles que celle proposée dans [Maicher and Witschel, 2004]. À titre d'exemple [Lu and Feng, 2009] présente un mécanisme de détection et de réso-

lution de conflits. Ce dernier est fondée sur des mesures de similarité définies dans [Maicher and Witschel, 2004]. Dans [Kim et al., 2007] une taxonomie des conflits a été proposée pour assister l'algorithme de fusion.

Notons, qu'hormis la méthode de fusion proposée dans CITOM [Ellouze et al., 2012], aucune approche ne prend en compte le multilinguisme. Aussi, un autre avantage de cette méthode est qu'elle intègre dans son processus de création de Topic Maps une autre opération d'évolution de Topic Map qui est la modification de la Topic Map par le biais de l'enrichissement du contenu qu'elle organise. Cependant, la technique associée à cette opération oblige la reconstruction de toute la Topic Map, ce qui du point de vue de la performance, n'est pas l'idéal.

Enfin, soulignons le fait qu'aucune proposition d'algorithme de fusion, n'intègre, pour assurer une fusion sémantique de Topic Maps, l'aspect variation du sens des termes ou concepts représentés par les topics.

## 3.5 Outils d'édition de Topic Map

L'objectif de n'importe quel outil d'édition de Topic Maps est la création, la modification ainsi que la visualisation de ses différents constituants. Quantités d'éditeurs et prototypes ont été développés dans ce sens. Les prototypes sont liés aux méthodes de construction citées ci-dessus. La plupart ne sont accessibles qu'à travers la présentation faite dans les publications relatives aux méthodes. Les éditeurs accessibles via les URLs les présentant, n'offrent, en réalité, aucune aide à la construction de Topic Maps. Parmi ces éditeurs, on peut citer la suite OKS (Ontopia Knowledge Suite) d'ontopia [Team, 2014] dotée de l'outil Omnigator/Vizigator pour la visualisation et la navigation et de l'outil Ontopoly pour la création de Topic Maps, TM4J [tm4j, 2006], Topic Map Designer [TMdesigner, 2006], TM4L [Ditcheva and Dicheva, 2007] et enfin Wandora [Team, 2014]. Ces éditeurs reposent sur le modele TMDM. Ils proposent la plupart du temps deux formes de visualisation : graphique et orientée classification. Un état de l'art détaillé et critique sur les

outils de visualisation de Topic Maps est présenté dans l'annexe A de cette thèse.

## 3.6 Conclusion

Après un rappel de la technologie des Topic Maps, nous avons dans ce chapitre dressé un état des lieux de la recherche associée à cette technologie.

Nous avons montré dans la section 3.1.4 que la portée des Topic Maps a dépassé strictement le cadre d'une utilisation humaine pour la recherche d'information. C'est aussi un outil utilisé par des applications de domaines divers et variés, notamment pour la prise de décision.

Nous avons aussi montré dans nos analyses des différents états de l'art présentés dans ce chapitre, que les problématiques liées à l'organisation de contenus multilingues sont toujours à l'ordre du jour et celles liées aux contenus pluridisciplinaires ont été omises par les chercheurs travaillant sur la construction, la fusion et la visualisation de Topic Maps.

Le chapitre 5 de cette thèse, réunit nos efforts dans la prise en charge de la variabilité des sens sous-jacents aux contenus multilingues et pluridisciplinaires. Il décrit une boîte à outils constituée de composants élémentaires qui, lorsqu'ils sont combinés, délivrent diverses approches de construction de Topic Maps, dites contextualisées, servant à l'organisation de contenus pluridisciplinaires et multilingues. Ce chapitre présente aussi un nouveau modèle de Topic Map étendu.

# Chapitre 4

# Conception et mise en œuvre de la structure du wiktionnaire des Sciences Humaines et Sociales (SHS)

Selon les définitions simplifiées des dictionnaires, les sciences humaines ont pour objet d'étude tout ce qui concerne les hommes, leurs histoires, leurs cultures, leurs réalisations et leurs comportements individuels et sociaux. Les sciences sociales, quant à elles, ont pour objet d'étude les sociétés humaines. Ces deux sciences regroupées sous le sigle SHS (Science Humaines et Sociales) rassemblent de ce fait plusieurs champs disciplinaires hétérogènes tels que la sociologie, l'économie, l'ethnologie, l'anthropologie, la psychologie, l'histoire, la géographie, la démographie, les sciences politiques, l'archéologie, la linguistique, les sciences administratives et les sciences de religion. Elles jouent un rôle primordial dans la compréhension et l'interprétation du contexte économique, culturel et social des populations. L'évolution de la recherche dans ce domaine passe inévitablement par l'échange et le partage des

connaissances entre les chercheurs.

Afin de promouvoir les échanges entre les pays du Maghreb et la France dans le domaine des sciences sociales et humaines, un projet de construction d'un contenu multilingue et multiculturel a été lancé par le FMSH[1] en collaboration avec des partenaires des pays du Maghreb et de la France[2]. Une fois réalisé, ce projet permettra de développer les échanges entre chercheurs maghrébins et leurs partenaires français et mettra en commun un ensemble de savoirs sur les deux cultures et les deux sociétés.

Dans ce projet, il était question, dans un premier temps, de construire un dictionnaire en ligne franco-maghrébin des SHS, inexistant à ce moment là. Dans un second temps, il était prévu de concevoir un système d'organisation de contenu qui prendrait en charge les contraintes multiculturelles et multilingues des SHS et qui permettrait aux chercheurs des deux rives de partager des connaissances associées au domaine des SHS.

Ce chapitre décrit la conception de la structure de ce dictionnaire électronique ainsi que sa mise en œuvre. La structure du système d'organisation de contenu proposé, en l'occurrence le modèle de Topic Map étendu fait l'objet du chapitre suivant. Dans ce même chapitre est présenté l'approche de peuplement d'une Topic Map SHS. Cette approche fait partie d'un ensemble d'approches de construction et d'évolution de Topic Map.

La structure de ce chapitre est calquée sur la démarche méthodologique adoptée dans le cadre de ce projet. Cette démarche comporte cinq étapes (voir figure 4.1) .

1. Un des acronyme de La Fondation Maison des sciences de l'homme (FMSH), http ://www.msh-paris.fr/

2. Les partenaires sont : Le FMSH, Le Cnam de Paris, et l'INI (Institut National d'Informatique) d'Alger.

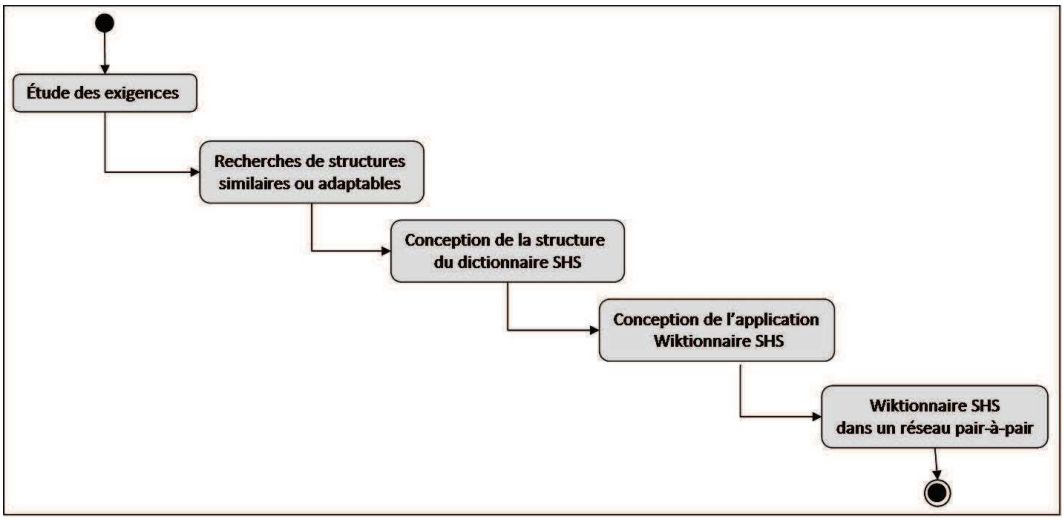


FIGURE 4.1 – Démarche adoptée dans la réalisation du dictionnaire des SHS

La première étape a consisté à prendre connaissance, à travers le cahier des charges du projet, des exigences associées à ce dictionnaire électronique. Nous avons, par la suite, effectué une recherche de structures de dictionnaire similaires à celle exigée par le projet ou encore adaptables à notre problématique. Les résultats de cette étape, nous ont amené, dans la phase suivante du projet, à une proposition de structure du dictionnaire SHS suivie du développement d'une application permettant son peuplement, son évolution et sa visualisation. Elle a été nommée wiktionnaire. Cette même application a été déployée sur un réseau pair-à-pair.

La section 1 de ce chapitre est relative à l'exposé des contraintes énoncées dans le Projet FSP-Maghreb quant à la structure et la mise en œuvre du dictionnaire. La seconde section 2 rend compte de notre recherche de dictionnaires similaires ou adaptables. La section 3 présente la structure du dictionnaire SHS. La section 4 est dédiée à la description de l'application Wiktionnaire SHS permettant d'alimenter, de visualiser et de faire évoluer le dictionnaire SHS. La section 5 concerne l'intégration de l'application dans un environnement pair-à-pair. La section 6 conclut ce chapitre.

## 4.1 Contraintes de structure et de mise en œuvre du dictionnaire électronique des SHS

Le projet de développement du dictionnaire électronique SHS a été initié dans l'objectif de renforcer l'échange et le partage des connaissances, entre les chercheurs des deux rives de la Méditerranée, dans le domaine des sciences humaines et sociales (SHS) et ce quels que soient leurs lieux géographiques de travail et/ou de résidence. Ce dictionnaire devrait, à court terme, pouvoir contenir les principaux termes SHS utilisés en France et dans les pays du Maghreb, préciser leurs usages par les deux sociétés et fournir leur traduction d'une langue à une autre. À long terme, il devrait englober les différentes langues du bassin méditerranéen. La conception de ce dictionnaire doit, conformément au cahier des charges du projet, prendre en compte les faits suivant :

- une entrée $A_k$ dans une langue source peut avoir plusieurs sens et donc plusieurs traductions $B_1$,... $B_j$, ... $B_m$ dans la langue cible. Cette même entrée $A_k$ peut être définie avec plusieurs éléments $A_1$,..., $A_i$, ... $A_n$ du schéma du dictionnaire (synonyme, antonyme, étymologie, expressions figées, hyperonyme, hyponyme, etc.). Ces éléments peuvent être, à leur tour, des entrées dans la même langue source. Par conséquent, ils peuvent avoir plusieurs sens dans cette même langue source et plusieurs traductions dans la langue cible (voir figure 4.2). Notons, à cet effet que, selon le sens de la traduction, une langue source peut aussi devenir cible et qu'une entrée dans une langue source peut ne pas avoir d'équivalent dans une langue cible.
- La signification attribuée à une entrée du dictionnaire SHS dépend du contexte de définition de cette entrée. Ce dernier est décrit par un ensemble fini et connu de paramètres contextuels qui varient d'une discipline à une autre. Parmi ces paramètres on peut citer les paramètres temporels et géographiques.
- l'ensemble des éléments servant à décrire une entrée doit faire partie de la

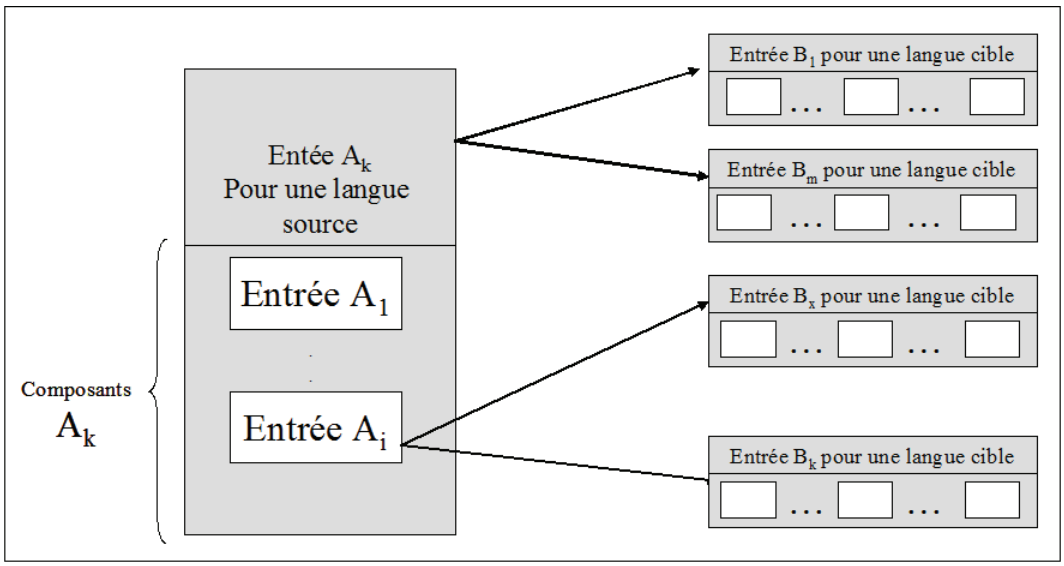


FIGURE 4.2 – Un extrait du schéma du dictionnaire des SHS

norme ISO 1951 [ISO1951, 2007]. Cette norme a été choisie de par sa structure générique indépendante des supports de publication.

Enfin, la réalisation de ce dictionnaire et de l'application permettant de le manipuler ont été, dans le cahier des charges, conditionnées par l'exploitation de la technologie Wiki pour les avantages qu'elle offre, dont la facilité dans la construction et la maintenance collaborative d'un contenu par des non informaticiens.

## 4.2 Recherche de structures de dictionnaires similaires ou adaptables

Une multitude de dictionnaires dédiés aux SHS existent. À titre d'exemple, nous pouvons citer les dictionnaires suivants :

- Dictionnaire des races, éthnie et culture [Bolaffi, 2003].
- Dictionnaire le referis : droit des affaires, comptable, gestion financière, fiscal, droit communautaire, sociale, gestion budgétaire [Paul-Jacques and Patrice, 1995].
- Dictionnaire multi-culturel de psychologie : problèmes, termes et concepts

[Hall, 2004].

– Dictionnaire des mots clés dans les interventions multiculturelles [Mio, 1999].

Le premier dictionnaire est pluridisciplinaire et multilingue. Il existe sous le format papier et sous le format PDF. Ses entrées sont accompagnées de références bibliographiques en langue anglaise. Il fournit des traductions d'entrées en langue italienne, allemande et française. Chaque définition associée à une entrée est datée.

Le second dictionnaire est aussi pluridisciplinaire mais monolingue. Ses entrées sont en langue française. Elles peuvent comporter une définition pour chaque discipline représentée dans ce dictionnaire (comptabilité, droit des affaires, gestion financière, social, gestion budgétaire, etc.).

Le dictionnaire multiculturel de psychologie est en langue anglaise. Chaque entrée dispose d'une ou de plusieurs définitions contextualisées (date et auteur de la définition).

Le dernier dictionnaire est un dictionnaire en langue anglaise, spécialisé dans le domaine de la psychothérapie. Il est disponible aussi bien sous le format papier qu'en PDF. Les définitions associées aux entrées sont référencées.

Ces dictionnaires couvrent partiellement le domaine des SHS. Ils sont tous non conformes à la norme ISO. Certains sont multilingues et pour la plupart, en format papier.

Il existe aussi, dans la littérature, plusieurs projets de construction de dictionnaires spécialisés. Parmi ces projets, on peut citer le projet PAPILLON [Mangeot, 2006], le projet DHYDRO [Descotte, 1999], le projet JMdict/EDICT [Bond, 2007] et enfin le projet SAIKAM [Ampornaramveth and Aizawa, 2001]. PAPILLON utilise le paradigme de construction collaborative de Linux pour l'édition collaborative de définition. Il offre, parmi les critères de recherche possibles, la restitution d'un terme à partir de sa lecture contextuelle. Dans le projet DHYDRO, un espace terminologique multilingue spécialisé dans le domaine de l'hydrographie a été construit. JMdict/EDICT propose un outil d'édition, à distance, d'une base terminologique

multilingue. SAIKAM est un dictionnaire en ligne dédié à la création de nouveau termes Thai à partir de termes japonais.

Notre exploration de la littérature nous a fait constater l'absence de projets d'élaboration de ressources sémantiques (dictionnaires ou autres) dédiées aux SHS ni même de projets pour lesquels la structure de la ressource sémantique pourrait être exploitée pour la mise en œuvre du dictionnaire exigé par le projet.

Ceci nous a amené à explorer la possibilité d'exploiter le Wiktionnaire actuel hébergé par la fondation WIKIMEDIA. Ce Wiktionnaire est une structure de dictionnaire ouvert, universel et fondé sur la technologie Wiki. Il permet, à des personnes autorisées, d'éditer, de publier facilement et rapidement des contenus en ligne et de les faire évoluer via des processus de travail collaboratif. Il offre aussi une gestion complète des versions, une gestion des historiques des contenus et enfin une gestion des notifications permettant à des personnes intéressées par des thèmes particuliers d'êtres alertées à chaque création, modification ou suppression de contenus en rapport avec leurs thématiques favorites. Il est structuré en articles [wikipédia, 2002]. Chaque article sert à décrire un terme et regroupe :

- une section principale dont l'objet est de décrire le terme dans la langue associée au Wiktionnaire (exemple : section de langue française pour un Wiktionnaire en langue française),
- zéro ou plusieurs sections de langues autres que celle de la langue du Wiktionnaire,
- une section catégorie permettant de classer le terme dans une ou plusieurs catégories parmi celles répertoriées,
- et enfin une section de liens interwikis permettant de faire des liens vers le même article dans les autres Wiktionnaires. Ces liens se font vers les articles ayant exactement le même titre que l'article, et non vers ses traductions.

La section principale propose :

- un ensemble obligatoire d'éléments de description de base : étymologie, une

ou plusieurs sections pour le type de mot (c'est-à-dire ses variations orthographiques, ses abréviations, le ou les termes dérivés, ses synonymes, ses antonymes, ses hyponymes, ses holonymes, ses méronymes, ses traductions, etc)

– et un ensemble d'éléments optionnels : la ou les prononciations, la ou les anagrammes, une section « à voir aussi » qui regroupe les liens en rapport avec le terme de l'article et une section « référence » permettant de donner les références utilisées lors de la rédaction de l'article.

Une section de langue autre que celle du Wiktionnaire est similaire à la section principale. Toutefois elle ne possède ni de section « Traduction », ni de sections « Hyperonymes », « Hyponymes », « Holonymes » et « Méronymes ».

Ce Wiktionnaire actuel ne répond pas aux spécificités du dictionnaire des SHS. D'une part, il ne dispose pas de système automatique de gestion des correspondances qui permettrait de gérer la complexité des renvois entre la langue source et la langue cible. Il est possible, à l'aide du Wiktionnaire actuel, de faire évoluer une entrée indépendamment des autres entrées auxquelles elle est liée. En d'autres termes, il est possible d'ajouter, dans un Wiktionnaire dédié à une langue A, une traduction d'un terme vers une langue B sans qu'il y ait répercussion de ce changement dans le Wiktionnaire dédié à la langue B. De plus, les liens interwikis ne peuvent s'établir qu'entre articles de même nom. Cela signifie qu'on ne pourra pas lier deux termes dont l'un est la traduction de l'autre si ces deux termes sont dans des Wikis différents. D'autre part, le schéma du Wiktionnaire actuel ne permet pas une recherche, par contexte, de la signification d'un terme. Cette fonctionnalité s'avère très importante dans le domaine des SHS.

Une autre version du Wiktionnaire existe. Il s'agit de OmegaWiki [OmegaWiki, 2009]. Il est basé sur une extension du MediaWiki. OmegaWiki, contrairement au Wiktionnaire actuel, réunit dans un même espace tous les Wiktionnaires des différentes langues. Cela permet de pallier l'inconvénient du Wiktionnaire actuel concernant la non répercussion des changements d'un Wiktionnaire d'une

langue sur celui d'une autre langue. En plus, du fait qu'OmegaWiki soit en lecture seulement, il conserve la structure du Wiktionnaire actuel et ne permet donc pas une recherche de termes par contexte.

En conclusion de cette section, nous pouvons constater l'absence de structure de dictionnaire similaire à celle imposée dans le cahier des charges ou de structure que l'on pourrait adapter pour satisfaire les exigences du cahier des charges. Ceci explique la poursuite du projet vers une proposition de mise en œuvre d'une structure que nous détaillons dans la section suivante.

## 4.3 Modèle conceptuel du dictionnaire SHS

Comme il est indiqué dans la section 4.1, une entrée du dictionnaire SHS peut avoir plusieurs descriptions. Chacune d'elles est applicable à un contexte donné décrit par un ensemble de paramètres de contexte dépendant d'une discipline. De plus, chacune de ces descriptions doit être conforme à la norme ISO 1951. Par conséquent, la conception de la structure de notre dictionnaire doit reposer sur des correspondances entre les éléments de départ (entrées) et leurs contextes de définition dans la langue source et les éléments d'arrivée (entrées) et leurs contextes de définition dans la/les langue(s) cible(s) selon un schéma qui pourrait contenir les éléments de la norme ISO 1951 suivants : définition, antonyme, synonymes, termes associés, informations orthographiques, prononciation, etc.

Avant de présenter le modèle conceptuel du dictionnaire SHS proposé pour répondre au cahier des charges du projet FSP-Magreb, commençons par préciser la notion de contexte associée aux descriptions des entrées du wiktionnaire projeté.

### 4.3.1 La définition du contexte associé au dictionnaire SHS

Pour [Brézillon, 1999] et [Svensson, 2009], la définition du terme contexte bien que présent dans plusieurs domaines, ne fait pas l'objet d'un consensus. À titre d'exemple :

- [Dey, 2001] définit le contexte, dans le cadre des applications sensibles au contexte, comme étant « n'importe quelle information pouvant être utilisée pour caractériser la situation d'une entité c'est à dire la situation d'une personne, d'un endroit, d'un objet qui peut être pertinent pour l'interaction entre un utilisateur et une application ».
- Dans le domaine de la reconnaissance des formes, [Brémond and Thonnat, 1997] proposent une définition du contexte à travers la description de différentes sortes d'informations manipulées par un processus.
- Selon [Bastien, 1992], le contexte, dans le domaine des ontologies, permet de définir les connaissances qui doivent être considérées, leurs conditions d'activation, leurs limites de validité et leur utilisation à un moment donné. Selon [Brézillon, 1999], le rôle du contexte dans ce même domaine est de fournir aux humains un meilleur contrôle de la connaissance.

Dans le cadre de ce travail, la notion de contexte est associée aux termes et aux concepts utilisés dans le domaine des SHS afin de préciser leurs sens. En effet, un terme ou concept dans le domaine des SHS, conformément au cahier des charges du projet, dispose de plusieurs définitions. Chaque définition décrit les différents sens du terme dans une discipline donnée des SHS. Un sens est lié à l'usage de ce terme ou concept durant une période donnée et pour des lieux géographiques précis. Ces éléments ou paramètres (discipline, période et lieu géographique), permettant de préciser la validité du sens d'un terme ou d'un concept, constituent le contexte d'une définition.

Notons que la variation du sens des termes n'est pas une problématique propre aux SHS. Elle concerne plusieurs autres domaines. Selon [Dury, 2000] de nombreuses sciences, pour ne pas dire toutes, empruntent des concepts à d'autres sciences, connexes ou non. L'auteur cite un certain nombre d'exemples dont le terme biosphère qui a d'abord été utilisé en biologie pour désigner un atome globuleux d'une existence hypothétique et qui serait la base unique de tous les corps organisés. Ce

terme a ensuite été utilisé au moyen d'un transfert sémantique, dans le domaine de la biogéographie pour désigner l'une des couches géochimiques de la sphère terrestre, constituée par la masse organique des êtres vivants. Enfin, au cours d'un second transfert, l'écologie a repris biosphère à son compte et l'utilise aujourd'hui pour décrire « la région de la planète qui renferme l'ensemble des êtres vivants et dans laquelle la vie est possible en permanence ». Ceci est aussi le cas du terme « Mercure » qui désigne un Dieu grec dans la mythologie romaine ou un élément chimique dans le domaine de la physique ou de la chimie, ou encore une planète dans le domaine de l'astronomie.

Les termes au sein d'une même discipline peuvent :

- Subir des transformations de sens dues à des causes diverses tels que les innovations technologiques et scientifiques ou encore le développement de la vie sociale, économique et culturelle [Dury, 2000]. Tel est le cas, par exemple du terme « chômer » qui, initialement voulait dire « ne pas travailler pendant la chaleur », a pris son sens actuel avec le développement de la société capitaliste et avec l'apparition des sans travail [Dury, 2000].
- accumuler des sens différents au fil de leur évolution. Citons comme exemple, le terme « Amérasien » définit dans le domaine de la psychologie (discipline des SHS) et qui faisait référence, après la deuxième guerre mondiale, à la descendance des couples dont la conjointe est japonaise et le conjoint un militaire américain. Ce même terme a subi un changement de sens après la guerre du Vietnam en 1975. Il désigne actuellement les personnes ayant une mère vietnamienne et un père américain.

Les termes peuvent aussi avoir un sens qui varie selon le lieu géographique d'usage. Dans ces mêmes lieux, ils peuvent avoir subi des variations de sens. C'est le cas du terme « Entrepreneur » dont le sens a changé d'une discipline à une autre, d'une période historique à une autre et d'un lieu géographique à un autre. C'est aussi le cas du terme « indien » qui peut designer soit un autochtone d'Amérique ou encore

un habitant de l'Inde.

Certains termes peuvent disparaître après avoir accumuler des sens. C'est le cas par exemple du terme « manufacture » qui est apparu au 16ième siècle. Au 17ième siècle, il désignait un établissement industriel. De nos jours il sort de l'usage.

Enfin, notons aussi le fait que des termes d'une langue ne puisse pas avoir d'équivalent dans une autre langue. C'est le cas, par exemple, du terme « amae ». Ce terme d'origine japonaise fait référence à un sentiment plaisant d'attachement ou de dépendance dans la relation mère-enfant. Selon [Doi and Bester, 1977], ce terme n'a pas d'équivalent dans les langues européennes.

Pour conclure cette section, nous avons, conjointement avec le FMSH, proposé une description du contexte d'un terme SHS. Ce contexte consiste en un ensemble de paramètres (méta-données) dont le paramètre « langue » et le paramètre discipline. Par conséquent, à une description d'un terme SHS donné sera associé un contexte de validité décrit à travers un ensemble de valeurs. Les deux premières valeurs correspondent aux deux paramètres invariants (la langue et la discipline). Les autres valeurs sont associées aux paramètres propres à la discipline associée à cette description. Ainsi, pour exprimer, par exemple, le fait que le terme « Amérasien » a les deux sens énoncés ci-dessus. On associera à l'entrée représentant ce terme deux descriptions. Chacune d'elle aura un contexte où la discipline, la langue et la période de validité ont été renseignées.

### 4.3.2 Le modèle conceptuel du dictionnaire électronique SHS

La description conceptuelle du dictionnaire SHS pourrait être représentée à l'aide d'un modèle de classes UML.

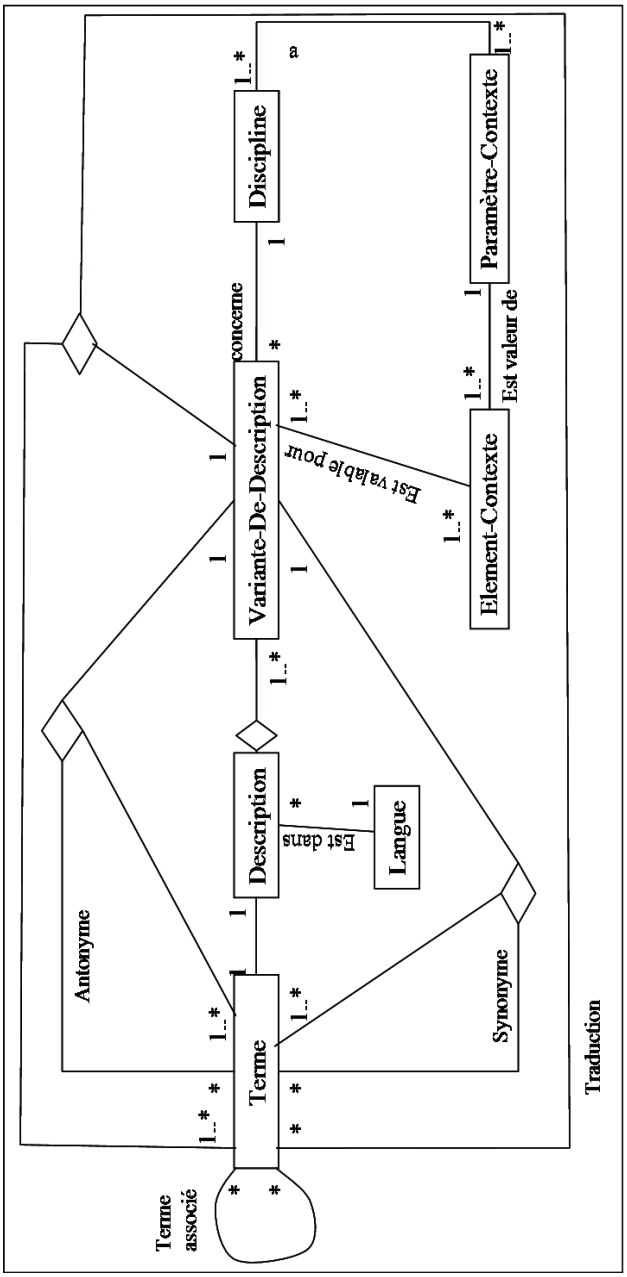


FIGURE 4.3 – Un extrait du modèle conceptuel du dictionnaire des SHS

La figure 4.3 présente un extrait de ce schéma conceptuel. Ce modèle montre qu'une description d'une entrée (terme) dans une langue donnée est construite par union des variantes de cette description. À chaque variante correspond un contexte défini par la discipline concernée et un ensemble d'éléments de contexte nommés « valeurs des paramètres de contexte ». Chaque discipline a ses propres paramètres de contexte. Chaque entrée décrite à l'aide d'une variante de description donnée peut avoir un synonyme associé à cette variante.

## 4.4 Présentation de l'application Wiktionnaire SHS

L'utilisation de la technologie Wiki étant une contrainte de mise en œuvre du dictionnaire SHS, nous avons procédé à un état de l'art nous permettant de choisir le type de wiki adéquat.

### 4.4.1 L'état de l'art sur les wikis

Nous avons recensé, plusieurs Wikis. WikiNi, Wiclear, DokuWiki, MediaWiki et les Wikis sémantiques en sont des exemples.

Les Wikis sémantiques tels que celui de [Krötzsch et al., 2006], KawaWiki [Kawamoto et al., 2006], IkeWiki [Schaffert et al., 2006], SweetWiki [Buffa et al., 2008], Kaukolu [Kiesel et al., 2006], KnowWE [Baumeister et al., 2011], 959NG [Kumar et al., 2012], AceWiki-GF [Fuchs et al., 2013], et RiceWiki [Zhang et al., 2014], sont des applications du web sémantique aux Wikis.

KawaWiki permet la création de pages Wikis, en utilisant des modèles en RDF, ainsi que l'interrogation à l'aide du langage SPARQL.

IkeWiki est un outil pour une construction formalisée et collaborative de contenus. Il offre des possibilités d'annotation de liens et de raisonnement.

SweetWiki annote sémantiquement les ressources d'un Wiki. Il supporte le tagging social et utilise des ontologies pour décrire le domaine et la structure du Wiki.

Il dispose aussi d'un éditeur WYSIWYG.

Kaukolu est un Wiki sémantique fondé sur JSPWiki. Il permet l'annotation, la création et l'édition de pages Wiki. Pour favoriser la création de nouvelles pages, il transforme les URIs en alias.

KnowWE renforce les architectures des Wikis sémantiques par une ontologie des tâches et un processus de raisonnement.

959NG est un wiki fondé sur le Mediawiki sémantique, pour éditer la taxonomie de phylum Nematoda[3] (sorte de vers) .

RiceWiki se base également sur le Mediawiki sémantique. Il permet d'éditer les différents gènes du riz.

AceWiki-GF est une extension de Acewiki. Il permet d'écrire un article en une langue spécifique et de le rendre disponible immédiatement en d'autres langues avec des traductions précises.

Le Mediawiki sémantique est une extension du MediaWiki. Il hérite des avantages du Mediawiki telles que la facilité d'édition de documents collaboratifs (minimum de pré-requis techniques) et l'évolutivité. Il permet aussi d'annoter les pages Wikis, leurs contenus et les liens entre elles et cela à l'aide de méta-données compréhensibles par une machine. De plus, pour des objectifs de navigation, les Mediawikis sémantiques et les Wikis sémantiques en général, à travers l'utilisation intensive des hyperliens, donne la possibilité, à un futur utilisateur d'avoir une vue globale sur une page et de « zoomer », en cas de besoin, sur une partie de son contenu.

Notre étude de l'état de l'art et sa confrontation avec les spécificités de notre Wiktionnaire des SHS, nous a permis de retenir, la technologie du mediaWiki sémantique pour mettre en œuvre l'application permettant la création, la manipulation et la visualisation du dictionnaire SHS.

3. Nematode. http ://en.wikipedia.org/wiki/Nematode

### 4.4.2 La conception et la mise œuvre de l'application Wiktionnaire SHS

Pour la mise en œuvre de l'application permettant de créer, manipuler et visualiser le dictionnaire SHS projeté, nous avons dans un premier temps effectué une rétro-conception du Mediawiki Sémantique pour pouvoir déduire son méta-modèle (voir figure 4.4) et procéder à la transformation du modèle conceptuel du dictionnaire en Wiktionnaire. Dans un second temps, nous avons réalisé l'application en question que nous avons nommée application Wiktionnaire SHS.

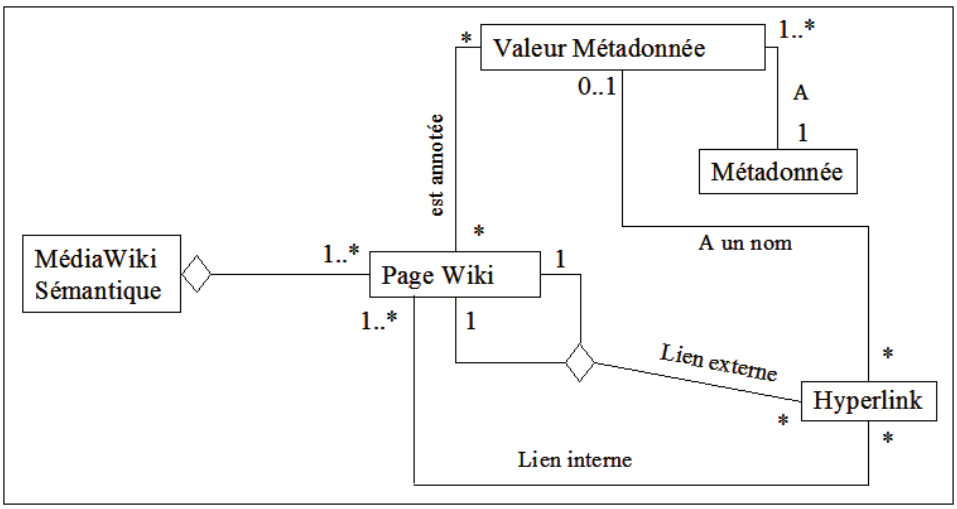


FIGURE 4.4 – Le métamodèle du Mediawiki sémantique.

Comme le montre la figure 4.4, un Mediawiki Sémantique est un ensemble de pages Wikis que l'on peut annoter. Une page Wiki peut être reliée à une autre page Wiki à travers des hyperliens externes. Les hyperliens peuvent aussi être utilisés à l'intérieur d'une page. Les hyperliens peuvent aussi être annotés.

Les correspondances effectuées entre les concepts du MediaWiki sémantique et ceux de notre dictionnaire SHS (voir figure 4.3) sont représentées dans le tableau 4.1. Comme l'indique de tableau, les différentes descriptions d'une entrée du dictionnaire seront représentées sous forme de pages Wikis. Les paramètres de contexte

| **Concepts du Wiktionnaire des SHS** | **Concepts du Mediawiki Sémantique** |
|---|---|
| Description | Page Wiki |
| Variante de description | Page Wiki |
| Elément de contexte | Valeur de la méta-donnée du paramètre de contexte |
| Langue | Méta-donnée |
| Discipline | Méta-donnée |
| Paramètre de contexte | Méta-donnée |
| Antonyme | Hyperlien |
| Terme associé | Hyperlien |
| Synonyme | Hyperlien |
| Traduction | Hyperlien |

TABLE 4.1 – Correspondance entre les concepts du dictionnaire SHS et ceux du Mediawiki sémantique

sont considérés comme des méta-données du Mediawiki sémantique. Un élément du contexte est une valeur d'une méta-donnée dans le Mediawiki sémantique. Tous les autres concepts (antonymes, termes associés, synonymes, traductions, etc.) sont transformés en des liens Wikis.

L'exemple de la figure 4.5 illustre la structure de notre dictionnaire transposé dans un Mediawiki Sémantique.

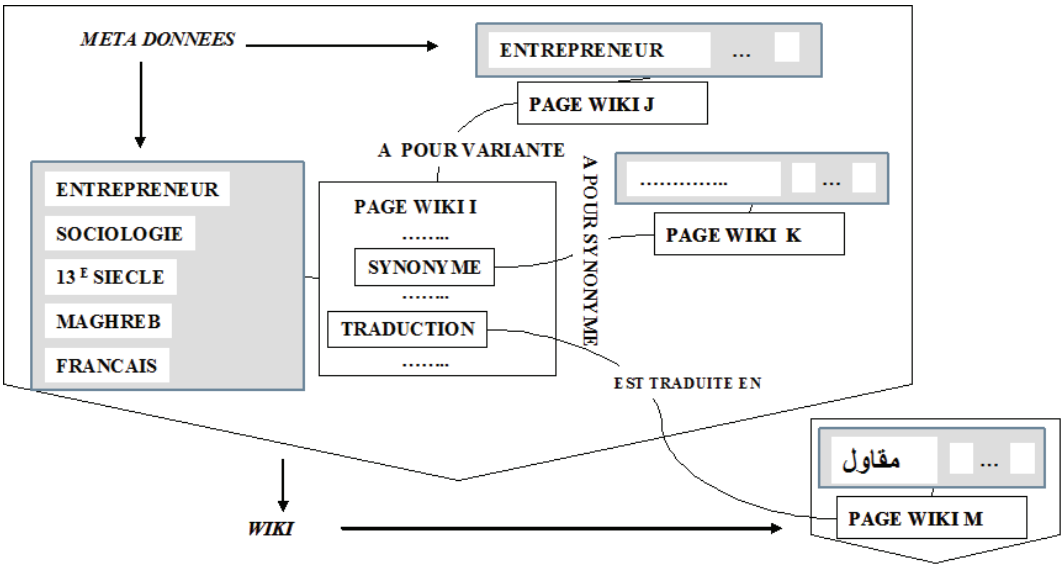


FIGURE 4.5 – Illustration de la structure du Wiktionnaire des SHS

Cette figure décrit une page Wiki pour une variante de la description du terme « Entrepreneur » qui est une entrée en Français du Wiktionnaire. Cette page est annotée par les valeurs de méta-données suivantes :

- « Entrepreneur » associé à la méta-donnée « terme »
- « Sociologie » qui correspond à une valeur de la méta-donnée « discipline »
- « Français » qui correspond à une valeur de la méta-donnée « langue »

Les valeurs « 13ième siècle » et « Maghreb » sont les valeurs respectives des paramètres temporels et géographiques. Ces deux paramètres représentent les éléments du contexte associés à la discipline « Sociologie ». Cette page Wiki, associée à une

variante de description du terme « Entrepreneur », est reliée à d'autres variantes via le lien « a pour variante ». De plus, cette variante contient un hyperlien « est traduite en » qui relie cette page à sa traduction en arabe pour le même contexte.

Pour la réalisation de l'application Wiktionnaire SHS, nous avons opté pour la construction d'un Wiki par langue et pour l'établissement des liens entre ces wikis. Un tel choix nous offre la possibilité de réaliser, dans un premier temps, un Wiktionnaire franco-arabe, extensible par la suite à d'autres langues et dialectes pratiqués dans le bassin méditerranéen.

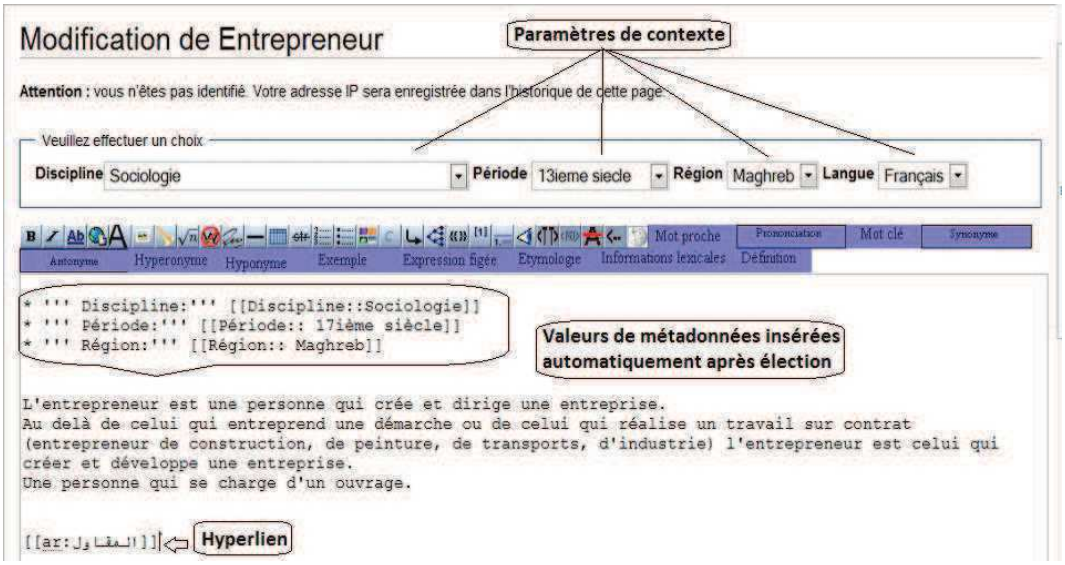


FIGURE 4.6 – Interface d'édition

L'éditeur de notre Wiktionnaire (figure 4.6) intègre, à l'heure actuelle, un sous ensemble des éléments de la norme ISO 1951. Son extension à l'ensemble des éléments de cette norme ou uniquement aux éléments utiles au domaine des SHS, est possible. L'utilisateur, via cet éditeur, peut annoter une page Wiki associée à une entrée du Wiktionnaire en utilisant les méta-données de son contexte de description. Il peut aussi compléter la description d'une entrée en utilisant des annotations associées aux éléments du schéma de la norme ISO 1951. Avant de saisir une description (dans une

langue donnée) d'une entrée, l'utilisateur doit fournir le contexte de définition de son entrée. En d'autres termes, il doit fournir la discipline, la langue concernée ainsi que les autres éléments de contexte qui rendront valide et spécialiseront sa description.

Selon le contexte fourni, le système propose soit de modifier une ancienne version de la description (dans le cas où l'entrée existe déjà sous le même contexte) ou encore de la créer. Durant la création ou la modification d'une description, l'utilisateur aura à utiliser les tags proposés pour ajouter éventuellement de nouveaux synonymes, antonymes, etc. Le MediaWiki sémantique se chargera, par la suite, de traduire ces tags en RDF.

L'utilisateur peut aussi consulter une description pour un contexte donné.

Le système, dans ce cas, lui fournira une description dans laquelle les hyperliens vers les synonymes, les antonymes, les termes associés et sa traduction apparaissent. Il peut demander à avoir la description globale d'une entrée. L'application se charge de le faire automatiquement en rassemblant, dans une seule page Wiki, les différentes variantes d'une entrée.


FIGURE 4.7 – Un exemple de consultation

La figure 4.7 est un exemple d'interface fournie à un utilisateur qui souhaite obte-

nir la description en Français du terme « Entrepreneur » pour un contexte décrit par le biais des valeurs saisies « Français, Sociologie, Maghreb, 13ième Siècle » associés aux paramètres de contexte.

## 4.5 Déploiement de l'application Wiktionnaire SHS sur un réseau pair-à-pair

Pour pallier les problèmes liés à la montée en charge des utilisateurs et des pannes réseaux, nous avons proposé d'intégrer l'application Wiktionnaire SHS dans un réseau pair-à-pair. De ce fait, les fonctionnalités attendues d'une telle entreprise consistent d'une part, en une édition collaborative en mode déconnecté de pages tout en maintenant la cohérence des données. Elles intègrent d'autre part, l'exécution de recherches avancées pour la construction de la fiche globale en faisant appel à des pairs spécifiques. Par ailleurs, cette extension devrait permettre le passage à l'échelle, la gestion des liens inter-wikis. Elle devrait aussi assurer la disponibilité des données.

Plusieurs projets exploitant la technologie wiki, le web sémantique et les réseaux pair-à-pair existent. Parmi eux citons Swooki [Rahhal et al., 2008], Distriwiki [Morris, 2007] et XwikiConcerto [Canals et al., 2008]. Swooki permet de fusionner des pages wikis contenant des annotations sémantiques en utilisant des algorithmes de fusion Wooto qui est l'adaptation de l'algorithme Woot [Oster et al., 2005]. Il permet aussi le travail en mode déconnecté, le passage à l'échelle. Il assure une amélioration des performances et une tolérance aux pannes.
DistriWiki est une solution pour remédier aux limites générées par les wikis sur une architecture client-serveur imposée par la nature centralisée des serveurs web. Pour cela, cette solution propose une décentralisation totale du réseau où chaque ordinateur agit comme un pair et stocke une copie redondante d'une page wiki où chaque page wiki (document) est identifiée d'une manière unique. Le but est de réduire la bande passante ainsi que le coût du matériel se rapportant au serveur central dans

l'architecture centralisée et les dépenses dues à sa maintenance.

En plus d'offrir des services dédiés à des interactions depuis des terminaux mobiles, XwikiConcerto propose un mécanisme de réplication multi-maître (plusieurs serveurs) avec fusion de données. Cette fusion supporte la modification simultanée des différentes copies d'un même objet. Ce mécanisme de fusion a pour but d'améliorer la disponibilité des services et d'accroître les performances du wiki. La cohérence des pages wiki y est également prise en charge en assurant la propagation et l'intégration des mises à jour des données dans les différentes répliques d'un même objet. Cette solution est sensée supporter un nombre illimité de nœuds et de répliques.

### 4.5.1 La description de l'architecture proposée

Notre proposition d'architecture est décrite dans la figure 4.8.

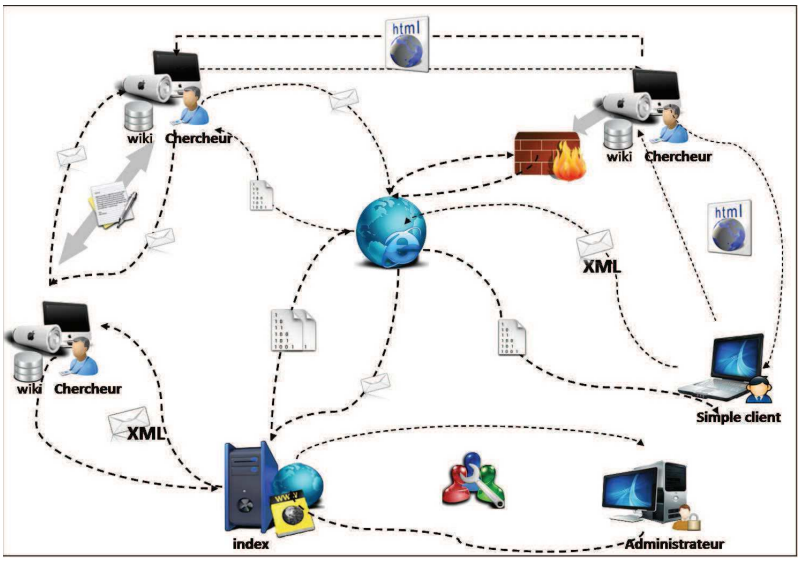


FIGURE 4.8 – Architecture générale de l'application Wiktionnaire SHS

C'est une architecture comportant :

– Un index qui permet une gestion des liens inter-wikis et le maintien de la cohérence du contenu. Il est chargé de l'indexation des pages et des chercheurs

qui les ont créées.

- Les pairs (peer) sont répartis en trois types d'utilisateurs :
  - 1 **les chercheurs** identifiés d'une manière unique à travers le réseau (utilisation d'un indice). Ils ont les droits d'édition et de consultation ;
  - 2 **les clients** (simples utilisateurs) pouvant se connecter au réseau de chercheurs grâce à l'index et effectuer uniquement des consultations sur l'ensemble des pages du réseau ;
  - 3 **e**t enfin, l'administrateur qui a le droit d'ajouter, de supprimer ou de modifier le profil des chercheurs ou encore de lancer ou d'arrêter le service d'indexation.

### 4.5.2 Les caractéristiques de l'application Wiktionnaire SHS distribué

L'application permettant d'utiliser le wiktionnaire est fondée sur l'architecture décrite ci-dessus. Elle permet l'édition collaborative aussi bien en mode connecté qu'en mode déconnecté. Le chercheur déconnecté de la plateforme peut éditer toutes ses pages en local, puis à sa reconnexion, effectuer une réconciliation de ses pages avec l'index. Cette réconciliation consiste en l'envoi d'un message à l'index en y spécifiant toutes les pages éditées lors de sa période de déconnexion. L'index se charge alors d'apporter les modifications chez les hébergeurs des copies de la page modifiée ou bien de les soumettre aux pairs détenteurs.

Une seconde caractéristique de notre application est sa capacité à gérer des liens inter-wiki c'est-à-dire des liens intégrés dans un contenu de la description d'une entrée du wiktionnaire (tels que les liens de synonymie et d'antonymie) faisant référence à des pages qui sont dans d'autres wikis.

Aussi, afin de garantir la disponibilité des données, le processus de réplication de pages est déclenché régulièrement par l'index. Il permet d'augmenter la disponibilité d'une page en la répliquant chez d'autres pairs chercheurs. Ces pairs sont choisis

en fonction de leurs taux de disponibilité sur le réseau ainsi que leurs domaines de recherche.

Pour remédier aux déconnexions des utilisateurs (chercheurs), qu'elles soient dues aux pannes du réseau, aux pannes système ou à la négligence humaine, nous avons intégré au niveau de l'index un processus qui se déclenche de façon régulière. Ce processus a pour objectif de vérifier la dernière date de connexion des chercheurs, Si cette date dépasse un certain seuil, un message de vérification de présence ($Test\_Connexion$) est alors envoyé à ce pair. Deux cas peuvent se présenter :

- 1- Si l'index reçoit une réponse ($Peer\_Connected$) de ce chercheur, cela voudra dire que le pair est toujours connecté. Dans ce cas de figure, l'index n'a qu'à mettre à jour la date de sa dernière connexion en mentionnant l'instant courant.
- 2- Si l'index ne reçoit pas de réponse au bout d'un certain temps, il déduit que ce chercheur est déconnecté et met le champ « connexion du chercheur » contenu dans l'index à « false ».

La présence de plusieurs copies de pages sur le réseau et l'aspect coopératif de l'application qui permet la création et la modification de pages par plusieurs chercheurs, engendre des incohérences au niveau des différentes répliques (copies). Pour remédier à ce type de problème, nous proposons la méthode suivante (voir figure 4.9).

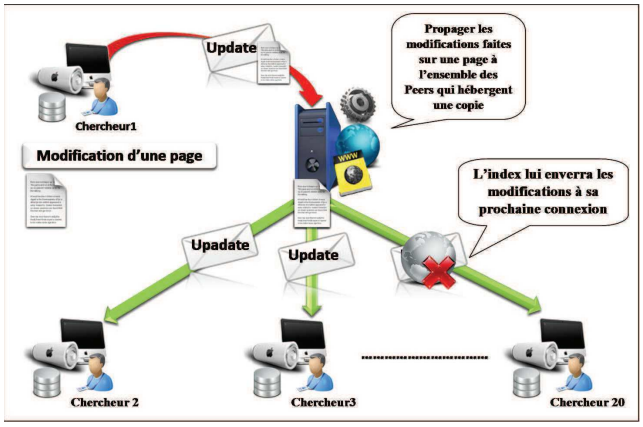


FIGURE 4.9 – Stratégie proposée pour maintenir la cohérence des données

Après chaque modification de page, le pair concerné envoie les mises à jour ainsi que les informations concernant son contexte à l'index. Ce dernier se charge de trouver les pairs connectés qui détiennent une copie de la page afin de leurs envoyer les modifications. Un pair peut détenir une copie de la page et être déconnecté au moment de la réconciliation. L'index met alors en attente ce pair et enregistre l'état des modifications. Ainsi, à sa prochaine connexion, le chercheur reçoit les modifications en question.

Notre application gère aussi la concurrence (édition simultanée d'une page) pour permettre une édition collaborative. Pour ce faire, nous avons opté pour une gestion à base de verrous. Toute page faisant l'objet d'une édition est verrouillée en écriture.

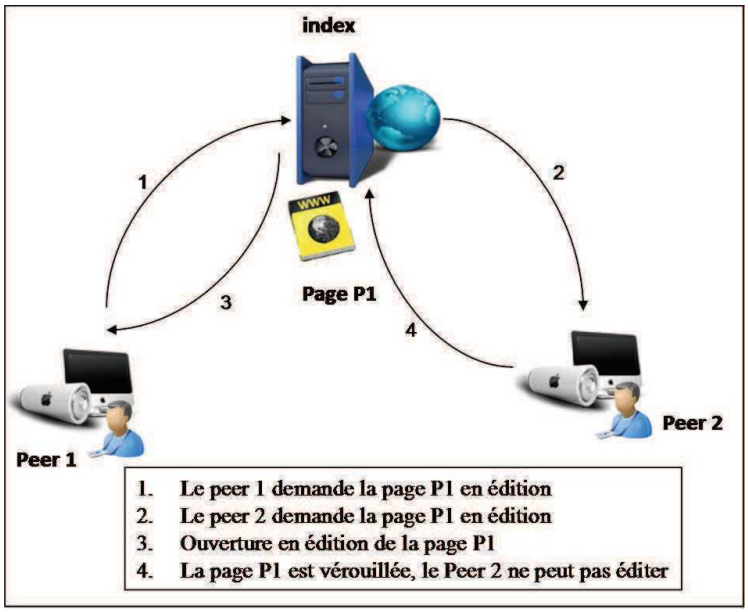


FIGURE 4.10 – Gestion des éditions concurrentes

Les pairs qui la demandent pendant cette période se voient refuser l'accès à celle-ci jusqu'à sa libération par le chercheur qui l'édite et ceci après la validation de sa modification. Les pairs mis en attente auront dès lors connaissance de la révision faite sur la page et peuvent décider de la réviser de nouveau selon le même principe (voir figure 4.10).

### 4.5.3 Le déploiement de l'application Wiktionnaire SHS sur un réseau pair-à-pair

Nous avons utilisé la plateforme JXTA pour le déploiement de l'application. Notre choix de JXTA est motivé par le fait que cette plateforme offre la possibilité à chaque composant d'un réseau de communiquer, de collaborer et de partager des ressources. Il assure également la communication et l'interconnection entre machines hétérogènes (par exemple machines de systèmes d'exploitation différents). Cette interconnexion est assurée à travers la création d'un réseau virtuel au-dessus du réseau physique, cachant ainsi la complexité de celui-ci.

L'une des interfaces de l'application est celle relative à l'édition des entrées du Wiktionnaire SHS.

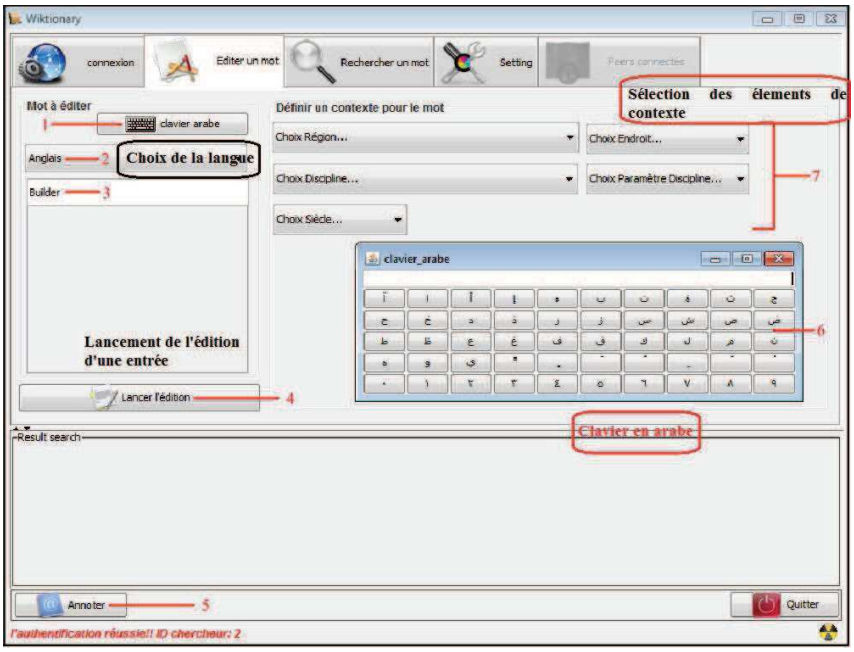


FIGURE 4.11 – Interface d'édition

Comme présenté dans la figure 4.11, l'interface offre un espace qui permet de choisir les éléments de contexte et offre également un clavier arabe pour permettre une saisie plus facile.

L'application offre aussi une interface de recherche personnalisée des entrées (les synonymes d'une entrée, mots-clé, exemple...) à travers tous les nœuds (pairs) connectés. Elle propose deux modes de recherche : une recherche par terme (entrée) selon un contexte donné et une recherche par terme sans spécification de contexte. Le second mode déclenchera la construction à la volée de la fiche globale (voir figure 4.12).

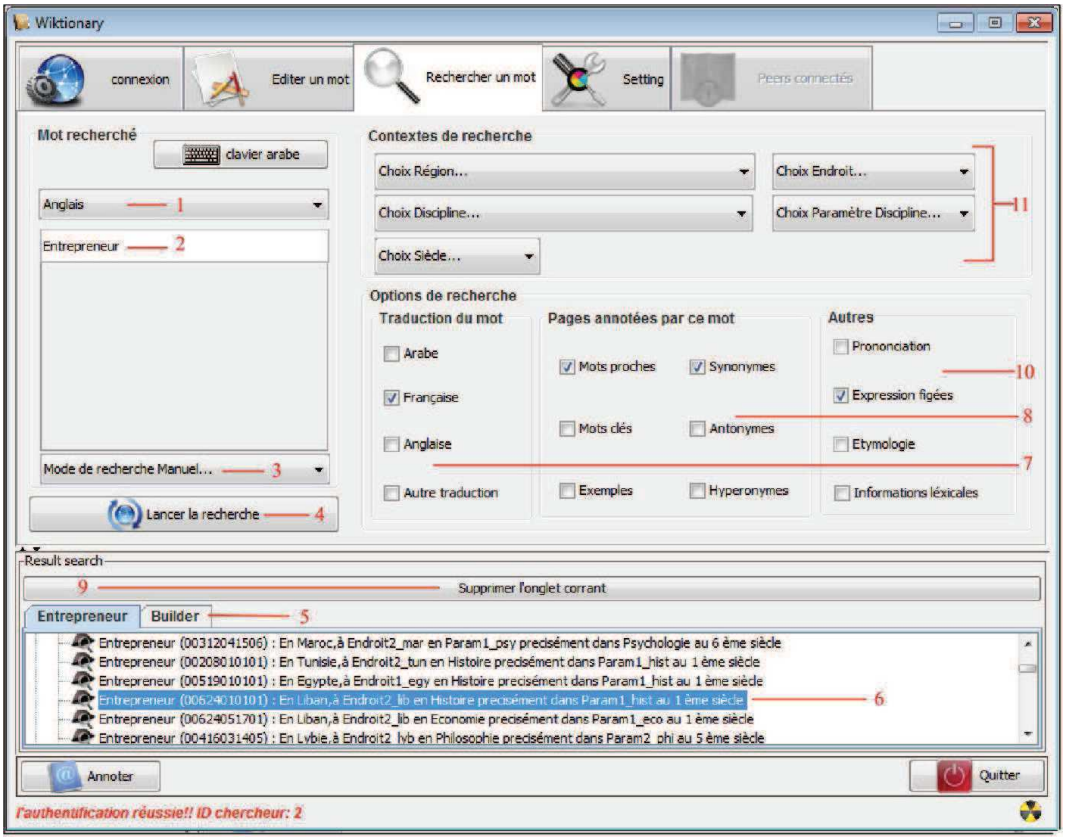


FIGURE 4.12 – Interface de recherche

## 4.6 Conclusion

Nous avons décrit dans ce chapitre la structure du dictionnaire en ligne des SHS. Pour sa réalisation nous avons utilisé, conformément au cahier des charges, la tech-

nologie Wiki pour une édition facile et collaborative de son contenu. Ce dictionnaire est multilingue et multiculturel dans le sens où il intègre des termes de langues différentes et il tient compte de la variation du sens des termes. Il est aussi conforme à la norme ISO 1951.

Sa structure générique permet de le réutiliser dans d'autres domaines que les SHS.

L'application Wiktionnaire SHS ainsi conçue a été mise en œuvre sous une architecture centralisée puis transposée vers une architecture réseaux de type pair-à-pair afin de permettre à des chercheurs des deux rives de la Méditérranée de collaborer pour son alimentation et de le consulter quelques soient leurs lieux de résidence. Lors de son alimentation, ce Wiktionnaire contribuera au développement des échanges entre chercheurs du bassin Méditerranéen et à la mise en commun d'un ensemble de savoirs sur les deux cultures et les deux sociétés.

# Chapitre 5

# Famille d'approches pour la construction et l'évolution de Topic Maps contextualisées

L'annotation et l'interconnexion de ressources informationnelles sont des aides précieuses pour la recherche d'information dans des contenus volumineux. Ces aides peuvent être offertes via une Topic Map à créer et à maintenir afin d'assurer sa durabilité.

La création d'une Topic Map associée à un contenu consiste en l'instanciation de son modèle (sa structure) compte tenu des ressources qu'elle organise. Le choix de la structure associée à la Topic Map impacte inévitablement la recherche d'information en terme de précision. Ceci explique les différentes propositions d'extension du modèle standard présentées dans le chapitre 3. Ce dernier nous a permis de constater que ni le modèle standard ni les extensions proposées ne permettaient d'organiser des contenus à la fois multilingues et pluridisciplinaires de telle sorte à offrir à l'utilisateur une recherche efficace et précise. La raison est que ces modèles ne prennent pas en compte la variabilité du sens des termes et des concepts au sein d'un domaine

ni même entre domaines. Pour répondre à ce besoin qui nous semble primordial, nous proposons un nouveau modèle de Topic Maps qu'on a baptisé modèle de Topic Maps contextualisées. Ce modèle permet d'associer des contextes de validité à des topics. Ces contextes ont la même description que celle présentée dans le chapitre précédent.

Le premier objectif de ce chapitre est de décrire le modèle de Topic Maps contextualisées. Cette description est suivie d'une présentation d'une famille d'approches permettant de créer et de faire évoluer une Topic Map contextualisée.

Le plan de ce chapitre s'articule donc de la façon suivante. La première section présente la structure sous-jacente aux Topic Maps contextualisées. Cette présentation est suivie par une présentation générale d'une famille de méthodes permettant de construire et faire évoluer les Topic Maps contextualisées. Elle sera suivi par une présentation des composants servant à la construction de cette famille de méthodes. La dernière section de ce chapitre conclut les différentes contributions.

## 5.1 Modèle de Topic Maps contextualisées

Pour gérer la variation du sens des termes, nous avons mis en œuvre dans le chapitre précèdent, le concept de contexte défini dans le cadre du Wiktionnaire SHS. Nous réutilisons ce concept tout au long de la suite de la thèse. Rappelons que nous l'avions défini comme un ensemble composé des deux paramètres stables qui sont la langue et la discipline et d'autres paramètres liés à la discipline telles que la zone géographique d'usage et la période de validité de la description associée au contexte. Nous noterons dans la suite de cette thèse un contexte par le n-uplet « langue, discipline, zone géographique, période ».

La prise en compte de la variation du sens, nous permet de mieux gérer l'hétérogénéité sémantique des ressources (c'est-à-dire la variation du sens des termes utilisés dans les ressources). Pour ce faire, nous représentons, dans la Topic Map, un sujet (topic), que nous relions à une ressource informationnelle, à travers ses diffé-

rents contextes de telle sorte que lorsqu'un sujet est associé à un contexte donné, il permettra d'indexer uniquement les ressources informationnelles où ce sujet apparaît sous ce contexte. Nous exploitons pour cela de façon intensive le mécanisme de réification proposé dans le TMDM, dans un premier temps pour construire des contextes et dans un second temps pour contextualiser des sujets. La figure 5.1 illustre notre propos. Dans cette figure, nous avons construit un topic « $Context_1$ » à partir des

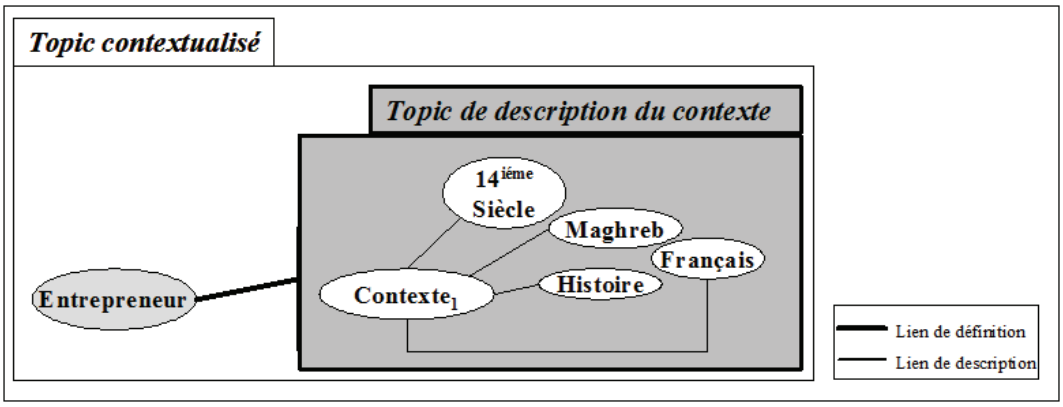


FIGURE 5.1 – Un exemple de réification dans une Topic Maps

topics « Français », « Histoire », « Maghreb », « 14ème siècle ». « $Context_1$ » est un topic de type « contexte ». Les topics « Français », « Histoire », « Maghreb » et « 14ème siècle » sont des topics de type « paramètre de contexte ». Les liens entre un contexte et un paramètre de contexte est un lien de description. Dans cette même figure, nous avons construit un topic contextualisé qui pourra être lié à une ou plusieurs ressources. Ce topic contextualisé a été construit en réifiant l'association « est défini pour » qui existe entre le topic non contextualisé « Entrepreneur » et le contexte « $Context_1$ ».

Une fois les topics contextualisés construits, nous établissons des associations entre topics contextualisés afin de permettre une navigation au sein des ressources. À titre d'exemple, dans la figure 5.2, le terme « Entrepreneur », placé dans le contexte « Français, Histoire, Maghreb, 14ème siècle » est traduit en « مقاول » dans le

contexte « العربية, تاريخ, المغرب, القرن الخامس عشر » dont l'interprétation en langue française est « Arabe, Histoire, Maghreb, 15ème siècle ».

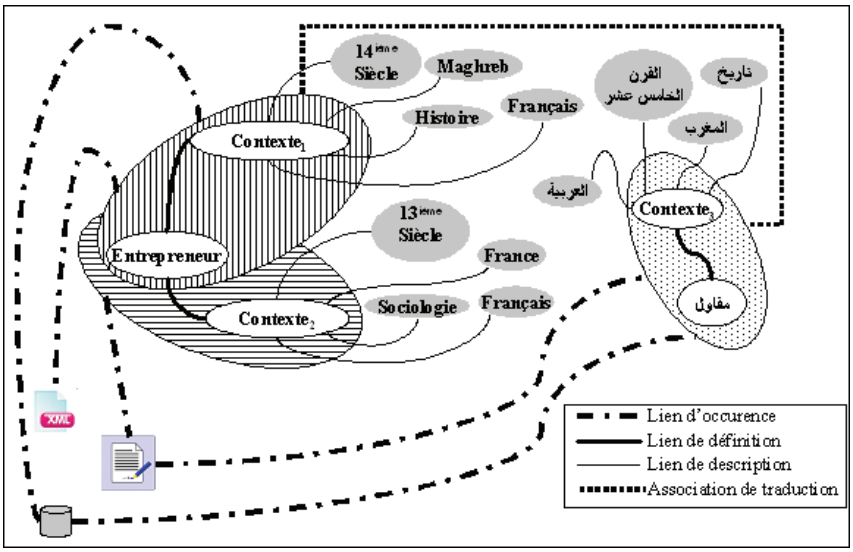


FIGURE 5.2 – Exemple de topics contextualisés

En résumé, la structure d'une Topic Map contextualisée se décompose en trois grands réseaux sémantiques placés sur une architecture à trois niveaux. Ces réseaux sémantiques sont reliés entre eux à l'aide du mécanisme de réification proposé dans le TMDM. La figure 5.3 illustre cette représentation.

On distingue un premier ensemble de topics qui regroupe les paramètres de contexte potentiels. Ces paramètres de contexte sont liés entre eux par des liens de type « est décrit par ». Leurs associations sont réifiées par des topics particuliers représentant les différents contextes. Le deuxième ensemble relie des contextes à des termes « non contextualisés » à l'aide le lien « est défini par ». Ces associations sont réifiées, à leur tour par des topics matérialisant, cette fois, des termes « contextualisés » qui indexent les ressources informationnelles. Ces termes contextualisés peuvent être liés entre eux par tout type de lien mentionné dans le TMDM.

Ce principe de contextualisation a été modélisé et a abouti à un méta-modèle conceptuel. Ce dernier, illustré dans la figure 5.4, décrit la structure abstraite des

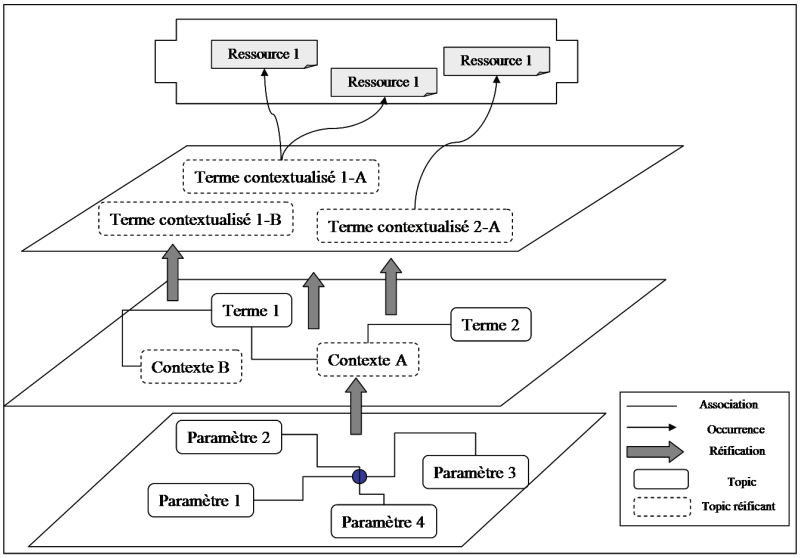


FIGURE 5.3 – Principe de contextualisation des topics

Topic Maps contextualisées.

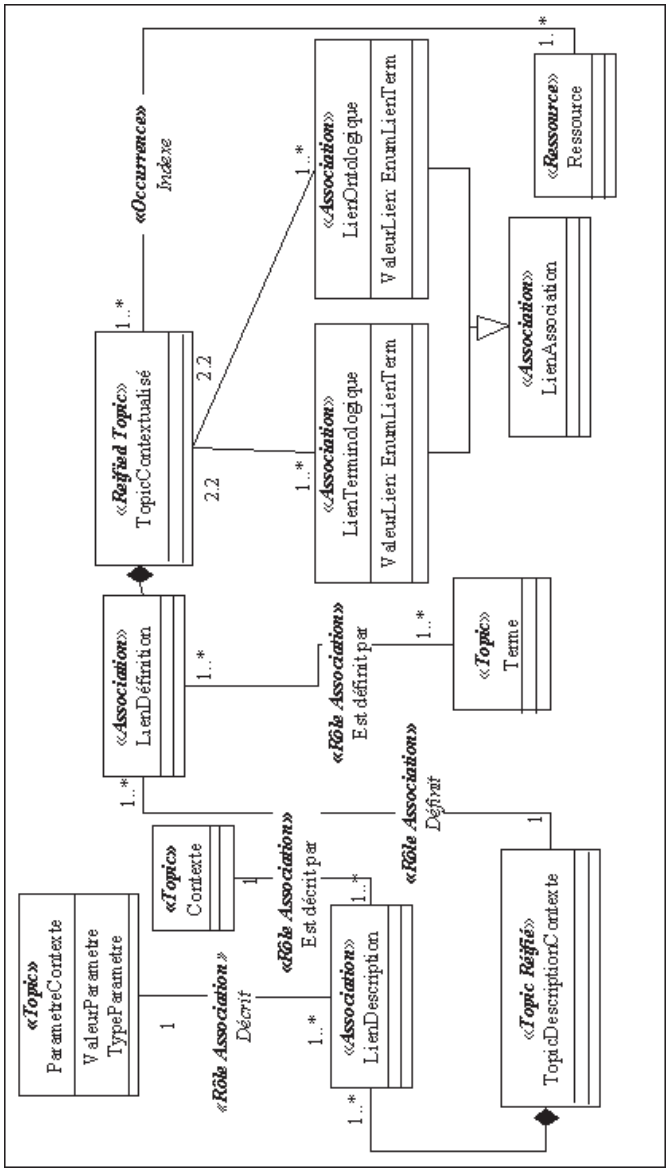


FIGURE 5.4 – Méta-modèle conceptuel de Topic Map contextualisée proposé

Cette modélisation conceptuelle sert de base théorique pour la mise au point du méta-modèle logique permettant d'instancier ce type de Topic Maps. Ce dernier est décrit dans le paragraphe suivant et porte le nom de « méta-modèle contextuel ».

**Remarque 1** : La contextualisation de topics et leurs associations à d'autres topics contextualisés ne peut être réalisée qu'avec le mécanisme de réification. En effet, bien que le concept de « scope », existant dans le TMDM, permet de définir des contextes, il ne permet pas de relier des topics contextualisés.

**Remarque 2** : Nous avons construit un topic appelé « context description topic » à partir d'un ensemble de paramètres de contexte dans un objectif de sa réutilisation dans la contextualisation de différents topics. Ce topic est, par conséquent, representé dans notre méta-modèle au même titre que les autres types de topics.

### 5.1.1 Le méta-modèle contextuel

Le méta-modèle conceptuel proposé ci-avant fait usage d'un profil UML permettant de masquer le détail de certains concepts tels que les rôles d'associations ou les occurrences. Il se concentre sur la structure nécessaire à la contextualisation. Pour implémenter ce méta-modèle tout en conservant le profil, il faut le compléter en précisant les valeurs marquées de chacun des stéréotypes. Pour ce faire, la réutilisation du profil Topic Maps de la spécification ODM a été envisagée dans un premier temps. Il s'est avéré que ce profil n'autorisait que la modélisation de la classification d'une Topic Map. L'approche de contextualisation suggérée s'intéresse peu à la classification, mais précise, en revanche, une structure particulière qui pousse le profil ODM hors des limites de son cadre d'utilisation.
Dans un second temps, nous avons envisagé de nous passer d'un profil UML tout en conservant un niveau d'abstraction satisfaisant permettant de modéliser notre approche de contextualisation. Une Topic Map contextualisée pouvant être considérée comme une Topic Map standard dans laquelle on a distingué les topics et les associations, nous permet d'envisager une autre solution pour la modélisation

logique de notre modèle conceptuel. Cette solution consiste à considérer une Topic Map contextualisée comme une spécialisation d'une Topic Map standard. La mise en œuvre de cette solution passe par le mécanisme de fusion des méta-modèles existant dans UML. L'extrait suivant, issu de la spécification UML [1], précise ce concept de *package merge* :

*« A package merge is a directed relationship between two packages that indicates that the contents of the two packages are to be combined. It is very similar to generalization in the sense that the source element conceptually adds the characteristics of the target element to its own characteristics resulting in an element that combines the characteristics of both ».*

[...] *« Conceptually, a package merge can be viewed as an operation that takes the contents of two packages and produces a new package that combines the contents of the packages involved in the merge. In terms of model semantics, there is no difference between a model with explicit package merges, and a model in which all the merges have been performed ».*

La spécialisation du modèle de données standard rentre parfaitement dans le cadre de cette définition. En héritant des caractéristiques du TMDM, le méta-modèle contextuel traduit une réelle combinaison des Topic Maps dites standards et de la structure contextuelle sous-jacente. Tous les éléments nécessaires, tels que les rôles d'association sont ainsi communiqués à travers des généralisations. La figure 5.5 fournit le méta-modèle contextuel résultant de cette fusion.

Toutes les spécialisations ne sont pas indiquées afin de ne pas alourdir le diagramme. La règle générale qui doit être observée est que les stéréotypes « topic », « association » et « occurrence » spécialisent respectivement, les classes Topic, Association et Occurrence. Notons que ces trois stéréotypes ont été conservés afin de préciser la sémantique du méta-modèle et ne s'accompagnent donc d'aucune valeur marquée.

1. OMG-UML. http ://www.omg.org/spec/UML/2.4.1/Infrastructure/PDF/

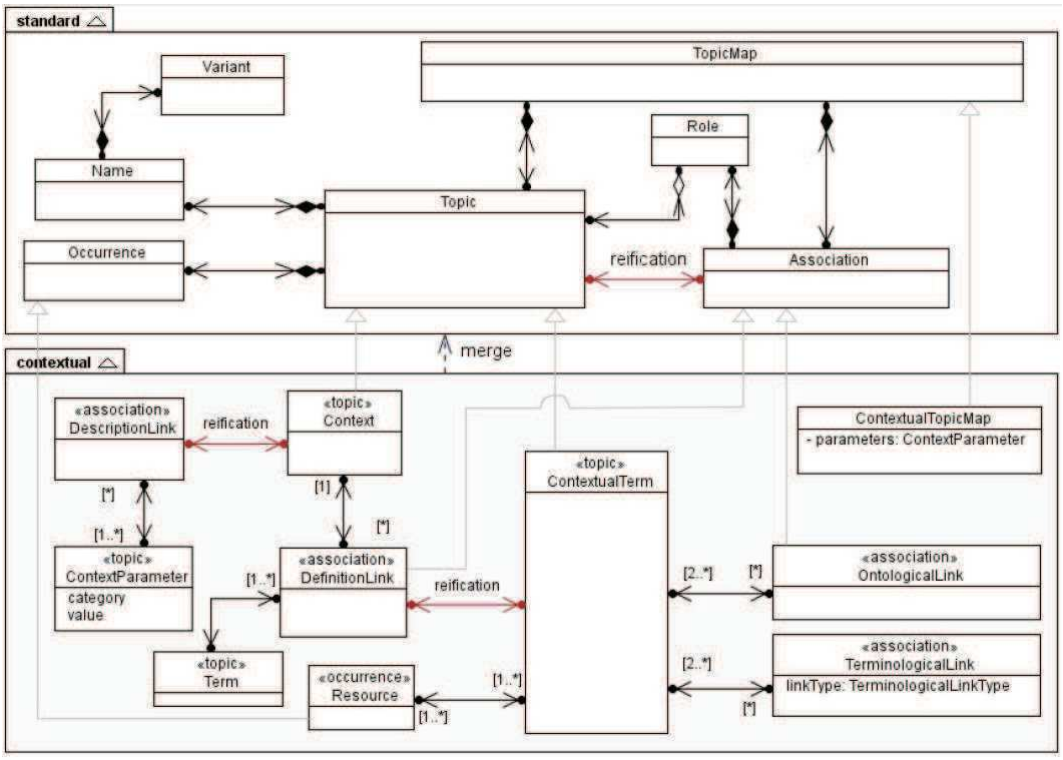


FIGURE 5.5 – Méta-modèle contextuel

Ce méta-modèle apporte, par ailleurs, une modification notable à la version conceptuelle pour rester dans le cadre imposé par le TMDM. Ce changement concerne la modélisation du contexte et de ses paramètres qui demandait de réifier plusieurs associations binaires par un même topic.

Le méta-modèle contextuel obtenu permet d'instancier une Topic Map contextualisée comprenant toutes les caractéristiques du modèle de données standard. Ce méta-modèle logique (méta-modèle contextuel) va constituer le cœur du prototype autour duquel il est possible de bâtir des fonctionnalités telle que celle de la création d'une Topic Map contextualisée.

## 5.2 Présentation générale des familles d'approches pour la construction et l'évolution d'une Topic Map contextualisée

L'instanciation d'un modèle de Topic Map, pour un contenu volumineux ne peut être effectuée sans automatisation. Une fois construite, elle doit à tout instant refléter le contenu qu'elle organise afin de satisfaire pleinement les besoins en recherche d'informations. Dans le cadre de cette thèse, nous nous intéressons à trois types de création de Topic Maps contextualisées :

- une création à partir d'un contenu multilingue et pluridisciplinaire,
- une création par fusion de Topic Maps contextualisées existantes,
- et enfin une création par ré-ingénierie d'une Topic Map standard associée à des ressources monolingues ayant trait à un domaine.

Nous nous intéressons aussi à deux types d'enrichissement de la Topic Map contextualisée. Le premier est un enrichissement suite à un processus de fusion de la Topic Map source avec une autre Topic Map contextualisée. Le second est un enrichissement dû à l'intégration de nouvelles ressources informationnelles.

Notre intérêt pour ces types de création et évolution de Topic Maps contextualisées est motivé par le fait que les approches correspondantes peuvent être construites par assemblage de composants élémentaires (méthodes). Un composant d'une approche est, la plupart du temps, réutilisé dans au moins une approche de cette famille de méthodes. Ainsi, comme le montre la figure 5.6, nous proposons :

- de créer à partir d'un contenu multilingues et pluridisciplinaires d'une Topic Map contextualisée en procédant à une annotation des ressources informationnelles composant le contenu suivie d'une génération de la Topic Map contextualisée, de son exportation, de son stockage et éventuellement de sa visualisation.
- de créer une Topic Map contextualisée par fusion deux Topic Maps contextua-

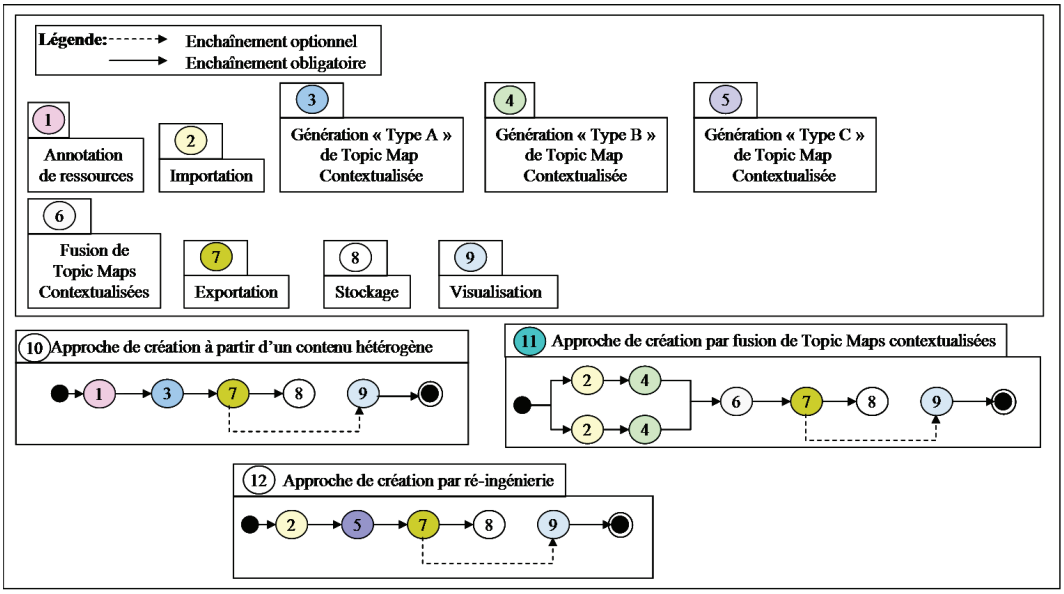


FIGURE 5.6 – Famille d'approches de construction de Topic Maps

lisées en important ces deux Topic Maps et en les fusionnant dans un premier temps puis, dans un second temps, en exportant la Topic Map résultante de la fusion en vue de son stockage et éventuellement de sa visualisation.

– ou encore de la créer par un processus de de ré-ingénierie d'une Topic Map standard associée à des ressources monolingues et monodisciplinaire. Ce processus consiste en l'importation de la Topic Map standard, en sa transformation en Topic Map contextualisée, en son exportation et stockage et éventuellement en sa visualisation. .

Nous utilisons aussi la même famille de méthodes pour les deux enrichissements cités ci-dessus, comme illustré dans la figure 5.7. En effet, pour enrichir une Topic Map contextualisée avec les instances d'une autre Topic Map contextualisée, nous proposons d'importer la Topic Map source et de la fusionner avec la Topic Map cible. La Topic Map enrichie est ensuite exportée en vue de son stockage et éventuellement visualisée. Ce processus revient à appliquer le second type de création énoncé plus haut. Dans le cas du second type d'enrichissement, nous proposons de construire

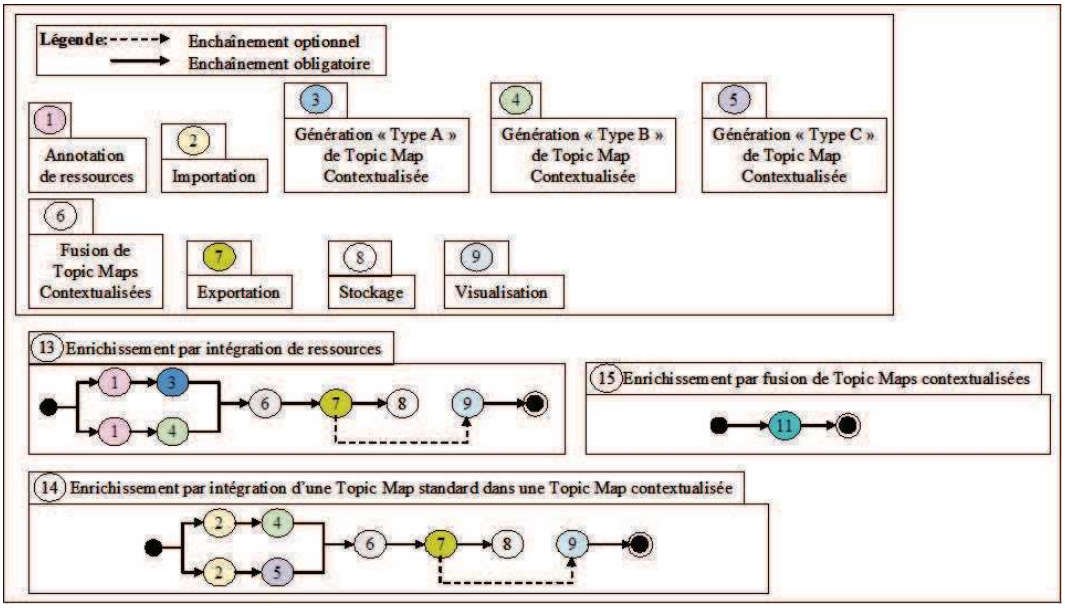


FIGURE 5.7 – Famille d'approches d'enrichissement de Topic Maps

la Topic Map contextualisée associée à la ressource à intégrer et de la fusionner avec la Topic Map cible. La Topic Map enrichie est ensuite exportée puis stockée et éventuellement visualisée.

Notons que la plupart des composants « méthodes » de cette famille suivent un processus dirigé par les modèles assurant ainsi une évolutivité des approches. Un rappel sur l'ingénierie dirigée par les modèles est fourni en annexe B de cette thèse.

Les sections suivantes présentent chacun des composants méthodes présentés dans les figures 5.7 et 5.6, hormis le composant visualisation qui fait l'objet du chapitre suivant.

## 5.3 Processus d'annotation de ressources informationnelles

Ce processus permet d'instancier le méta-modèle d'annotation de ressources décrit ci-dessous (voir figure 5.10). Ce dernier regroupe toutes les métadonnées attri-

buées, semi-automatiquement, par ce composant, à chacune des ressources informationnelles. Il consiste, dans un premier temps, en l'extraction pour chaque ressource informationnelle composant le contenu, d'un ensemble de termes représentatifs de cette ressource puis, dans un second temps, en leur contextualisation (voir figure 5.8) .

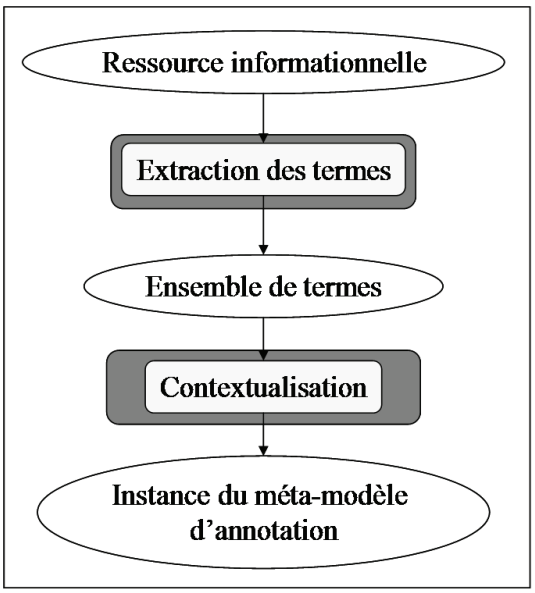


FIGURE 5.8 – Processus d'annotation des ressources informationnelles

Ce processus s'appuie, pour ce qui est de l'extraction des termes, sur un ensemble de techniques d'extraction existantes adaptées à la nature du contenu. En d'autres termes, si le contenu est totalement composé de documents textuels multilingues, les techniques et les outils de traitement du langage naturel tels que ceux citer dans [Ellouze et al., 2012] pourront être utilisés.

La contextualisation des termes est, quant à elle, effectuée manuellement par un expert. Les termes contextualisés sont décrits par un ensemble de déclarations RDF. Chaque déclaration est un triplet « sujet, prédicat, objet ». Pour contextualiser un terme et décrire un contexte comme une aggrégation de paramètres de contexte, nous utilisons le mécanisme de réification. Pour ce faire, nous simulons la contextualisation

à l'aide du constructeur RDF « statement » et nous exprimons l'aggrégation à travers le constructeur « collection ».

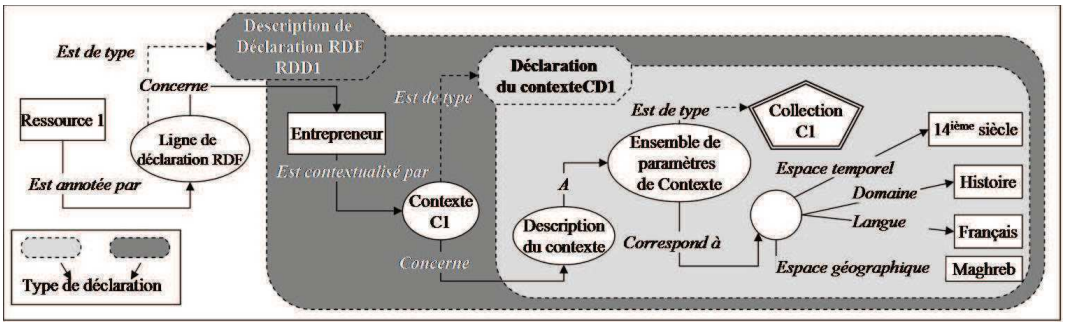


FIGURE 5.9 – Un exemple du graphe RDF décrivant une métadonnée contenue dans une ressource

À titre d'exemple, la figure 5.9 est un graphe RDF décrivant le fait que le terme « Entrepreneur » est présent dans la ressource « $ressource_1$ » et sa description dans cette ressource est valide dans le contexte défini par les paramètres « Français, Histoire, Maghreb, 14ème siècle ». Dans cette figure, cette déclaration est nommée « LigneDeclarationRDF ». Elle est enregistrée dans le méta-modèle d'annotation des ressources décrit dans la figure 5.10. Ce méta-modèle exploite le profil UML afin d'étendre UML aux concepts de RDF. Dans ce métamodèle, sont cités des exemples de paramètres de contexte. Ils sont à adapter en fonction des disciplines représentées par le contenu.

## 5.4 Processus de génération « Type A » d'une Topic Map contextualisée

Ce processus permet d'instancier le méta-modèle contextuel, décrit dans le paragraphe 5.1.1, à partir d'annotations ; une instance de ce méta-modèle étant une Topic Map contextualisée correspondante aux ressources annotées.

Cette instanciation est réalisée en trois étapes (voir figure 5.11). La première

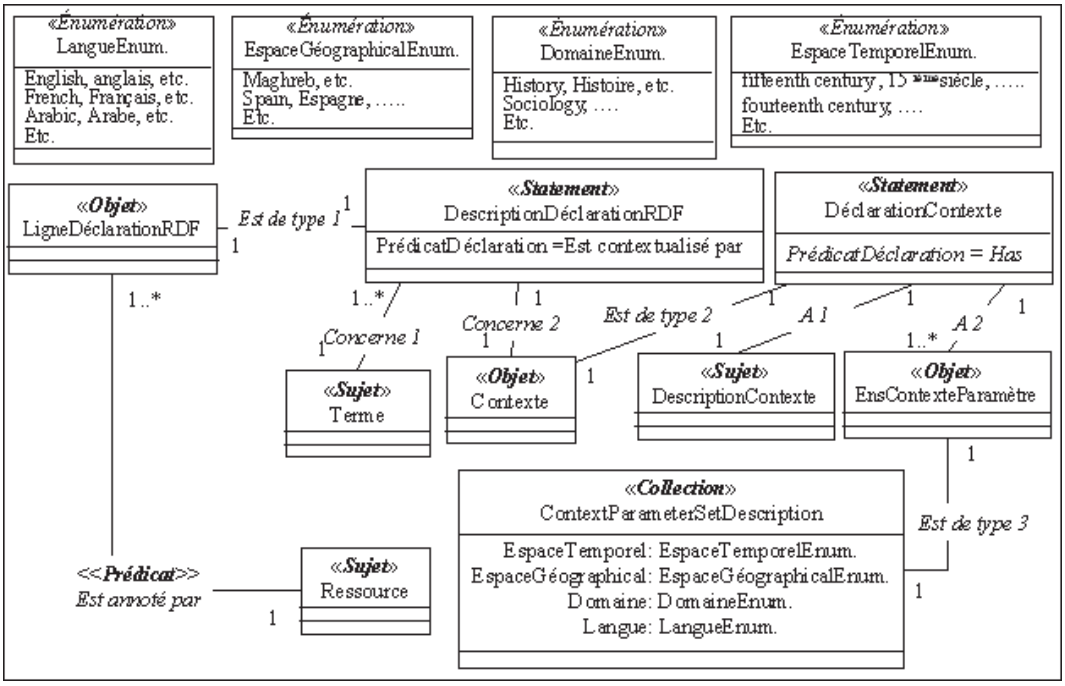


FIGURE 5.10 – Méta-modèle d'annotation

étape permet la génération du squelette de la Topic Map contextualisée (c'est-à-dire ses différents types de topics, et les occurrences des topics contextualisés). Les deux étapes suivantes permettent la construction du réseau sémantique à l'aide de la dérivation de liens entre topics contextualisés.

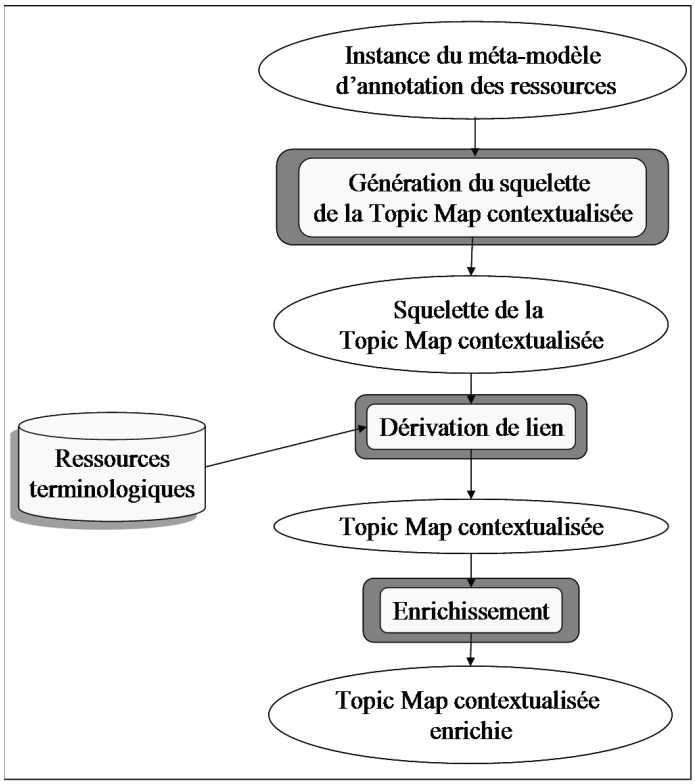


FIGURE 5.11 – Processus de génération d'une Topic Map contextualisée

À l'issue de la première étape, une instance du méta-modèle contextuel est générée. Dans cette instance apparaissent uniquement les topics et les occurrences de la Topic Map contextualisée. La génération de cette instance du méta-modèle conceptuel est réalisée par application des règles de transformation de modèles (M2M dans le sens de l'IDM) s'appuyant sur la mise en correspondance des concepts des deux modèles (voir tableau 5.1). La seconde étape du processus complète l'instance du méta-modèle contextuel par l'ajout de liens terminologiques. C'est une étape qu'on peut semi-automatiser si toutefois l'on dispose d'une ressource terminologique multilingue associée au domaine tel qu'un dictionnaire électronique multilingues et spécialisé. Nous devons, dans ce cas, mettre en œuvre un algorithme permettant de rechercher un lien terminologique entre deux termes et le proposer à l'expert du

| **Concepts du méta-modèle d'annotation de ressources** | **Concepts du méta-modèle contextuel d'une Topic Map contextualisée** |
|---|---|
| Classe ContextParameterSetDescription | Classe ContextParameter |
| Attribut EspaceTemporel,<br>Attribut EspaceGéographqiue,<br>Attribut Domaine,<br>Attribut Langue | Classe ContextParameter |
| Classe ContextParameterSet,<br>Classe Contexte | contexte |
| Association « Isoftype2 » | Reification |
| Association « Has1 »,<br>Association « Has2 »,<br>Classe ContextDeclaration,<br>Classe DescriptionContexte,<br>Valeur « Has »,<br>Association « Concern 1 » | DescriptionLink<br><br>DescriptionLink |
| Classe Term | term |
| Valeur « Is contextualized by » | DefinitionLink |
| Classe DescriptionDeclarationRDF,<br>Classe RDF DeclarationLine<br>Classe RDF DeclarationLine | ContextualTerm |
| Association « Is of type1 »<br>Association « Concern 2 » | Reification |
| Classe Resource | Resource |
| Association « Is annotated by » | occurence |

TABLE 5.1 – Mise en correspondance des concepts du méta-modèle d'annotations avec le méta-modèle contextuel

domaine pour un éventuel ajout de liens terminologique entre deux topics contextualisés. Si l'on dispose d'une ressource terminologique (tel que le wiktionnaire SHS présenté dans le chapitre 4 de cette thèse) où les termes ont été contextualisés alors le processus de déduction de liens entre topics contextualisés est totalement automatique. Enfin, dans le cas où aucune ressource terminologique n'est disponible, cette étape demeure manuelle.

La troisième et dernière étape de ce processus de génération d'une Topic Map contextualisée est une étape manuelle où l'instance du modèle contextuel résultat de l'étape précédente est enrichie par des liens d'associations fournis par un expert du domaine.

Une fois la Topic Map générée, elle est exportée en vue d'être stockeé et éventuellement visualisée

## 5.5 Processus d'importation et d'exportation d'une Topic Map

Le processus d'importation se charge d'instancier le méta-modèle standard à partir d'une Topic Map sérialisée dans un format d'échange donné.

La figure 5.12 donne une vue d'ensemble du processus d'importation qui comprend une étape correspondante à une transformation texte à modèle (T2M dans le sens de l'IDM). Elle permet de créer une Topic Map sous le format standard à partir d'un format d'échange textuel. Pour ce faire, une instance du modèle associé au format d'échange est d'abord créée, puis transformée de façon directe en une instance du modèle standard. La première phase de cette transformation T2M est, la plupart du temps, réalisée par des utilitaires disponibles sur le marché, compte tenu du fait que la plupart des formats d'échange sont standardisés (XTM, RDF, etc. ).

Le processus d'exportation correspond au traitement inverse de l'importation (voir figure 5.13). Il permet de sauvegarder à des fins de partage une Topic Map

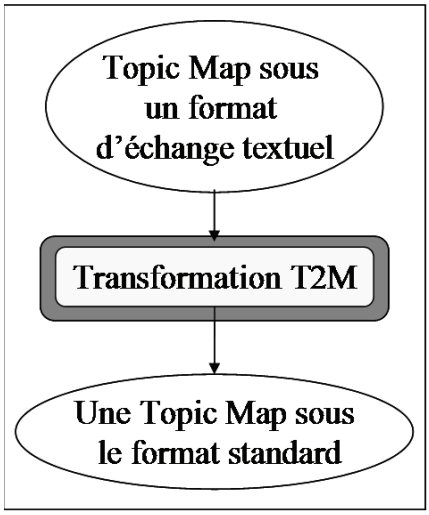


FIGURE 5.12 – Processus d'importation d'une Topic Map contextualisée

contextualisée. Cette exportation implique une sérialisation du modèle contextuel vers un format d'échange donné.

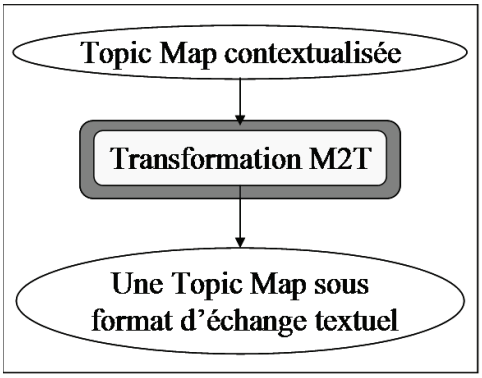


FIGURE 5.13 – Processus d'exportation d'une Topic Map contextualisée

Ce processus repose sur une transformation unique de modèles de type Modèle to Text (M2T dans le sens de l'IDM). Ce processus est simple à mettre en place du fait que le modèle contextuel est une spécialisation du modèle standard. Il se base sur les gabarits de formats d'échange afin de fournir un document textuel. L'instanciation de ce gabarit se fait en parcourant la partie standard du modèle contextuel.

## 5.6 Processus de génération « Type B » d'une Topic Map contextualisée

Ce processus convertit une description d'une Topic Map contextualisée exprimée sous le format du modèle standard, vers une description sous le format du modèle contextuel. Il se déroule en deux étapes comme illustré dans la figure 5.14.

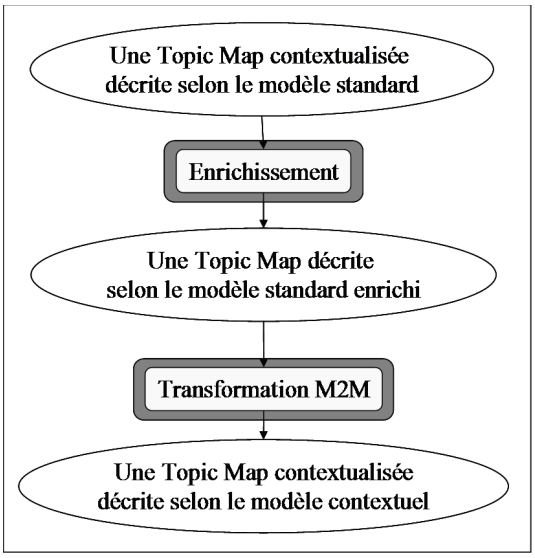


FIGURE 5.14 – Processus de génération « Type B » d'une Topic Map contextualisée

La première étape du processus consiste en la création d'une Topic Map sous un format standard enrichi. L'enrichissement correspond à l'identification de la structure contextuelle sous-jacentes du fait que le modèle standard ne fait pas la distinction entre les différents topics. Les informations concernant cette Topic Map sont alors stockées dans une instance du modèle standard enrichi, qui n'est autre que le modèle standard, auquel on ajouté des propriétés qualifiant les associations et les topics. L'algorithme utilisé pour la détection de la structure contextuelle sous-jacente est fourni en annexe C de cette thèse. La seconde étape de ce processus est une transformation Modèle à Modèle (M2M dans le sens de l'IDM). Elle permet de fournir une description d'une Topic Map contextualisée conforme au modèle contextuel à partir

d'une description sous le format standard enrichi. En d'autres termes, il s'agit d'appliquer les règles de transformation de schéma qui, selon la qualification attribuée à un élément de l'instance du modèle standard enrichi, déduisent un élément d'une instance du modèle contextuel. À titre d'exemple, on peut citer la règle suivante : « si un topic d'une instance du modèle standard a été qualifié de « paramètre » alors ajouter à l'instance du méta-modèle contextuel un topic de type « paramètre de contexte ».

## 5.7 Processus de génération « Type C » d'une Topic Map contextualisée

Il s'agit dans ce processus de procéder à la ré-ingénierie d'une Topic Map standard en une Topic Map contextualisée (voir figure 5.15). Comme le processus de génération « Type B », ce processus, est avant tout, un processus de transformation des schémas. Ce qui le distingue du processus « Type B » est le sujet de la source et, par voie de conséquence, l'étape de préparation de la transformation.

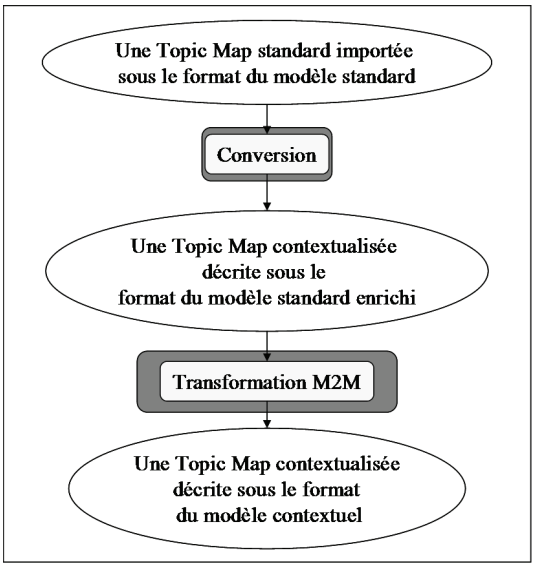


FIGURE 5.15 – Processus de génération « Type C » d'une Topic Map contextualisée

En effet, alors que dans le processus de génération, on procède à une identification de la structure contextuelle, dans ce processus, il s'agit de la construire. Cette construction implique la ré-indexation des ressources. La construction de la structure contextuelle nécessite l'intervention de l'expert dont le rôle est de proposer les paramètres de contexte associé à chaque topic pour la discipline associée au contenu à organisée ; la langue étant aussi une donnée de ce processus.

## 5.8 Fusion de deux Topic Maps contextualisées

Ce processus de fusion est semblable au processus de fusion de schémas connu dans le domaine des systèmes d'information. Ce dernier propose une intégration en quatre étapes [Batini et al., 1986] : une étape de pré-intégration dont l'objectif est d'uniformiser les schémas sources sous un même formalisme, une étape de comparaison qui permet d'identifier les relations inter-schémas sources, une étape de fusion qui rassemble les schémas sources en un schéma intégré selon les résultats de l'étape précédente et les règles d'intégration existantes, et enfin, une étape de restructuration du schéma intégré.

Notre processus, démarrant avec des Topic Maps de même structure, la première étape s'avère inutile.

Comme pour la plupart des systèmes de fusion de Topic Maps existants dans la littérature, notre processus effectue sa comparaison en priorité entre les topics. De plus, l'architecture par niveau d'une Topic Map contextualisée impose, non seulement une comparaison de topics de même niveau mais aussi un ordonnancement des comparaisons et des traitements des conflits allant du niveau inférieur vers le niveau supérieur. Ainsi, si l'on considère deux Topic Maps contextualisées $TMC_1$ et $TMC_2$, nous comparons d'abord les paramètres de contextes et les termes non contextualisés de $TMC_1$ avec ceux de $TMC_2$, et traitons les conflits décelés. Par la suite, vient le tour des contextes puis le tour des topics contextualisés. Dans l'état actuel de notre recherche, les comparaisons effectuées sont purement syntaxiques. En

d'autres termes, deux topics quelconques ne portant pas le même nom sont considérés comme non similaires. Nous prévoyons de traiter les similarités sémantiques entre topics dans nos travaux futurs.

Nous proposons donc de démarrer notre fusion par une création d'une Topic Map $TMC_3$ image de $TMC_2$, avant de procéder à la fusion des structures contextuelles et des réseaux sémantiques (voir figure 5.16). Cette étape constitue l'étape de pré-intégration.

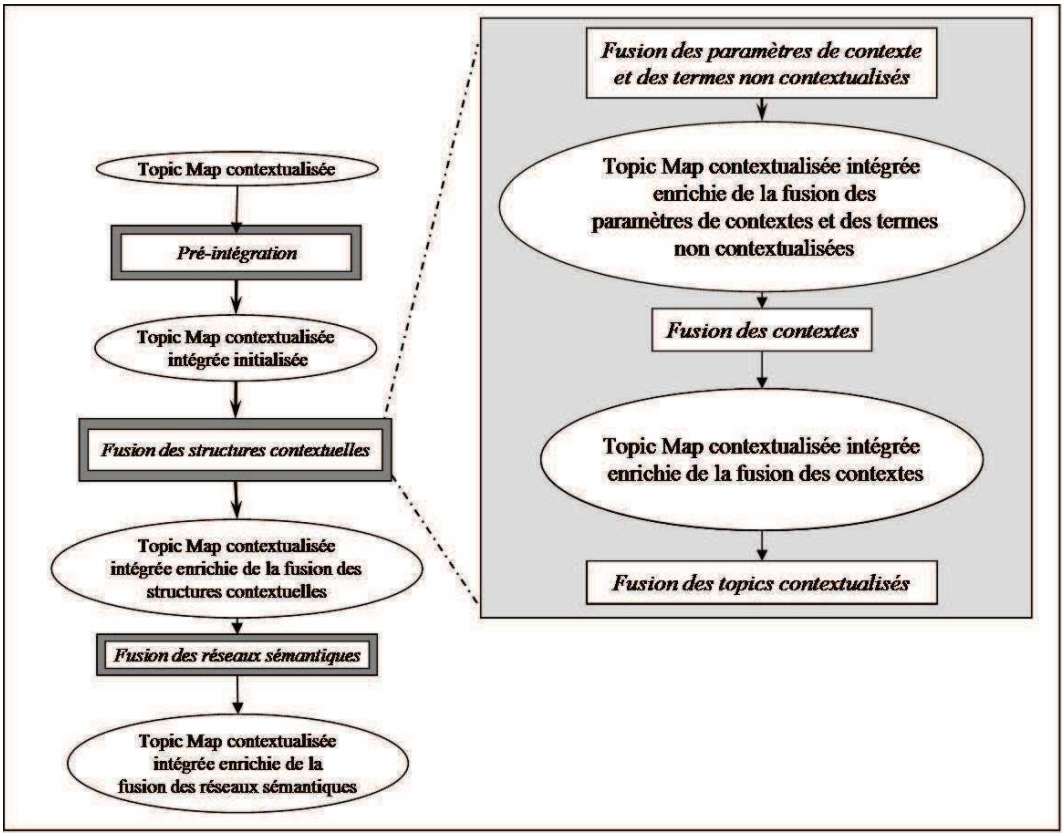


FIGURE 5.16 – Processus de fusion de Topic Maps contextualisées

L'étape suivante concerne la fusion des structures contextuelles. C'est une étape à trois phases. Elle épouse l'architecture de la Topic Map du point de vue contextualisation. Sa première phase concerne la fusion des paramètres de contexte et des

termes non contextualisés. Elle se réduit à l'enrichissement de $TMC_3$ par tous les paramètres de contexte et tous les termes non contextualisés de $TMC_2$ ne se trouvant pas dans $TMC_1$.

La seconde phase est dédiée aux appariements des contextes. À l'issue de cette comparaison, nous pouvons aboutir aux situations suivantes :

a) deux contextes appartenant chacun à $TMC_1$ et $TMC_2$ portent le même nom et sont construits à partir du même ensemble de paramètres de contexte.

b) deux contextes appartenant chacun à $TMC_1$ et $TMC_2$ portent le même nom mais ne sont pas construits à partir du même ensemble de paramètres de contexte.

c) deux contextes appartenant chacun à $TMC_1$ et $TMC_2$ ne portent pas le même nom mais sont construits à partir des mêmes paramètres de contexte.

d) deux contextes appartenant chacun à $TMC_1$ et $TMC_2$ ne portent pas le même nom et ne sont pas construits à partir des mêmes paramètres de contexte.

Les deux situations de conflit sont celles correspondant aux situations b) et c). Pour ce qui est de la situation b), le processus de fusion assigne un autre nom au contexte de $TMC_2$. Ce nom n'existe ni dans $TMC_1$ ni dans $TMC_2$. La situation c) est résolue par l'attribution au contexte de $TMC_2$, du nom associé au contexte de $TMC_1$.

Une fois toutes les situations de conflit associées aux contextes résolues, le processus ajoute à $TMC_3$ tous les contextes existants dans $TMC_2$ mais qui ne sont pas dans $TMC_1$.

La dernière phase de l'étape de fusion de la structure contextuelle consiste en la détection et le traitement des conflits existants entre topics contextualisés. Comme pour la phase précédente, l'appariement des topics contextualisés peut mener aux situations suivantes :

e) deux topics contextualisés appartenant chacun à $TMC_1$ et $TMC_2$ portent le même nom et réifient des associations identiques (associations liant les mêmes contextes et les mêmes termes non contextualisés),

f) deux topics contextualisés appartenant chacun à $TMC_1$ et $TMC_2$ portent le même nom mais ne réifient pas des associations identiques,

g) deux topics contextualisés appartenant chacun à $TMC_1$ et $TMC_2$ ne portent pas le même nom et mais réifient les mêmes associations,

h) deux topics contextualisés appartenant chacun à $TMC_1$ et $TMC_2$ ne portent pas le même nom et ne réifient pas les mêmes associations.

Seules les situations f) et g) sont considérées conflictuelles. La résolution du conflit associé à la situation f) passe par l'attribution d'un autre nom au topic contextualisé de $TMC_2$. Dans le cas où la situation g) se présente, le processus attribue au topic contextualisé de $TMC_2$, le nom associé au topic contextualisé de $TMC_1$.

Une fois toutes les situations de conflits associées aux topics contextualisés résolues, le processus ajoute à $TMC_3$ tous les topics contextualisés existants dans $TMC_2$ mais qui ne sont pas dans $TMC_1$. Il est évident que les topics contextualisés sont ajoutés avec leurs occurrences dans $TMC_2$.

La dernière étape du processus, concerne la génération du réseau sémantique intégré. Comme pour les différents types de topics, l'appariement des liens d'association entre topics contextualisés est purement syntaxique. Des évolutions de cette étape sont à prévoir dans nos prochains travaux de recherche. La fusion du réseau sémantique se réduit donc à la récupération, dans la Topic Map $TMC_3$, de toutes les associations existantes dans $TMC_2$ entre topics contextualisés mais absente dans $TMC_1$.

## 5.9 Conclusion

Ce chapitre constitue l'un des piliers de cette thèse. Il réunit une grande partie de nos contributions de recherche dans l'organisation de contenus hétérogènes. Il a permis de décrire un nouveau modèle de Topic Map permettant l'organisation et l'accès à des contenus multilingues et pluridisciplinaires et un ensemble de méthodes

pour sa construction et son évolution. Ces méthodes ont été conçues par assemblage d'approches élémentaires, en exploitant les principes de l'IDM. D'autres approches concernant la construction d'une Topic Map peuvent être envisagées. Elles n'ont pas fait l'objet de ce travail de recherche. De plus, uniquement quelques mécanismes d'évolution de Topic Map ont été conçus. Ils concernent tous l'aspect enrichissement d'une Topic Map. Les mécanismes d'appauvrissement tels que l'élagage d'éléments de Topic Map induits de l'élagage de ressources ou encore ceux induisant l'élagage de ressources ne font pas partie de cette thèse.

En revanche, les approches permettant la visualisation d'une Topic Map contextualisée font l'objet du prochain chapitre.

# Chapitre 6

# Visualisation d'une Topic Map contextualisée

Le modèle de Topic Map permet de rassembler des ressources informationnelles autour d'une structure riche sémantiquement. Cette structure sous forme de réseau sémantique visualisable, la plupart du temps, par le biais de representations graphiques favorise la navigation au sein des contenus et rend selon, l'étude faite dans [Dudycz, 2011], la recherche d'informations utile.

Le modèle de Topic Maps contextualisées proposé, dans cette thèse, pour l'organisation de contenus pluridisciplinaires et multilingues n'est exploitable qu'à côté d'une représentation qui éviterait les effets néfastes (lisibilité moindre du graphe) liés à la démultiplication de topics dûe à la contextualisation des sujets.

Le présent chapitre tente de proposer une solution de représentation graphique dédiée en introduisant le concept de vue synthétique. Ce concept repose sur la combinaison de deux techniques de visualisation. La première est fondée sur la classification. La seconde est une technique orientée graphe. La mise en œuvre de cette solution fait partie du chapitre 7.

Ce chapitre présente aussi une approche de construction et de gestion de vues personnalisées qui contribue à la lisibilité du réseau sémantique et qui offre à l'utili-

sateur une Topic Map partielle adaptée à ses besoins en recherche d'informations.

La section 1 de ce chapitre est donc dédiée à une présentation des techniques de visualisation existantes dans la littérature. Elle est suivie de la description du concept de vue synthétique et de son integration dans le modèle de Topic Map contextualisé. La section 3 de ce chapitre présente, après une description du concept de vue personnalisée, l'approche de construction et de gestion de vues personnalisées préconisée.

## 6.1 État de l'art sur les techniques de visualisation

Il existe plusieurs travaux de recherche sur le rendu graphique d'un graphe sémantique. Les graphes sémantiques peuvent être présentés sous forme de classification comme ceux présentés dans [Noy et al., 2001], [Shneiderman, 1992], [Le Grand and Soto, 2003] ou sous forme graphique comme celui de [ontopia, 2001].

### 6.1.1 La visualisation orientée classification

Cette catégorie se concentre sur la visualisation de la classification, de sa hiérarchie et des relations qui existent entre les différentes catégories.

Ce type de visualisation peut présenter les classifications sous forme de liste indentée. Son principe est de mettre en évidence les relations hiérarchiques entre les classes et parfois leurs instances.

Ces procédés vont du simple inventaire des types disponibles à des représentations plus avancées mettant, par exemple, en relief les classes selon le nombre de concepts qui les utilise.

La figure 6.1 montre un exemple de visualisation dans sa forme verticale. Ce type de visualisation a l'avantage de prendre peu de place et reste très efficace pour la navigation au sein des classifications. Des outils tels que Protégé [Noy et al., 2001] ou OntoEdit [Sure et al., 2003] l'ont adoptés.

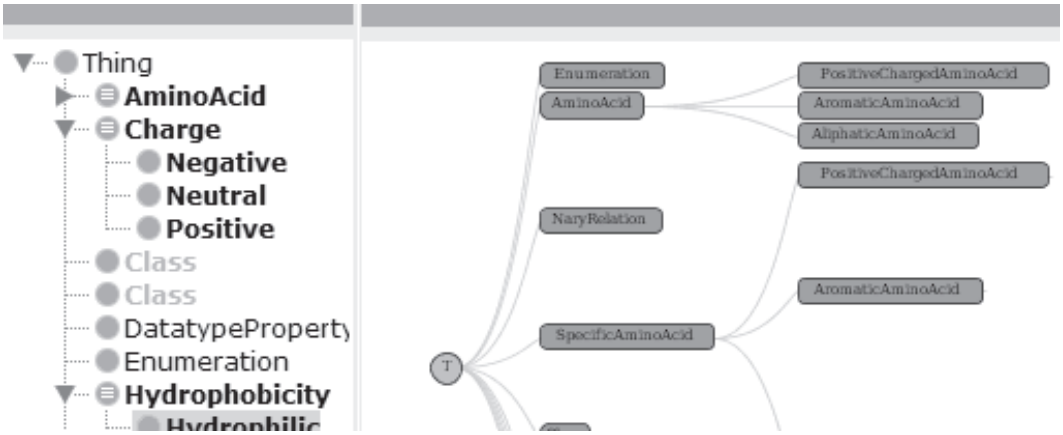


FIGURE 6.1 – Techniques de visualisation selon les listes indentées

Une autre technique de la même catégorie consiste en la subdivision de l'espace disponible en fonction de l'importance des classes. La taille de chaque zone traduit le nombre de sous-classes ou d'instances d'une classe particulière. Cette technique est utile pour des utilisateurs souhaitant visualiser et analyser les principales classes à travers des regroupements particuliers (figure 6.2). Des outils tels que TreeMaps [Shneiderman, 1992] ou Information Slices [Andrews and Heidegger, 1998] proposent ce type de visualisation dont la compréhension n'est pas des plus intuitives.

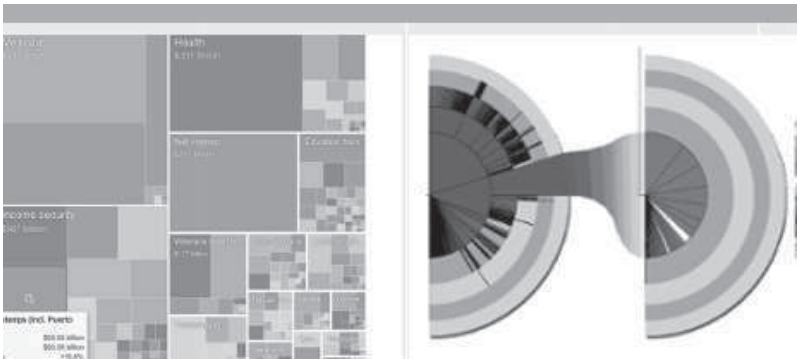

FIGURE 6.2 – Techniques de visualisation remplissant l'espace

La figure 7.4 illustre à travers l'application CropCircles [Parsia et al., 2005] une autre technique de visualisation orientée classification qui consiste à faire varier le

niveau de détail dans la présentation ; ceci en affichant à l'utilisateur selon ses préférences, une vue détaillée d'un nœud ou d'une zone spécifique.

En revanche, la représentation de la terminologie dans son ensemble n'est pas l'objectif premier de cette approche. L'agrandissement entraîne l'affichage de données supplémentaires, tandis que la réduction les regroupe progressivement. La taille des ensembles dépend du nombre d'éléments qu'ils englobent.

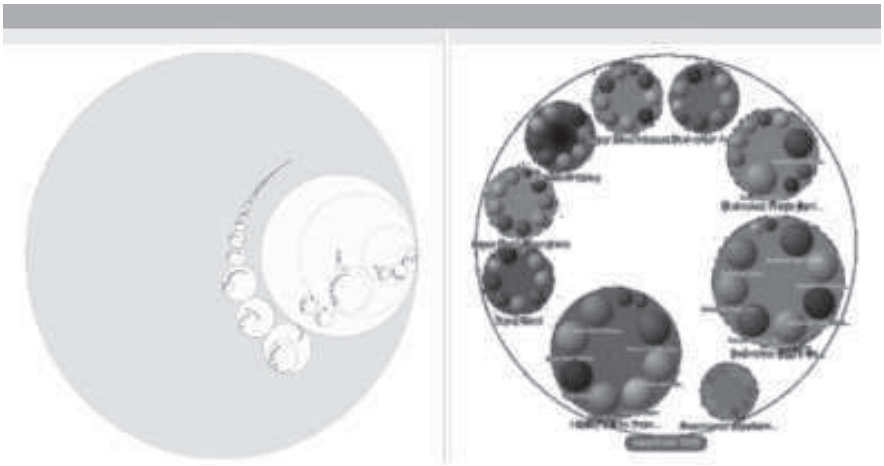

FIGURE 6.3 – Technique de visualisation utilisant un zoom (CropCircle)

La dernière technique de visualisation est appelée « paysage ». Elle répartit les objets dans une structure navigable telle qu'une carte 2D ou un monde en 3D. Les instances ou les classes sont symbolisées par des blocs dont la taille reflète l'importance en regard de divers critères.

Dans la figure 6.4, qui représente une visualisation avec le prototype Virtual City [Le Grand and Soto, 2003] illustre cette technique qui se combine avec des niveaux de détails variés et une vue de haut. Ce genre de fonctionnalité complémentaire semble essentiel pour naviguer plus efficacement.

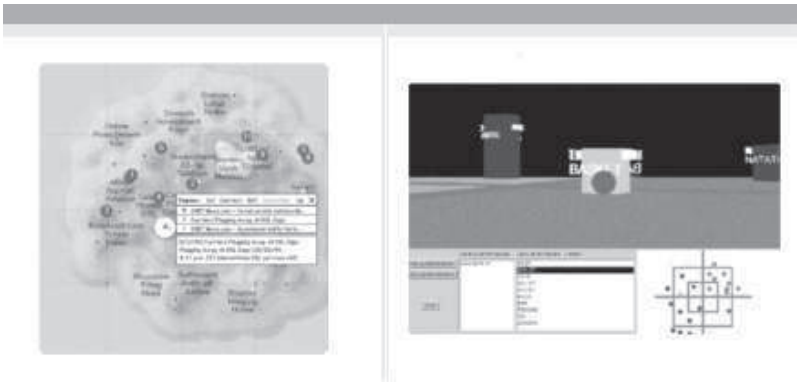

FIGURE 6.4 – Techniques de visualisation basées sur la métaphore du paysage

### 6.1.2 La visualisation orientée graphe

Ce type de visualisation matérialise les constituants du réseau sémantique sous forme de nœuds et d'arêtes. Cette approche, plus générique, peut également s'appliquer à la classification.

La représentation peut être sous forme de graphe simple (voir figure 6.5). Elle se prête bien à l'affichage d'un réseau sémantique. Toutefois, l'agencement des nœuds et des métadonnées devient crucial pour conserver une bonne visibilité, d'où son appellation : « technique orienté graphe distordu ».

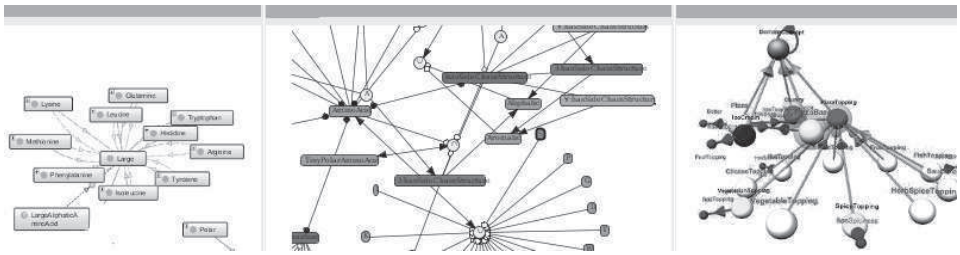

FIGURE 6.5 – Techniques de visualisation utilisant des graphes 2D ou 3D

À cet effet, différents algorithmes de placement automatique des nœuds sont généralement proposés pour obtenir diverses mises en page. L'espace est, cependant, rarement exploité de manière optimale. Dès que le nombre de nœuds augmente, la

visibilité en pâtit. Par conséquent, cette technique orientée graphe distordu, n'adresse que des graphes de petites tailles ou des graphes dont on ne veut visualiser qu'une partie réduite.

Une seconde technique de visualisation graphique, appelée graphe de distorsion (voir figure 6.6), exploite un graphe dont le rendu est déformé afin d'afficher un nœud et son voisinage le plus ou moins proche. La distorsion s'applique sur la taille des nœuds et leurs relations qui sont placés au sein d'un disque de visualisation. Les nœuds sélectionnés ou détaillés sont placés au centre et agrandis ; les nœuds environnants sont progressivement réduits en fonction de leur proximité, jusqu'à disparaître en atteignant la limite du disque. Cet agencement permet ainsi de visualiser un grand nombre d'éléments et de naviguer efficacement dans le graphe ; ce qui présente un intérêt non négligeable. Il faut toutefois noter que le réajustement continu de la taille des nœuds et leur déplacement régulier peut troubler certains utilisateurs. Néanmoins, cette technique s'avère appropriée et couramment, recommandée pour l'affichage des instances d'un réseau sémantique.

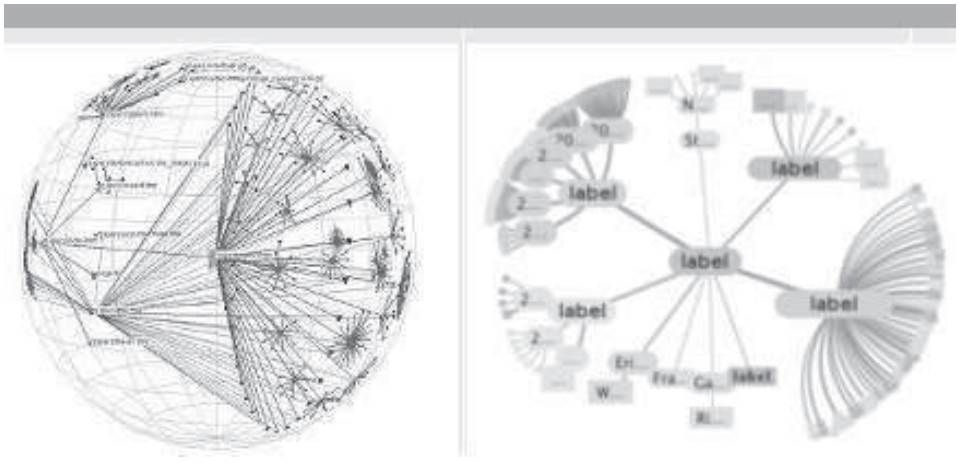


FIGURE 6.6 – Techniques de visualisation utilisant des graphes distordus (3D et 2D)

Ce principe de visualisation utilise des algorithmes, dits hyperboliques, permettant de placer de nombreux nœuds sur l'écran sans réduire la visibilité.

## 6.2 Approche de visualisation d'une Topic Map contextualisée

Ce bref état de l'art permet d'aborder les problématiques qui touchent, plus spécialement, à la technologie Topic Maps et au principe de contextualisation.

Cette contextualisation des termes entraîne une démultiplication des concepts qui ne doit pas nuire à la bonne visibilité du graphe. Une solution de cette problématique peut prendre forme avec un regroupement des termes contextualisés au sein de leur terme de base respectif. Cela aboutit à une vue synthétique faisant abstraction des contextes et permettant de présenter un graphe simplifié et compréhensible. Cette agglomération exige, toutefois, l'accès à différents niveaux de détails pour retrouver les informations contextuelles sous-jacentes, c'est-à-dire les valeurs des paramètres de contexte.

La visualisation de réseaux sémantiques est reconnue être relativement bien gérée par les graphes distordus [Katifori et al., 2007]. Nous proposons donc d'utiliser cette technique, pour représenter une vue synthétique de la Topic Map contextualisée. L'accès aux différents niveaux de détail pour retrouver les informations contextuelles pourrait se faire moyennant la technique de visualisation en liste indentée. Cette dernière s'avère, selon [Katifori et al., 2007], la plus adaptée pour représenter des agencements arborescents.

Cette approche, qui combine deux techniques de visualisation, offre ainsi divers moyens pour appréhender la contextualisation des termes. L'exemple de la figure 6.7 illustre l'utilisation de cette approche pour une Topic Map contextualisée.

Dans cette figure, nous présentons une vue synthétique composée des « terme 1 » et « terme 2 ». Celle-ci est créée à partir du regroupement des associations des quatre termes contextualisés sous-jacents. L'utilisateur peut ensuite obtenir un niveau de détail supplémentaire par le biais de boites de dialogues précisant les informations contextuelles. Notons que la vue détaillée reste disponible, si un usager souhaite

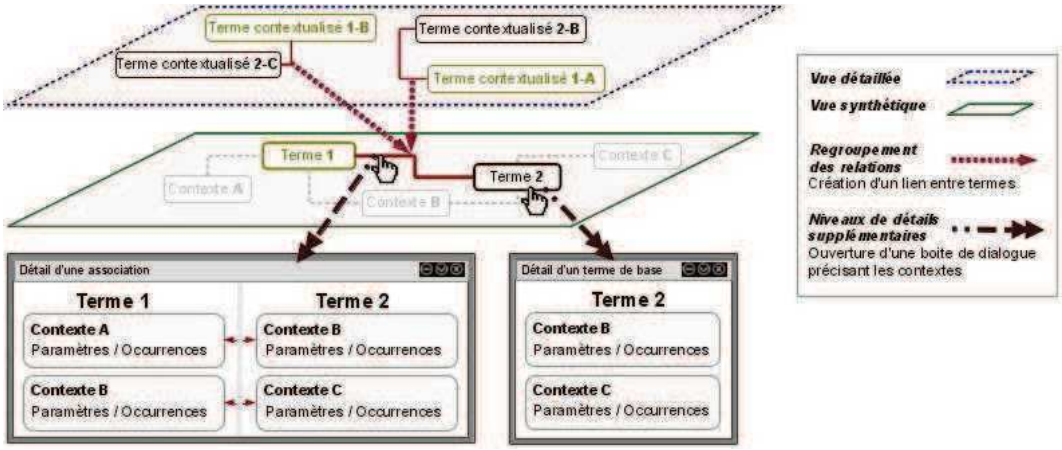


FIGURE 6.7 – Principe de la vue synthétique d'une Topic Map contextualisée

afficher directement les termes contextualisés et leurs relations sous forme de graphe.

La mise en œuvre de cette approche suppose une intégration de la vue synthétique dans le méta-modèle contextuel. La modélisation correspondante est décrite dans le paragraphe suivant. Elle respecte le cadre imposé par le TMDM.

### 6.2.1 L'intégration des vues synthétiques dans le modèle de Topic Maps contextualisées

Dans une vue synthétique, toutes les relations entre termes contextualisés sont rassemblées dans un lien unique entre les termes de base concernés (les termes non contextualisés). Cette liaison résume ainsi plusieurs associations contextuelles et traduit la proximité sémantique des termes « non contextualisés ». Ce système de regroupement s'applique aussi bien à des associations entre termes contextualisés construits à partir de mêmes termes « non contextualisés », qu'à des associations n-aires mettant en jeu plus de deux ou plusieurs termes contextualisés. Ainsi, comme le montre la figure 6.8, lorsque deux termes ont été contextualisés à partir d'un même terme « non contextualisé », la vue synthétique correspondante est une association reflexive entre les termes contextualisés. De même, deux liens binaires $\{(T_1C_x, T_2C_x), (T_1C_y,$

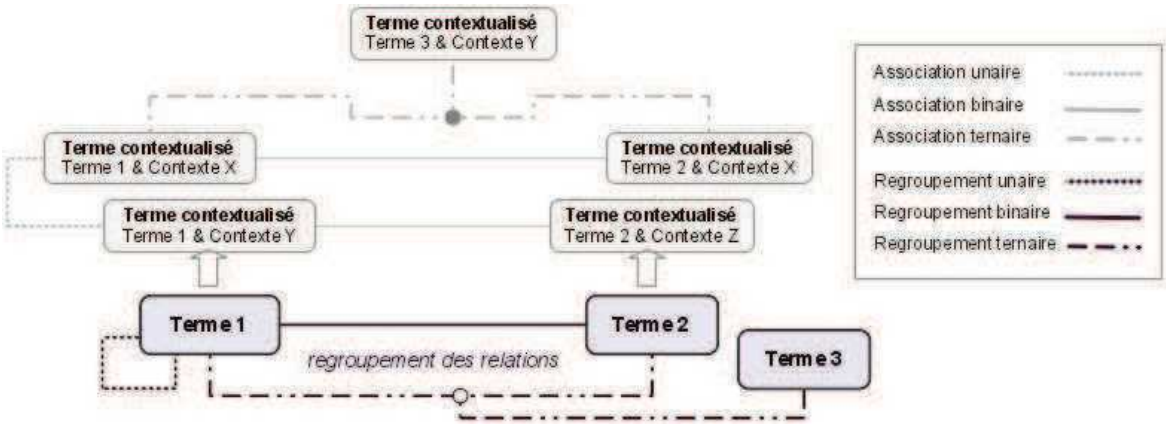


FIGURE 6.8 – Principe de regroupement des termes contextualisés et de leurs relations

$T_2C_z$)} entraînent la création d'une nouvelle association entre les termes T1 et T2.

Pour qu'une vue synthétique puisse être visualisée, il faut la stocker et donc l'intégrer dans le méta-modèle contextuel. Il faut aussi la lier, dans le modèle, à la portion de Topic Map qu'elle synthétise. Ce lien est réalisé moyennant l'exploitation du mécanisme de réification. En d'autres termes, un topic réifiant pourra être créé. Il comportera, tel qu'illustrer par la figure 6.9, autant d'associations n-aires que de liaisons entre termes contextualisés constituant la vue synthétique. De plus,

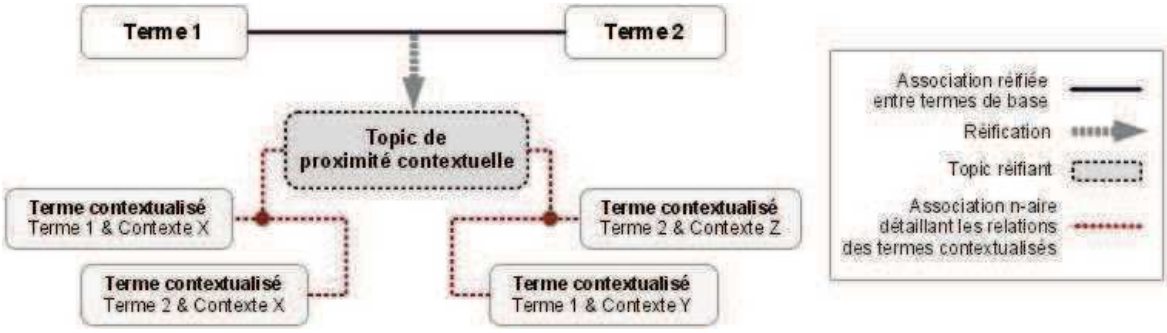


FIGURE 6.9 – Principe de réification des associations entre termes

afin de conserver la nature des relations, chacune d'elles reprend sa catégorisation originale et donc le même type. Pour détailler une union entre termes, il suffit alors

de récupérer le topic réifiant et d'accéder aux membres de ses associations.

Cette solution est conforme au modèle standard des Topic Maps, puisqu'elle est fondée uniquement sur la réification. Elle a été privilégiée, de surcroît, parce qu'elle gère tous les types envisageables d'associations.

L'intégration de la vue synthétique dans le méta-modèle contextuel est matérialisée par l'ajout de trois concepts : le concept de « TermLink », le concept de « RelatedTerm » ainsi que le concept de « RelatedTermLink » (voir figure 6.10). Le concept TermLink permet de réunir toutes les associations contribuant à la construction de la vue synthétique. Le concept « relatedTerm » permet de modéliser le topic réifiant qui a permis de construire la vue synthétique. Enfin, le concept « RelatedTermLink » permet de détailler chacune des associations sous-jacentes en liant le topic réifiant avec les termes contextualisés concernés.

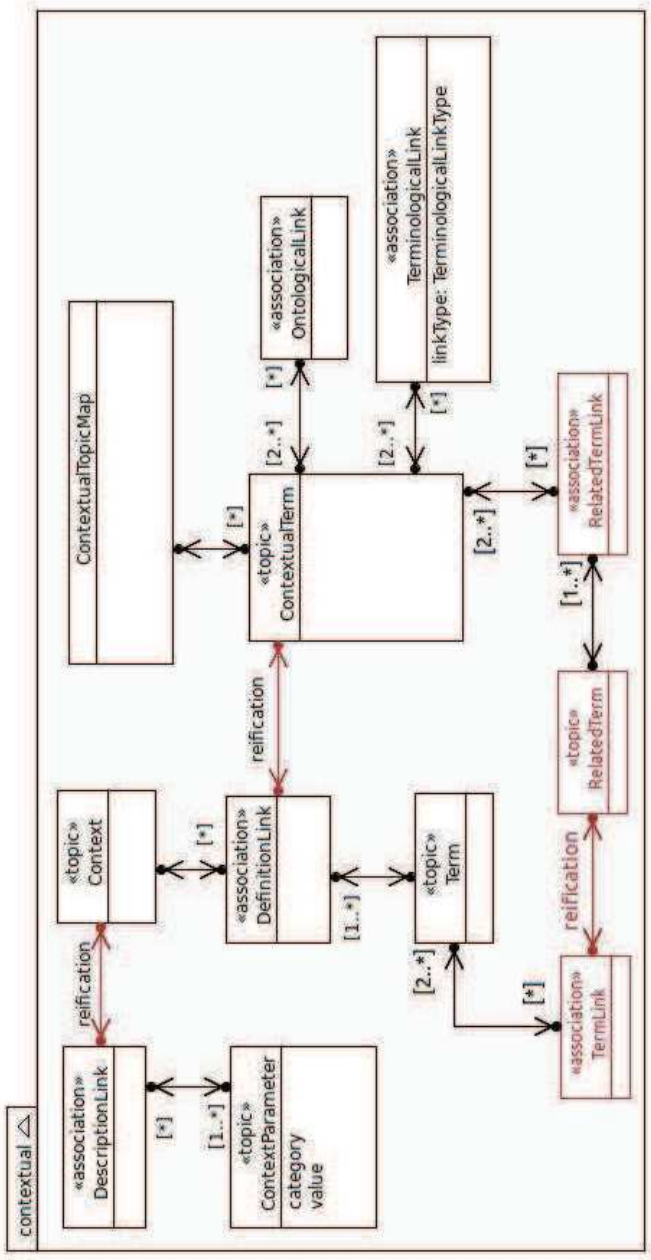


FIGURE 6.10 – Métamodèle contextuel intégrant la vue synthétique

L'algorithme permettant de créer et/ou de modifier automatiquement des vues synthétiques est fourni en l'annexe C de cette thèse. Il est exécuté lors du peuplement de la topic map. Il consiste à instancier ou à modifier les classes représentant les concepts associés à une vue synthétique. La modification des classes est opérée lorsqu'une vue synthétique existe et qu'il s'agit simplement de l'enrichir à travers l'ajout de détail ou l'ajout d'associations entre terme contextualisés.

## 6.3 Personnalisation d'une Topic Map contextualisée

L'idée véhiculée à travers cette notion de personnalisation d'une Topic Map est qu'un utilisateur peut, de par ses centres d'intérêt, vouloir focaliser sa recherche d'informations sur une portion du contenu qu'il juge lui-même pertinent à explorer. Par conséquent, offrir un contenu filtré à priori selon le centre d'intêret d'un utilisateur est, à notre avis, primordiale. Ce filtrage de contenu à priori induit un filtrage de la Topic Map au moment de sa visualisation. Cette fonctionnalité, selon [Shneiderman, 1996], est l'une des fonctionnalités que doit supporter un système de visualisation efficient et adapté à des ressources sémantiques.

Nous proposons donc une approche de personnalisation d'une Topic Map dont l'objectif est la facilitation de l'obtention de ses différents points de vue. De façon naturelle, nous proposons un filtrage fondé sur les paramètres de contexte. Un point de vue de la Topic Map sera, dans ce cas, une vue de la Topic Map selectionnée sur la base de valeurs attribuées à un ou plusieurs paramètres de contexte. A titre d'exemple, si un utilisateur souhaite se focaliser sur un contenu français et arabe relatif à la psychologie, le système est améne à lui fournir la Topic Map qui permettrait de naviguer au sein de ce contenu. Pour cela, il procède au filtrage de la Topic Map en ne considérant que les termes contextualisés qui ont pour valeur du paramètre « langue » la valeur « arabe » ou la valeur « français » et pour valeur de paramètre discipline la valeur « psychologie ».

Ce type de besoin aboutit à la définition d'un autre concept qui est celui de

« vue personnalisée » qui intègre dans sa définition la notion de vue synthétique. En d'autres termes, une vue personnalisée, au même titre qu'une Topic Map globale, peut être visualisée sous forme de vue synthétique.

Le volume d'une Topic Map pouvant être conséquent, le procédé de filtrage peut causer des temps de latence importants avant l'affichage d'une vue. Pour que les analyses selon différents critères puissent se faire sereinement, il faut veiller à ce que les traitements restent efficients. L'optimisation des performances passe, entre autres, par la sauvegarde des vues déjà demandées dans le but d'éviter des traitements inutiles, lorsqu'elles seront, de nouveau, sollicitées. La figure 6.11 illustre ce mécanisme de cache orchestrant la création, l'enregistrement et la récupération des vues archivées. La création des vues se fait de manière transparente pour l'utilisateur. Lorsqu'il sélectionne les paramètres qu'il souhaite retenir, le système tente de trouver une vue déjà enregistrée et, le cas échéant, en créer une nouvelle qu'il met en archive.

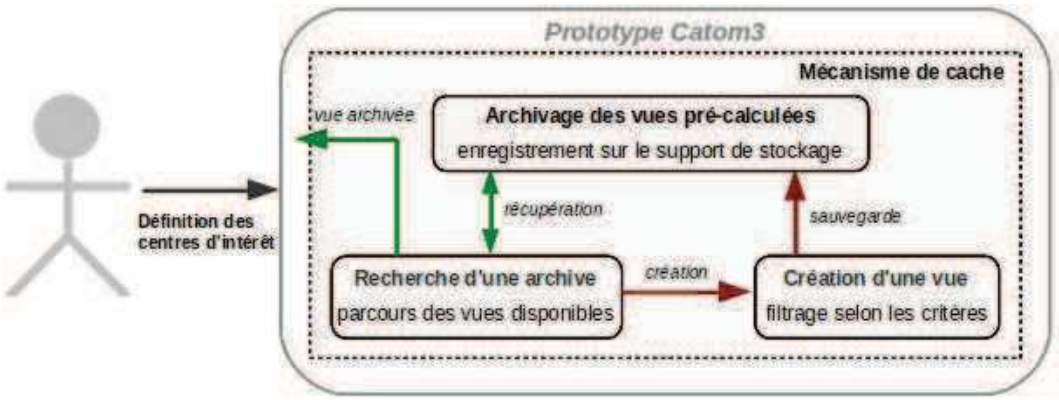


FIGURE 6.11 – Mécanisme de cache des vues personnalisées

Ce mécanisme exploite la méthode d'analyse formelle de concepts (AFC) qui permet d'améliorer les processus de recherche et de création des vues. Cette méthode [Bernhard and Rudolf, 1999] est une approche de structuration hiérarchique de différents sous-ensembles. Elle prend pour base un contexte, dit formel, qui englobe

| | Histoire | Sociologie | XIV siècle |
|---|---|---|---|
| TopicMap (tm) | X | X | X |
| Vue A | X | | |
| Vue B | | X | X |
| Vue C | X | X | |

TABLE 6.1 – Formalisation d'un contexte formel par la méthode AFC

un ensemble d'objets partageant des attributs communs. Le tableau 6.1 fournit un exemple de contexte formel basé sur notre problématique de hiérarchisation des vues. Les colonnes du tableau représentent les attributs formels et les lignes des objets formels. À partir de ce contexte, des concepts formels sont identifiés. Chacun d'eux correspond à une combinaison unique d'objets liés à l'ensemble des attributs qu'ils ont en commun. Pour reprendre l'exemple du tableau 6.1, seuls les objets « tm » et « vue C » utilisent les attributs « histoire » et « sociologie ». Cela conduit à la création d'un concept composé des objets « (tm, vue A) » et ses attributs « (histoire, sociologie) ». De la même façon, les objets « tm » et « vue B » partagent les caractéristiques « sociologie » et « XIV siècle » ce qui donne un nouveau concept dont, pour reprendre la terminologie en vigueur, l'extension est « (tm, vue B) » et l'intention « (sociologie, XIV siècle) ». Les concepts identifiés sont ensuite hiérarchisés dans une structure particulière nommée « treillis de concepts ». Un treillis est un ensemble partiellement ordonné dans lequel tout couple d'éléments possède toujours une borne supérieure et une borne inférieure. Cette structuration s'appuie sur la théorie des relations d'ordre. Elle produit un ordonnancement hiérarchique des sous-ensembles qui peut être visualisé dans un graphe dont les nœuds représentent les concepts. Ceux-ci sont agencés en fonction de leur rapport d'inclusion du type $concept1 \subseteq concept2$. La figure 6.12 montre le treillis obtenu à partir du contexte formel présenté dans le tableau 6.1. Comme le souligne cette figure, le treillis répartit les concepts en fonction de leurs attributs, et ce, du plus petit ensemble (vide) à

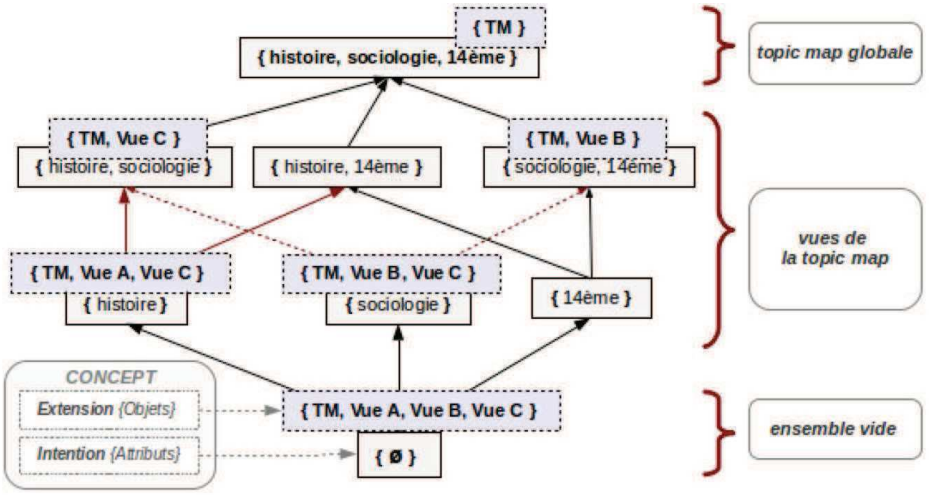


FIGURE 6.12 – Structure hiérarchique du treillis issu d'un contexte formel

l'ensemble le plus grand (Topic Map globale). Cette structure, ainsi que le contexte formel qui en est la source, ouvre diverses perspectives d'optimisation des processus de création et de recherche des vues.

### 6.3.1 L'approche de création d'une vue personnalisée

La production d'une nouvelle vue peut se faire en filtrant systématiquement la Topic Map globale. Cependant, dans le cas d'une structure dense composée de très nombreux topics, cette approche s'avère peu efficace et coûteuse en temps de traitement. Une autre démarche consiste à rechercher si une vue représentant un sous-ensemble juste un peu plus grand a déjà été archivée. Si cette vue existe, l'algorithme peut se baser sur celle-ci et limiter, de cette façon, le nombre d'éléments à traiter. La figure 6.13 répertorie les principales étapes de création optimisée d'une vue. L'élaboration commence par l'insertion des paramètres de contexte sélectionnés dans le contexte formel. Dans un second temps, le treillis correspondant est calculé afin d'identifier le sous-ensemble le plus proche en terme d'attributs. Les traitements nécessaires au filtrage sont ensuite effectués avant d'archiver la vue demandée.

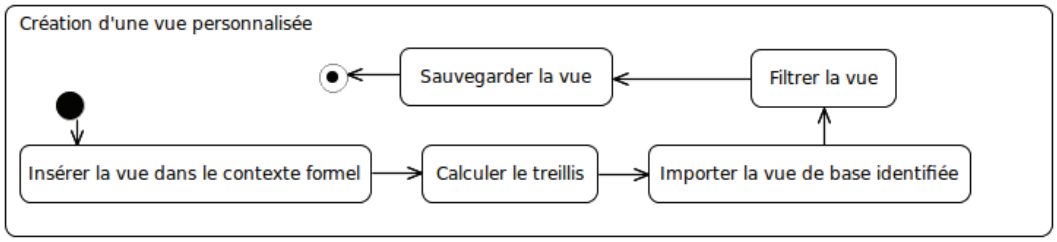


FIGURE 6.13 – Processus de création d'une vue basé sur la méthode AFC

### 6.3.2 La recherche de vues personnalisées disponibles

Lorsqu'un utilisateur soumet une série de paramètres de contexte, le système tente, en premier lieu, de rechercher si une vue correspondante a préalablement été archivée. Un contexte formel de la méthode AFC, comme celui du tableau 6.1, fournit alors une structure appropriée pour identifier rapidement les vues disponibles.

Comme il est indiqué plus haut, ce contexte formel est constamment mis à jour lors de la création d'une nouvelle vue et comporte donc toutes les informations nécessaires à la sélection de vues représentant un sous-ensemble égal ou juste un peu plus grand que celui recherché.

L'algorithme précise qu'il faut tester les attributs associés aux objets afin de vérifier qu'ils font bien tous partie des paramètres sélectionnés. Si c'est le cas, le système en déduit que la vue a déjà été archivée. L'algorithme correspondant à cette recherche de vues personnalisée est fourni ci-dessous.

**Algorithm 1** Recherche de Vue Disponible

Recherche si une vue est disponible dans le contexte formel en fonction des paramètres selectionnés
- **Entrées :** parametres : Liste de chaine, contexteFormel : Liste de Liste de chaine objet/attributs
- **Sorties :** chemin de la vue : chaine chemin de la vue sur le support de stockage
- classe Liste

  **attributs** entête : ListElement premier element , curseur : ListElement element en cours

  **fonctions** suivant(), premier(), dernier(), nombreElement(), horsListe() sous-algorithmes non detailles
- **type** ListElement = agrégat
- valeur : Info type de base ou complexe ,
- suivant : ListElement reference a un autre element

  **variables**
- *objet : Liste de Chaine contient les attributs de l'objet*
- *attribut : Chaine , attributeFound : Booléen , nbAttributeFound : Entier*

```
while contexteFormel do
    - Vérification de l'ensemble des objets du contexte formel
    objet ← contexteFormel.suivant()
    nbAttributeFound ← 0) {Nécessite au moins le même nombre d'attributs que de paramètres selectionnés pour vérifier la vue}
    if objet.nombreElement()==parametres.nombreElement() then
        while non objet.horsList() do
            attribut ← objet.suivant());
            attributeFound ← Faux;
            parametres.premier(); {Réinitialise la liste avant recherche}
            {vérifie que l'attribut fait partie des paramètres selectionnés}
            while Il existe parametres do
                if parametres.suivant() = attribut then
                    attributeFound ← Vrai;
                    nbAttributeFoundFound ← nbAttributeFound + 1;
                    parametres.dernier(); {arrête de la boucle}
                end if
            end while
            if non attributeFound then
                objet.dernier(); {si un des attribut n'a pas ete trouvé, alors on teste un autre objet}
            else
                if nbAttributeFound == parametres.nombreElement() then
                    retourner objet.cheminDeLaVue ; {retour de la méta-donnees indiquant l'emplacement de la vue}
                end if
            end if
        end if
    end while
end while
```

## 6.4 Conclusion

La visualisation des Topic Maps -et plus largement des ressources sémantiques- est une tâche complexe du fait des multiples éléments qui les composent. Les différentes techniques d'affichage demandent ainsi d'être combinées pour améliorer l'expérience utilisateur et tirer pleinement profit des connaissances modélisées. Dans le

cadre de notre travail de recherche, nous avons proposé de combiner à la fois la technique de visualisation à base de listes indentées pour visualiser les classifications et la technique« orientée graphes distordus » pour le réseau sémantique que represente la Topic Map. Cette combinaison de technique permet de traiter la problématique due à la contextualisation et qui entraîne une surabondance d'informations nuisant ainsi à l'intelligibilité du graphe. Elle est appliquée à la vue synthétique de la Topic Map et aux vues personnalisées de la Topic Map. Une vue synthétique permet le regroupement de termes contextualisés afin d'élaborer une vue plus simple dite « hors contexte ». Une vue personnalisée permet de filtrer le contenu en fonction du centre d'intêret de l'utilisateur. Ce concept de vue personnalisé est un concept différent de celui de la méthode CITOM [Ellouze et al., 2012] et du modèle SocioTM [Sasha and Rudan, 2008]. Notre élagage des topics en vue de fournir une vue personnalisée repose sur le centre d'intêret de l'utilisateur. Celui de CITOM est fondé sur l'analyse des méta-données associées aux topics. Quant à celui de SocioTM, il est fondé sur des mesures de pertinence associées aux éléments de la Topic Map. Cependant, à notre avis, une exploration de la combinaison des trois approches est à étudier.

La gestion des vues personnalisées a, en outre, été optimisée par l'utilisation d'un système de cache associé à la méthode d'analyse formelle de concepts. Le système de cache, bien qu'indispensable à nos yeux, pose le problème de gestion des vues archivées. Nous prévoyons, à cet effet, renseigner les vues stockées par des métadonnées d'usage (information concernant l'utilisation des vues par les utilisateurs) nous permettant de décider de leur élagage.

Le chapitre suivant presentera la mise en œuvre de l'approche de visualisation exposée dans ce chapitre.

# Chapitre 7

# Plateforme pour la construction, l'évolution et la visualisation de Topic Maps contextualisées

Ce chapitre présente quelques aspects importants du développement de la plateforme dédiée à la construction, l'évolution, et la visualisation de Topic Maps contextualisées.

Nous nous sommes appuyés pour le développement de la première version de ce prototype sur les concepts fondamentaux de l'IDM tel que l'encrage des modèles au centre du système, la transformation de modèles et la génération d'une partie du code.

La réalisation repose sur le cadriciel EMF (Eclipse modeling Framework) [Steinberg et al., 2008] permettant de produire, manipuler et transformer les métamodèles. Une synthèse sur ce cadriciel est présentée dans l'annexe B.

Le plan de ce chapitre est structuré comme suit. La première section de ce chapitre donne une vue d'ensemble de la plateforme et décrit le contexte de son développement. Les sections suivantes présentent quelques aspects techniques liés au

développement des modules constituant l'architecture de ce prototype.

## 7.1 Présentation générale du prototype

La première version de la plateforme vise à satisfaire quelques exigences fonctionnelles décrites à travers le diagramme des cas de UML de la figure 7.1.

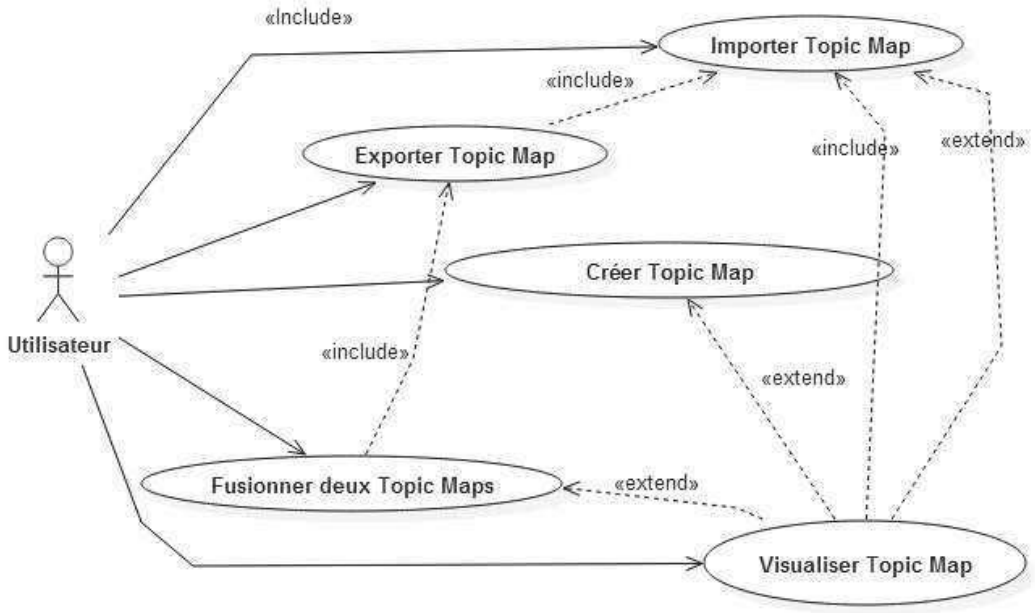


FIGURE 7.1 – Cas d'utilisation de la plateforme

Les scénarios associés aux cas d'utilisation mentionnés dans cette figure et qui ont fait l'objet de développement sont :

- le scénario nominal du cas « Importer Topic Map »
- le scénario nominal « Visualiser la vue synthétique d'une Topic Map » du cas « Visualiser »
- le scénario alternatif « visualiser la vue personnalisée avec affichage synthétique ou détaillé de la vue » du cas « Visualiser »
- le scénario nominal du cas « Exporter »

– Le scénario nominal du cas « Créer Topic Map »
– Et enfin le scénario nominal du cas « Fusionner deux Topic Maps »

Quelque uns de ces scénarios ont été développés par un ingénieur Cnam dans le cadre de la préparation de son mémoire d'ingénieur.

Dans sa prochaine version, cette plateforme pourrait renfermer, à court terme, les autres scénarios de cas décrit dans ce diagramme et à moyen terme, d'autres fonctionnalités relatives aux autres scénarios de besoins en création et en évolution décrits dans le chapitre 5 de cette thèse (création par ré-ingénierie, intégration de nouvelles ressources , etc. ).

L'architecture logicielle associée à cette version de prototype est fournit dans la figure 7.2. Dans cette architecture (voir figure 7.2) sont distinguées :

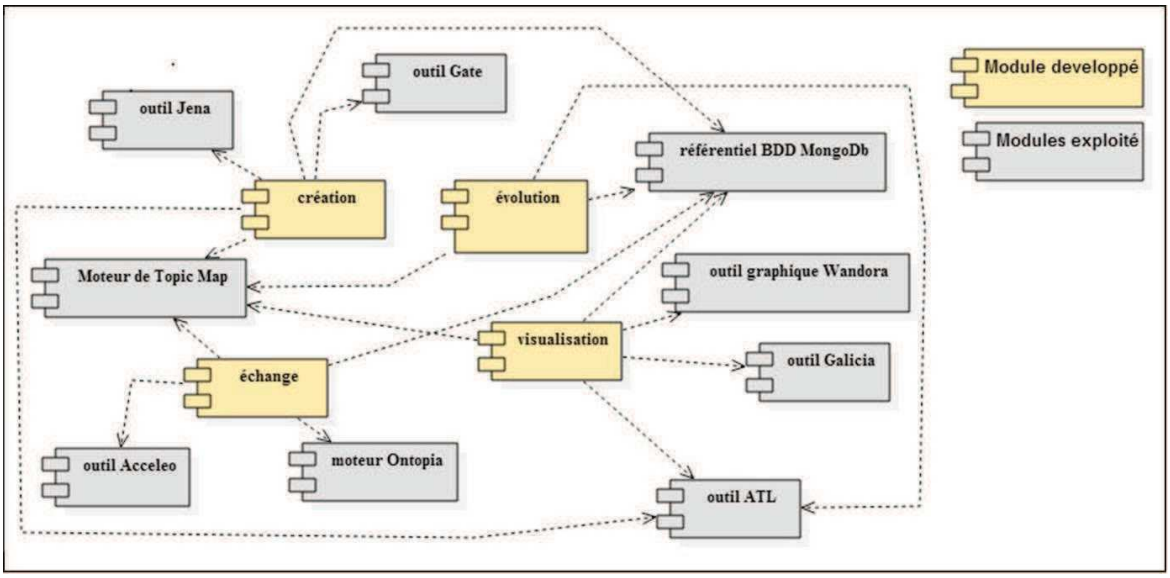


FIGURE 7.2 – Architecture logicielle

– les modules internes à la plateforme : Le moteur Topic Map, et les modules d'échange, de visualisation, de création et d'évolution
– de ceux qui sont sollicités : Wandora, Ontopia, Acceleo, ATL, Galicia, Gate, MongoDB.

Le moteur Topic Map est le cœur de la plateforme. Il permet le stockage des méta-modèles standard et contextuel sous le format Ecore et leur instanciation et manipulation. Une instance d'un méta-modèle est un fichier XMI. Rappelons que le méta-modèle standard a été enrichi par des propriétés permettant de distinguer les différents types de topics dans une Topic Map contextualisée (voir pour cela, dans le chapitre 5, la description des modules représentant respectivement le processus de génération « Type B » et « type C » de Topic Maps contextualisées)

Le module « échange » regroupe les modules « import » et « export » conçus selon les deux approches décrites dans le chapitre 5 de cette thèse. La mise en œuvre de ces modules a concerné le format d'échange XTM.

Le module import est interfacé avec l'outil Ontopia pour la réalisation des transformations T2M. Cet outil a été choisi de part sa fonctionnalité relative à la validation d'un document XTM ( un document XTM est valide s'il est conforme au standard XTM).

Le module export, quant à lui, est interfacé avec le générateur Acceleo [1] de MDA pour la réalisation des transformations M2T et la base de données MongoDb pour le stockage des documents XTM.

Acceleo implémente le standard « MOF model to text » [2] de l'Omg qui spécifie comment transformer des modèles MOF en texte. Cette norme s'appuie sur la technique de transformation basée sur des gabarits textuels. La figure 7.3 donne un aperçu d'un gabarit consacré à XTM, comportant des expressions OCL pour récupérer les données du modèle à transformer.

Le module « Visualisation » est chargé de la gestion des vues synthétiques et personnalisées. Il se charge aussi de réaliser le pont entre le prototype et l'outil de visualisation Wandora. Il exploite aussi l'outil ATL pour les transformations M2M, l'outil Galicia pour la construction des graphes de concepts et la base de données

1. http ://www.acceleo.org.
2. MOF Model to Text Transformation Language. Disponible sur http ://www.omg.org/spec/MOFM2T

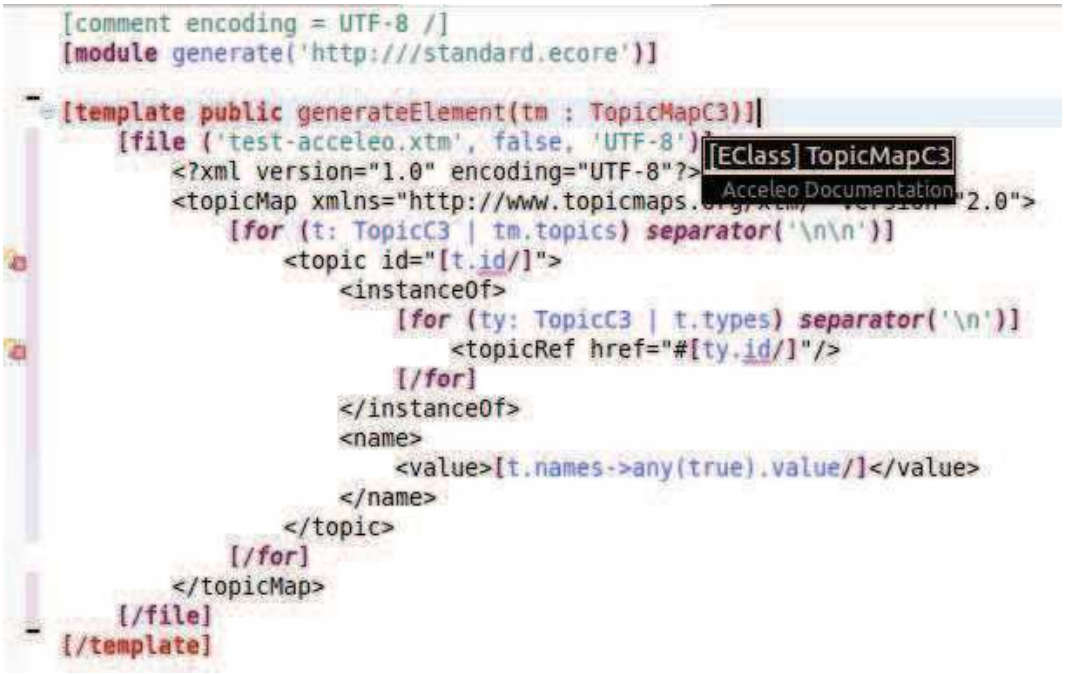


FIGURE 7.3 – Aperçu d'un gabarit Acceleo dédié au format XTM

MongoDB en tant que système de stockage de vues (globale et personnalisées) sérialisées en XTM.

Le module « Création de Topic Map » regroupera à court terme l'implémentation de toutes les approches de création décrites dans le chapitre 5. À ce stade de

développement, seule la méthode de création à partir de contenus a été mise en œuvre. Elle est nommée $CATOM^3$ (Construction Automatique de Topic Maps Multilingues et Multiculturelles). Ce module, utilise à l'heure actuelle l'outil Gate pour l'extraction de termes, MongoDb pour le stockage des Topic Maps générées et l'outil ATL pour les transformations M2M.

La même remarque peut être faite pour le module évolution qui ne regroupe pour l'instant que le module correspondant à la mise en œuvre de la fusion de Topic Maps contextualisées. Le module logiciel « Visualisation » est chargé de la gestion des vues synthétiques et personnalisées. Il se charge aussi de réaliser le pont entre

le prototype et l'outil de visualisation Wandora. Notons que dans cette version du prototype, le composant méthode « génération Type B » de Topic Maps contextualisées décrit dans le chapitre 5 a aussi été mis en œuvre pour satisfaire les besoins du module visualisation. Pour une raison de simplicité de l'architecture nous avons préféré l'intégrer pour l'instant dans le module logiciel « import » décrit ci-avant. De ce fait ce module a aussi nécessité l'utilisation d'ATL pour les transformations M2M présentées dans le composant méthode associé.

## 7.2 Description du module d'échange

Comme dit précédemment, le module d'échange regroupe deux composants : le module d'importation et le module d'exportation. Le module d'importation met en œuvre le composant méthode « export » et le composant méthode « génération type B de Topic Maps contextusalisées ». La figure 7.4 présente le méta-modèle associé

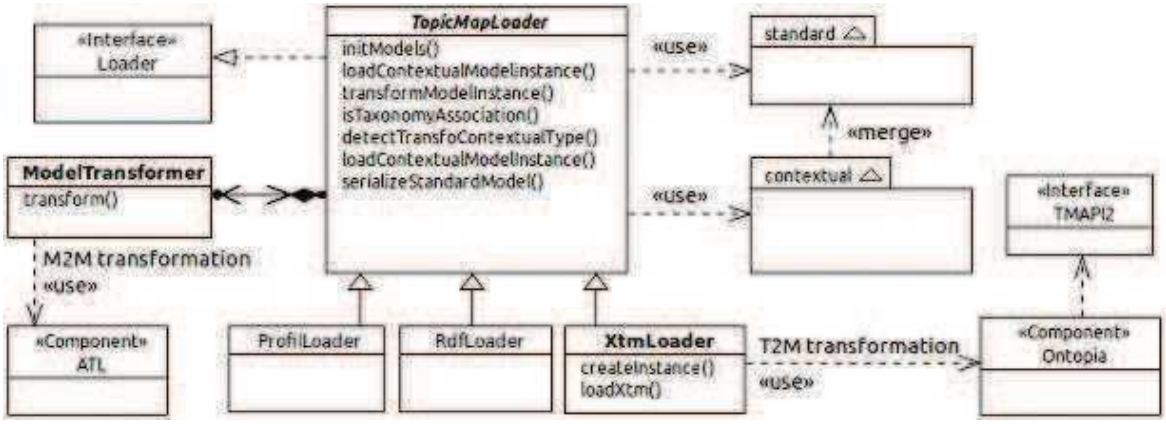


FIGURE 7.4 – Méta-modèle du module d'importation

au module « importation ».

Ce méta-modèle exploite le patron de conception « Strategy » qui facilite la prise en charge d'autres formats de sérialisation et permet ainsi d'anticiper les évolutions du module. Tous les formats disponibles étendent une même classe abstraite « To-

picMapLoader » qui concentre les comportements communs. L'ajout d'une interface (Loader) est considéré comme une bonne pratique permettant de fixer le mode d'utilisation d'un module. L'interface de programmation exposée est visible de l'extérieur et masque l'implémentation concrète des fonctionnalités. Les transformations M2M, réalisées dans le cadre de la mise en œuvre du composant méthode « génération type B de Topic Map contextusalisée », sont traitées par la classe « ModelTransformer » qui fait office de pont avec l'outil ATL. Les figures 7.5 et 7.6 sont des exemple de règles de transformation sous ATL. La première règle duplique l'ensemble des topics du modèle standard dans le modèle contextuel. La seconde règle s'applique aux associations entre un terme et un contexte. D'autres exemples de règle de transformation M2M sont fournis en annexe D de cette thèse.

```
[nom]
   Topic

[transformation]
   source Standard: Topic
   cible  Contextual: Topic
   (
      id                ← id
      itemIdentifiers   ← itemIdentifiers
      subjectIdentifiers ← subjectIdentifiers
      subjectLocators   ← subjectLocators
      occurrences       ← occurrences
      reified           ← reified
      types             ← types
      names             ← names
   )
```

FIGURE 7.5 – Exemple de première règle de transformation

Le méta-modèle associé au module d'exportation de Topic Map est présenté dans la figure 7.7. Il utilise la même structure que celle du module d'importation. Il exploite aussi le patron « Strategy » pour simplifier l'intégration de nouveaux formats. La transformation de modèles M2T est gérée par la classe « ModelTransformer » qui initialise le composant Acceleo et le gabarit nécessaire avant de lancer l'exécution du traitement.

**[nom]**
DefinitionLink

**[extension]**
Règle Association

**[condition d'application]**
Méta-donnée transfoType = 'DefinitionLink'

**[transformation]**
**source** Standard:Association
**cible** Contextual: DefinitionLink
(
context ← membre du rôle pour le topic Context
term ← membre du rôle pour le topic Term
contextualTerm ← reifier
)

**[post-traitement]**
ajouter context.definitionLinks ← cible
ajouter term.definitionLinks ← cible
contextualTerm definitionLink ← cible

FIGURE 7.6 – Exemple de deuxième règle de transformation

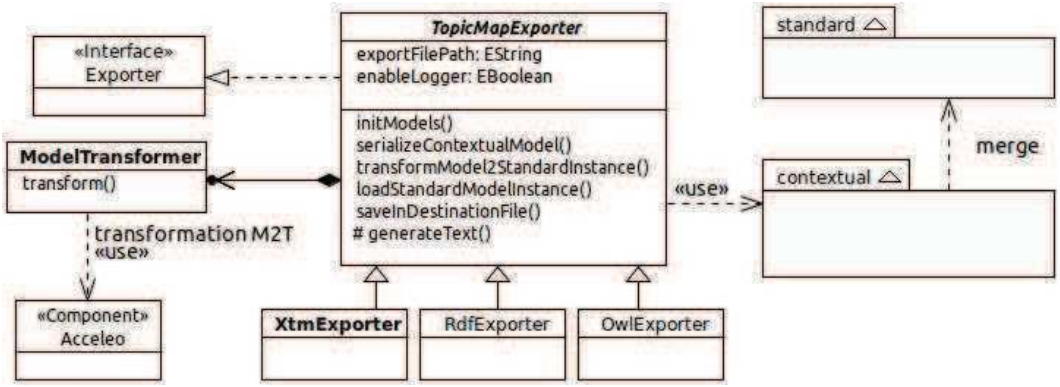


FIGURE 7.7 – Méta-modèle du module d'exportation

## 7.3 Description du module de visualisation

Une étape préalable à la mise en œuvre du module de visualisation a été le choix d'un outil permettant de prendre non seulement en charge les spécificités de notre modèle de Topic Maps mais aussi la technique de visualisation adoptée dans le chapitre précédent. Une étude comparative des outils de visualisation existants,

fondées sur les critères proposés par [Shneiderman, 1996] a été faite et a aboutit à la sélection des outils Wandora et Treebolic. Notre choix s'est finalement porté sur l'outil Wandora qui contrairement à Treebolic est dédié aux Topic Maps. C'est aussi un outil extensible qu'on peut personnaliser. L'annexe A présente cette partie du travail.

Néanmoins, Wandora a fait l'objet d'une adaptation pour, d'une part, prendre en compte la réification et, d'autre part, pour permettre l'identification de la structure contextuelle d'une Topic Map contextualisée, le chargement automatique d'une Topic Map et la proposition de différents niveaux de détails d'une Topic Map contextualisée.

Le paragraphe qui suit relate les différentes adaptations effectuées sur Wandora pour prendre en charge les exigences dictées par le modèle de Topic Maps contextualisées ainsi que celles liées à nos propositions en terme de visualisation. Il est suivi de deux paragraphes décrivant quelques aspects liés à la mise œuvre du module visualisation. Le premier paragraphe décrit l'interfaçage de Wandora étendu avec le prototype. Le second paragraphe a trait au stockage et à la gestion des vues.

### 7.3.1 Les adaptations effectuées sur l'outil Wandora

Wandora n'implémentant pas l'ensemble des spécifications du TMDM (en premier lieu la réification), la prise en charge de la réification a nécessité un ajustement de son modèle de données par ajout de deux nouvelles propriétés : « reified » et « reifier » (voir figure 7.8). Ces propriétés référencent respectivement l'élément réifié et le topic réifiant. De plus, les cardinalités précisent qu'un élément ne peut être réifié que par un unique objet de type topic et réciproquement.

L'ajout de ces propriétés a requis une modification des différents types de Topic Maps proposées dans l'outil. Cela passe notamment par l'implémentation des opérations des trois classes abstraites « Topic Map », « Topic » et « Association ». Des logiciels de rétro-conception pour cartographier l'application ont permis d'effectuer

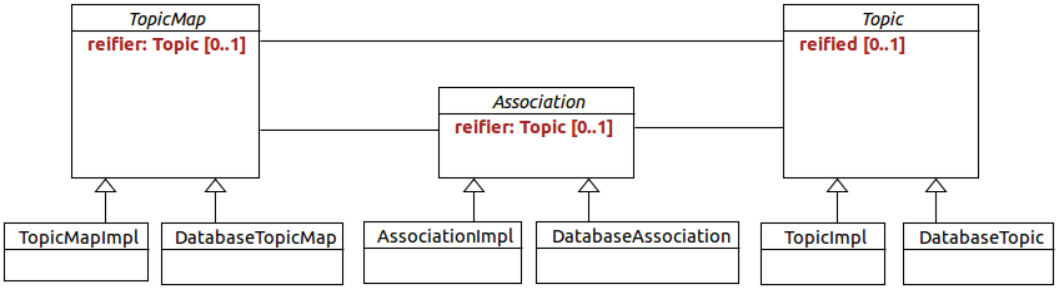


FIGURE 7.8 – Méta-modèle des Topic Maps dans Wandora

cette tâche plus aisément sachant que Wandora propose de nombreuses variantes de stockage des Topic Maps. L'initialisation de ces propriétés a nécessité, par ailleurs, la modification du module d'importation et, plus spécialement, de son analyseur syntaxique. La figure 7.9 schématise l'ajustement du processus de chargement d'une Topic Map sérialisée au format XTM 2.0. Une fois la Topic Map sérialisée et ses

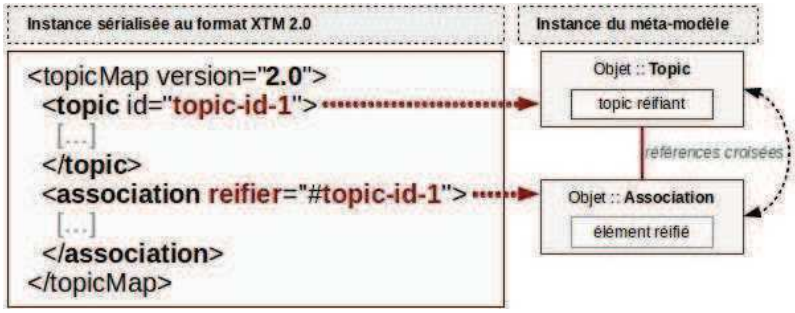


FIGURE 7.9 – Ajustement du processus d'importation de Wandora

réifications chargées, l'API du système graphique permet de parcourir le modèle de données afin de récupérer son contenu. Toutefois, Wandora ne dispose d'aucun levier pour identifier la structure contextuelle et discerner précisément les données à afficher. Cette problématique a requis donc un moyen pour distinguer les éléments contextuels tels que les contextes, les termes ou encore les associations qui les relient. Pour ce faire, nous avons exploité l'API du modèle de données de Wandora permettant d'obtenir des opérations pour tester ou récupérer des éléments en fonction de

leur type afin de prendre en charge la problématique d'identification de la structure contextuelle.

Le chargement automatique d'une Topic Map et l'affichage du graphe de la vue synthétique ont requis, quant à eux, le développement d'une nouvelle extension. La figure 7.10 présente le méta-modèle de ce module (LoadContextualTopicMap) et ses dépendances avec les autres classes du système. Le chargement automatisé s'appuie

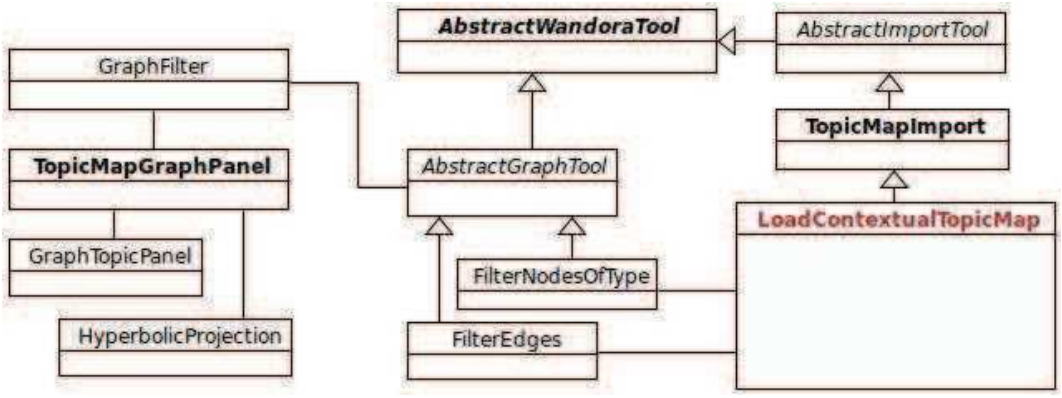


FIGURE 7.10 – Méta-modèle de l'outil d'import et d'affichage du graphe

sur le mécanisme d'héritage qui est applicable aux extensions. Ce dispositif, relevant du paradigme objet, permet ainsi de spécialiser le module original d'importation. Il est alors possible d'instancier immédiatement le modèle de données puis de configurer le rendu du graphe avant d'afficher la Topic Map. La personnalisation du graphe est facilitée par des classes d'origine dédiées au filtrage des nœuds et arêtes, selon leur classification. L'apposition des classes prédéfinies permet ici d'appliquer un filtrage particulier sur les topics de type « Term » ou « ContextualTerm ». Il en résulte un graphe élagué représentant la vue synthétique. Une fois l'outil adapté au chargement automatisé des Topic Maps et l'affichage du graphe de la vue synthétique, il a fallu, ensuite, réfléchir sur une solution permettant l'affichage de niveaux de détails supplémentaires. La solution proposée a été une mise en place de boites de dialogues (voir figure 7.11) que gérera la classe « TopicMapGraphPanel ».

Un tel choix permet de lister les contextes un à un sans surcharger l'interface.

Les contextes sont développés en indiquant la catégorie et la valeur des paramètres associés. Les occurrences sont, quant à elles, listées en précisant pour chacune le type de la ressource ainsi que son URI. La mise en page adoptée regroupe les termes

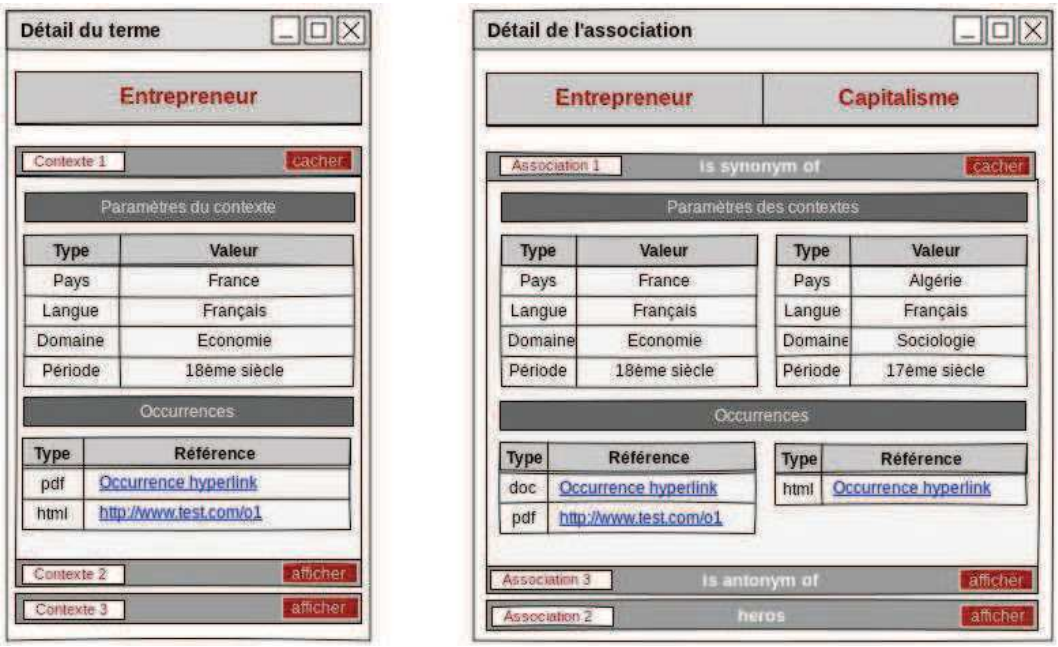


FIGURE 7.11 – Boites de dialogues détaillant termes et associations

contextualisés dans un système d'accordéon qui se déplie et se replie. L'affichage du détail d'une association entre termes réutilise la même mise en page en y ajoutant une colonne pour chaque membre de l'association. La boite de dialogue en question, illustrée sur la droite de la figure 7.11, reprend le système d'accordéon qui cette fois liste chacune des associations en précisant leurs natures, les contextes en relation et les occurrences qui s'y rapportent. L'obtention de ce niveau de détail complémentaire repose sur deux modules qui parcourent le modèle de données, à partir de l'association ou du terme sélectionné. Ils récupèrent progressivement les contextes disponibles et les occurrences concernées avant de les mettre en page. La figure 7.12 présente le diagramme de classe du module dédié aux termes (DetailTerm).

Le second module d'extension, consacré au détail des associations entre termes,

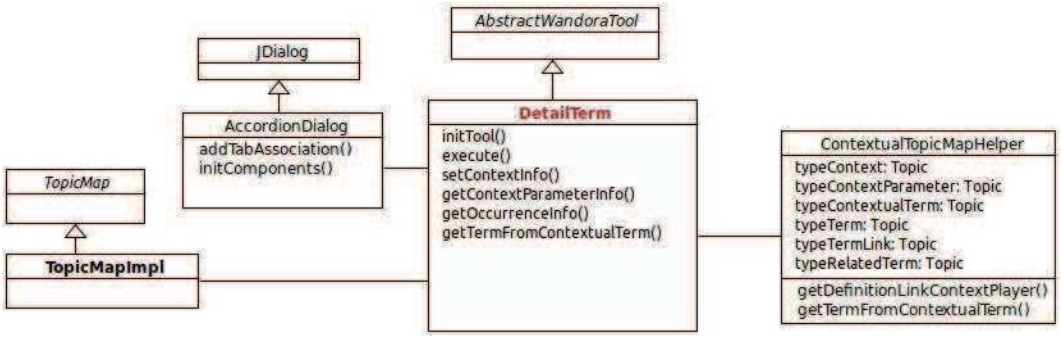


FIGURE 7.12 – Méta-modèle de l'outil d'affichage du détail d'un terme

possède exactement la même structure que celle présentée dans la figure 7.12. La seule différence réside dans son comportement qui prend en compte les différents membres des associations.

Deux autres adaptations mineures ont été effectuées sur Wandora. La première a concerné la visualisation de la taxonomie des paramètres de contexte qui n'est possible sous Wandora que si elle est intégrée dans celle du système en tant que sous-classe du type racine ($Wandora\_class$). La seconde concerne les noms attribués aux topics contextualisés. Contrairement à ce qui est préconisé dans la norme, Wandora exige l'attribution de noms différents aux topics. Cette contrainte bien qu'inexistante dans la norme, nous a obligé à suffixer les noms identiques avec un index unique.

### 7.3.2 L'interfaçage de l'outil wandora étendu avec le prototype

L'interfaçage de l'outil Wandora étendu et non son intégration, nous permet d'interfacer, si possible, dans l'avenir, un autre outil de visualisation plus adapté, sans changements majeurs sur le prototype. Sa mise en œuvre s'est concrétisée par la conception et la réalisation d'un module spécifique assurant le pont entre le prototype et Wandora. Il s'attache à lancer le sous-système et à lui fournir une Topic Map sérialisée, selon ses besoins et contraintes spécifiques. La figure 7.13 illustre le méta-modèle de ce composant.

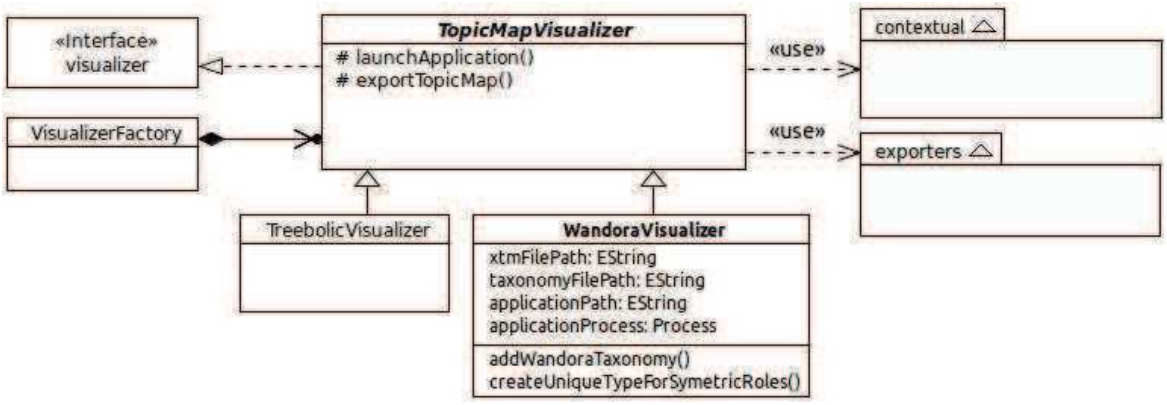


FIGURE 7.13 – Méta-modèle du module de lancement du sous-système graphique

Quoique Wandora ait été retenu, la modélisation exploite le patron de conception « Strategy » afin de simplifier l'intégration d'autres applications graphiques.

### 7.3.3 Le stockage, la création et la gestion des vues

Dans la réalisation du prototype, nous nous sommes fixés deux contraintes : l'utilisateur peut charger une vue locale ou distante et le système doit être capable d'exploiter plusieurs supports de stockage. Pour répondre à ces contraintes, nous avons, dans un premier temps, mis en place un mécanisme permettant de retrouver une vue d'une Topic Map. Ce mécanisme consiste en l'annotation de chaque vue du contexte formel associé à une Topic Map donnée, par une métadonnée précisant l'emplacement des fichiers sérialisés sur le support de stockage. Cette solution laisse une grande marge de manœuvre en terme de choix technologiques dans la mesure où elle peut représenter un chemin d'accès vers un fichier local, une URI vers un système de fichier distant ou encore un identifiant d'enregistrement en base de données. Dans un second temps, nous avons procédé au choix d'une architecture de stockage des vues. De toutes les propositions faites, celle qui a été retenu est celle correspondante à un stockage local sur le poste de l'utilisateur (voir figure 7.14). Le choix de cette solution est guidé par sa simplicité de mise en œuvre et non par une étude de

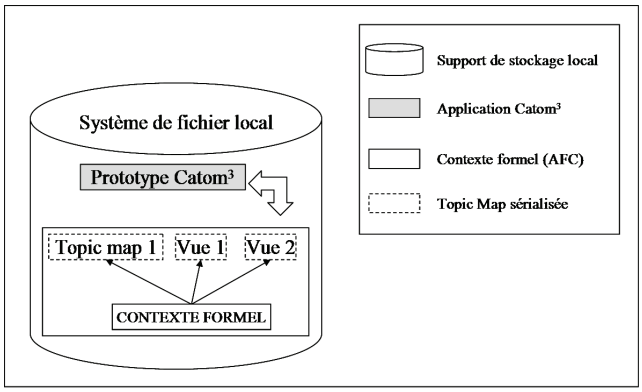


FIGURE 7.14 – Stockage local sur le poste utilisateur

performance qu'on a d'ailleurs pas réalisé compte tenu du temps imparti pour le prototypage. Telle que le montre la figure 7.14, l'ensemble des fichiers matérialisant les Topic Maps sérialisées et les contextes formels sont enregistrés sur le poste de l'utilisateur. Sur cette figure sont schématisées les relations entre l'application et ces fichiers ainsi que le rôle central du contexte formel. En effet, ce dernier assure le lien entre le prototype et les vues disponibles. Au chargement d'une Topic Map, le logiciel se charge de générer un contexte par défaut comprenant la Topic Map globale. Il s'appuie ensuite sur cette structure pour créer et retrouver les vues dans le système de fichiers.

Pour la création des vues, leur consultation et gestion, nous nous appuyons sur l'outil Galicia qui permet de calculer un contexte formel. La méthode d'analyse formelle ainsi que la création des vues sont gérés respectivement par les classes « Lattice » et « TermLinkViewBuilder » du méta-modèle proposé pour la gestion des vues (voir figure 7.15).

## 7.4 Description des modules de création et de fusion

Le module de création englobe deux sous-modules : annotation et génération de la Topic Map. Il est interfacé avec les modules d'exportation et visualisation. Son

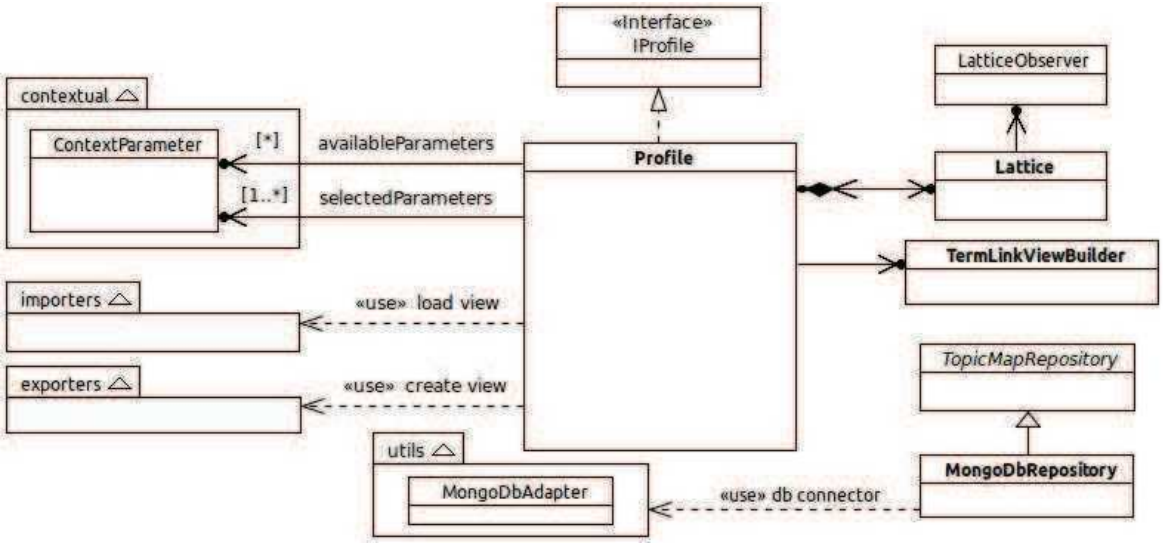


FIGURE 7.15 – Méta-modèle du module de gestion des centres d'intérêt

développement a concerné un contenu textuel. Les deux sous-modules le constituant correspondent respectivement à la mise en œuvre du composant méthode « annotation des ressources » et du composant méthode « génération type A d'une Topic Map contextualisée » décrits au chapitre 5 de cette thèse.

Le module annotation des ressources exploite le module Gazetter proposé par la plateforme GATE pour l'extraction de termes. Via une interface développée dans ce module, les termes sont contextualisés par l'utilisateur pour être traduit, par la suite, par le système en déclarations RDF à l'aide de l'outil Jena. Ces déclarations RDF instancient le méta-modèle d'annotation décrit au chapitre 5. La figure 7.16 présente l'interface d'édition des termes contextualisés. Elle permet de visualiser le document annoté et d'associer à chaque sélection de terme les éléments de contexte. Ces éléments une fois choisis, contextualisent le terme auquel on peut associer via des liens d'associations d'autres termes.

Le module « Génération de Topic Map » a nécessité dans un premier temps la mise œuvre d'une transformation M2M pour l'instanciation du méta-modèle contextuel. Cette transformation a été réalisée selon des règles de transformation exprimées

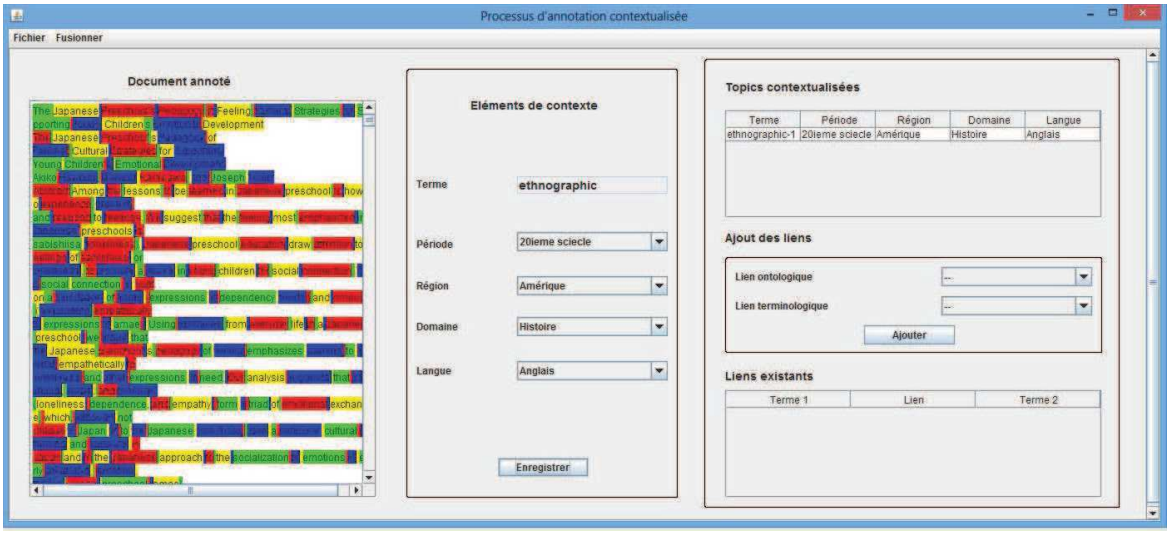


FIGURE 7.16 – Interface d'édition des termes contextualisés

en ATL. La figure 7.17 fournit un exemple de règle utilisée. Elle décrit la transforma-

**Règle *ContextParameter()***

Les paramètres de contexte demandent

l'initialisation de leur valeur et de leur catégorie.

```
[nom]
    ContextParameter
[extension]
    Règle topic
[condition d'application]
    source.transfoType = 'ContextParameter'
[transformation]
    source Annotation:ContextParameterSetDescription
    cible  Contextual:ContextParameter
    (
        category ← nom du premier paramètre de contexte
        value    ← nom du topic
    )
```

FIGURE 7.17 – Règle de transformation du module de création

tion de la classe « ContextParameterSetDescription » du méta-modèle d'annotation en la classe « contexteParameter » du méta-modèle contextuel de la Topic Map. En

l'absence de ressources terminologiques au moment de la réalisation du prototype, les deux étapes suivant cette transformation et permettant de construire le réseau sémantique de la Topic Map contextualisée ont été réduit à un simple module permettant à l'utilisateur de mettre en place des liens entre topics contextualisés (voir figure 7.18).

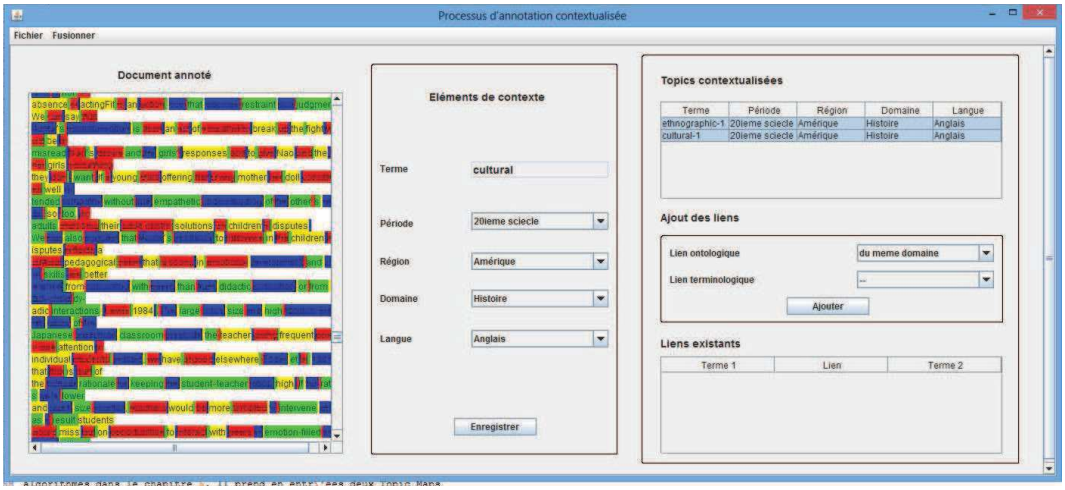


FIGURE 7.18 – Interface de rajout de lien entre termes contextualisés

Le module de fusion, intégre deux Topic Maps contextualisées importées à l'aide du module importation et renvoie le résulat de l'intégration au module exportation qui se charge de sa transformation en XTM. Ce module implémente l'algorithme de fusion du composant méthode « fusion » décrit au chapitre 5 de cette thèse. La figure 7.19, présente une vue synthétique d'une portion de Topic Map dédiée aux SHS. Seuls les topics non contextualisés sont affichés. L'accès aux niveaux de détails des topics contextualisés et des associations sont possible via le menu associé à chacun des topics.

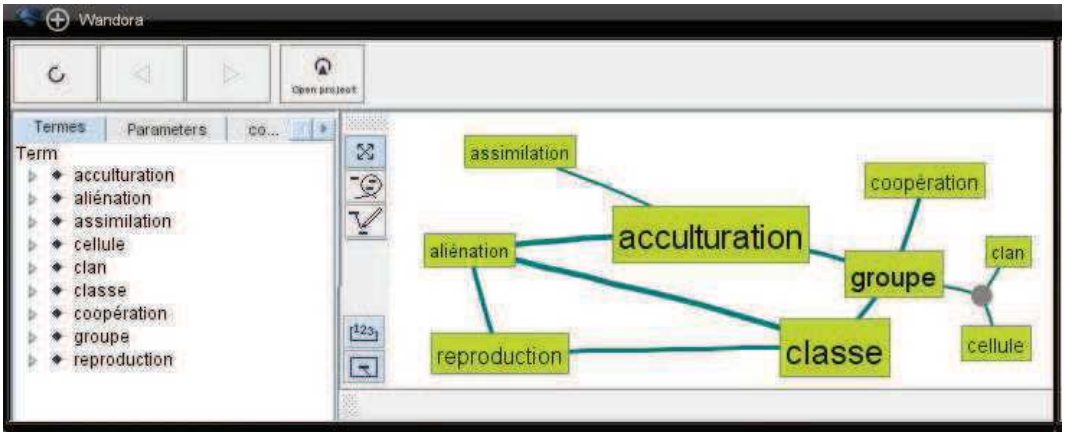


FIGURE 7.19 – Vusualisation de la Topic Map contextualisée des SHS

## 7.5 Conclusion

Ce chapitre a été consacré à la présentation de quelques aspects liés à la mise en œuvre de la plateforme de création, évolution et visualisation de Topic Maps contextualisées.

La première version de la plateforme vise à satisfaire quelques exigences fonctionnelles décrites dans le diagramme des cas de UML.

Cette plateforme a été développée selon les principes de l'IDM afin d'assurer son évolutivité et d'assurer l'évolutivité des approches qu'elle supporte. Lors du processus de développement, nous avons veiller à ne pas être dépendant des outils exploités ni des systèmes de stockage. Nous avons certes mis en œuvre uniquement une approche de création et une approche d'évolution. Cependant, de par l'architecture de la plateforme, certains modules développés peuvent être réutilisés pour la mise en œuvre d'autres approches.

# Chapitre 8

# Conclusion et perspectives

Pour permettre un meilleur accès aux contenus et une recherche précise des informations, les ressources informationnelles doivent être enrichies de métadonnées. Ces métadonnées correspondent la plupart du temps aux termes et concepts se trouvant dans les ressources. Elles sont actuellement spécifiées selon deux standards : le modèle RDF et le modèle de Topic Map. La spécificité du modèle de Topic Map réside dans le fait qu'il est un réseau de liens sémantiques entre les métadonnées. Cette structuration en réseau permet une navigation facile et sélective au sein des contenus. Du fait de l'avantage que présente ce modèle, nous avons proposé de l'utiliser dans l'organisation de contenus multilingues et pluridisciplinaires dont la spécificité est la variation du sens des termes et concepts utilisés dans ce type de contenus.

L'absence de modèle de Topic Maps adapté à des contenus multilingues et multidisciplinaires, nous a amené, après un état de l'art détaillé, à proposer un nouveau modèle de Topic Maps qui est avant tout une extension du modèle standard. Nous l'avons nommé « modèle de Topic Maps contextualisées ».

Les Topic Maps, au regard de leurs objectifs et utilisations, s'adressent à des contenus volumineux et en perpétuelle évolution. Les construire et les faire évoluer manuellement est une tâche quasiment impossible. Il est donc nécessaire de mettre en œuvre des approches satisfaisant ces objectifs. Bien que les Topic Maps soient utiles

en termes de recherche d'information, elles posent une problématique de visualisation du réseau sémantique qu'il ne faut pas négliger. Une proposition de modèle, sans mise en œuvre d'approches permettant son instanciation ainsi que l'enrichissement et la visualisation de son réseau sémantique, aurait été, de notre point de vue, une contribution incomplète.

Enfin, la construction de Topic Maps nécessite, la plupart du temps, l'utilisation de ressources sémantiques décrivant le domaine couvert par le contenu qu'elles organisent. Notre étude de l'état de l'art sur les modèles de ressources sémantiques décrivant un ou plusieurs domaines, nous a permis de constater l'absence de ressources intégrant la variation du sens des termes. Notre proposition de structure générique du Wiktionnaire, conçue pour les sciences sociales et humaines mais adaptable à d'autres domaines, est, à notre avis, une réponse à ce besoin.

Pour résumer, nos contributions de recherche ont été de plusieurs ordres. La première contribution a concerné la définition d'une structure de wiktionnaire (dictionnaire électronique développé selon la technologie wiki) permettant de recenser les termes de plusieurs domaines et de fournir une description contextualisée de ces termes. Ce wiktionnaire est conforme à la norme ISO 1951. Bien que dédié à l'heure actuelle aux sciences sociales et humaines, sa structure est suffisamment générique pour couvrir d'autres domaines.

La seconde contribution est une application permettant d'instancier, de mettre à jour et de consulter le wiktionnaire. Un déploiement de cette application sur un réseau pair à pair a aussi été mis en œuvre. Cette solution permet de pallier les problèmes liés à la montée en charge et aux pannes réseaux.

La troisième contribution est le modèle de Topic Maps contextualisé. Ce modèle est une extension du modèle standard. Sa structure permet de décrire des sujets à sens variés.

La quatrième contribution est une famille d'approches pour la construction et l'enrichissement d'une Topic Map contextualisée. Ces approches peuvent être construites

par assemblage de composants méthodes décrits dans cette thèse.

La cinquième contribution est la proposition de deux mécanismes de visualisation de Topic Maps contextualisées. Le premier mécanisme permet d'obtenir une vue synthétique d'une Topic Map ou d'une portion de Topic Maps. Le second est un mécanisme permettant une visualisation personnalisée d'une Topic Map. Ce second mécanisme est accompagné d'une proposition de matérialisation de ces vues ainsi que de leur gestion.

La dernière contribution est d'ordre pratique. Elle concerne la mise en œuvre d'une plateforme de construction et de gestion de Topic Maps contextualisées. Son développement a suivi les principes de l'IDM (Ingénierie Dirigée par les Modèles).

**Perspectives de recherche**

Notre travail peut être poursuivi dans au moins trois axes qu'on peut classer par ordre de priorité de la façon suivante :

**Axe 1** : amélioration du processus de fusion qui, pour l'instant, effectue un appariement syntaxique. Pour ce faire, nous comptons réutiliser et adapter les travaux existants dans le domaine des systèmes d'information.

**Axe 2** : amélioration du processus de création de Topic Maps contextualisées. Dans cet axe, il s'agira essentiellement d'améliorer le processus de génération de Topic Maps contextualisées qui, pour l'instant, faute de disponibilité de ressources terminologiques adaptées à notre contexte, s'appuie essentiellement sur l'expert du domaine. Il est évident que le peuplement du wiktionnaire contribuera à l'automatisation de ce processus dés qu'il s'agira de construire une Topic Map propre au domaine des SHS. Une autre amélioration concerne la phase de dérivation des liens sémantiques qui est, pour l'instant, totalement manuelle. Nous pensons que l'utilisation d'ontologies de domaines permettrait d'améliorer le degré d'automatisation du processus.

**Axe 3** : amélioration de la qualité de la Topic Map ainsi que celle des vues personnalisées, en introduisant des métadonnées qui fournissent des renseignements

quant à l'usage de la Topic Map et des vues. Cela permettra de décider de la pertinence des vues et des constituants de la Topic Map. Une fois ces métadonnées spécifiées, un système décisionnel peut être envisagé pour décider de l'élagage des vues et des constituants de la Topic Map. Pour cette problématique, nous comptons profiter des travaux réalisés au sein de notre équipe.

Deux perspectives, d'ordre pratique cette fois, peuvent être menées en parallèle avec les perspectives de recherche citées plus haut. La première perspective concerne l'intégration dans la plateforme des approches citées dans le chapitre 5 de cette thèse et qui n'ont pas fait l'objet d'une mise en œuvre. La deuxième perspective est relative à l'évaluation réelle de la montée en charge du nombre d'utilisateurs du wiktionnaire.

Chapitre 9

# Liste des publications de cette thèse

| | |
|---|---|
| [1] | Lydia Khelifa, Nadira Lammari, Jacky Akoka, Thouraya Bouabana-Tebibel. Building Contextualized Topic Maps. 19th IBIMA (International Business Information Management Association) Conference on Innovation Vision 2020. Barcelone, Espagne. 12-13 Novembre 2012. |
| [2] | Lydia Khelifa, Nadira Lammari, Hammou Fadili, Jacky Akoka. A Wiki-Oriented Online Dictionary for Human and Social Sciences. Fifth Workshop on Semantic Wikis Linking Data and People (SemWiki2010) co-located with ESWC 2010 (European Semantic Web Conference). Hersonissos, Crête, Grèce. 31 Mai 2010. |
| [3] | Lydia Khelifa, Rafik Mezil, Aziz Si-Mohammed, Thouraya Bouabana-Tebibel. A Human and Social Sciences Wiktionary in a Peer-to-Peer Network. In IEEE International Conference on Machine and Web Intelligence (ICMWI 2010). Alger, Algerie. 3-5 octobre 2010., |
| [4] | Lydia Khelifa, Nadira Lammari, Hammou Fadili, Jacky Akoka. Un Wiktionnaire Multilingue et Multiculturel pour les Sciences Sociales et Humaines. Le Web Social 2010 En conjonction avec 10ème Conférence Internationale Francophone sur l'Extraction et la Gestion des Connaissances (EGC 2010). Hammamet, Tunisie, 26 Janvier 2010 |
| [5] | Hammou Fadili, Nadira Lammari, Lydia Khelifa, Jacky Akoka. Un Wiktionnaire pour les Sciences Sociales et Humaines. Le dictionnaire électronique. Quelles perspectives pour les sciences humaines et sociales. 2007. |

# Annexe A

# Outils de visualisation

Quantité d'outils, dédiés ou non aux Topic Maps, sont à même d'afficher des structures modélisant la connaissance. En revanche, peu d'entre eux se révèlent efficace en termes de visibilité et d'ergonomie. De surcroît, beaucoup de prototypes prometteurs sur le papier ne sont accessibles qu'à travers la présentation qui en est faite.

Cette annexe se rapporte donc uniquement aux solutions qui ont pu être testées concrètement. La première section de cette annexe répertorie les outils étudiés. La seconde section présente une étude comparative permettant de sélectionner un sous ensemble d'outils répondant à notre problématique de visualisation. La dernière section approfondie la précédente étude afin de sélectionner l'outil à utiliser pour notre développement de plateforme de gestion de Topic Maps contextualisées.

## A.1 Panorama des outils de visualisation

Ci-après une brève présentation des outils étudiés.

**H3 Viewer**

Cet outil proposé par [Munzner, 1998] utilise un graphe hyperbolique en 3D confiné au sein d'une sphère dans laquelle la navigation se fait par rotation et translation.

L'interactivité est limitée. Il n'est ainsi pas possible d'effectuer des zooms ni de masquer des branches du graphe. Le libellé des nœuds rend, par ailleurs, la vue confuse dès qu'il y a une concentration d'informations. Cet outil trouve principalement son utilité dans l'étude de la vue d'ensemble d'une large structure.

**Ontopia Vizdesktop**

L'outil [ontopia, 2001] propose une vue détaillée sous forme de graphe qui s'avère vite encombrée. Il exploite un système de navigation qui ne comprend que des translations ainsi qu'un unique algorithme de mise en page automatique qui empêche tout ajustement manuel. De plus, les interactions avec le graphe entraînent un bouleversement du rendu et perturbent fortement la représentation cognitive. Vizdesktop se destine donc essentiellement à l'affichage de petites parties d'un graphe dans un contexte d'édition.

**Cytoscape**

L'application Cytoscape ([Smoot et al., 2011] et [Smoot et al., 2013]) est orientée bio-informatique. Elle permet l'analyse et la visualisation de réseaux variés tels que ceux des molécules ou des ontologies. Le logiciel utilise un graphe vectoriel et plusieurs algorithmes de mise en page automatique. La navigation est efficace et repose sur une vue de haut globale liée à des translations et des zooms. L'application met l'accent sur la personnalisation graphique des nœuds, arcs et libellés afin de faire ressortir des propriétés du réseau. Cependant, l'interactivité avec le graphe reste difficile à l'image du masquage de certaines branches.

**TMNav**

TMNav [Dicheva and Dichev, 2006] est un outil dédié aux Topic Maps. Il propose trois techniques : un graphe hyperbolique, un graphe standard et une liste indentée de la classification. Le graphe hyperbolique est de bonne qualité et peut être en partie personnalisé. TMNav utilise un système de filtrage avancé par requêtes et

propose un historique des topics visités. Bien que relativement efficace, cet outil, dont la dernière évolution date de 2004, ne prend pas en compte le modèle standard (TMDM) publiée en 2006.

**Protege Ontograph**

Ontograph [Falconer, 2010] est un module d'extension de l'éditeur Protege qui permet de visualiser une ontologie sous forme de graphe. Il fournit uniquement une vue détaillée associée à une liste indentée de la classification. La mise en page automatique est adaptée à des structures hiérarchiques et nécessite souvent des ajustements manuels. L'interaction avec les nœuds provoque, par ailleurs, un déplacement de tous les éléments ; ce qui demeure très perturbant. Ontograph s'avère pratique pour visualiser une partie de l'ontologie et trouve donc son utilité dans un cadre d'édition.

**Wilmascope**

Wilmascope [Dwyer and Eckersley, 2004] produit des graphes en 3D où l'on peut naviguer au moyen de rotations, de translations et de zooms. Les nœuds, libellés et arcs sont personnalisables graphiquement. L'interactivité est relativement bonne, mais bien que le rendu soit relativement bon, un bon rendu la vue en 3d reste déstabilisante en raison des multiples points de vue dûes aux rotations et aux zooms.

**OWL2Prefuse**

OWL2Prefuse [Jethro and Giles, 2007] matérialise des ontologies Owl sous forme de graphe ou d'arbre hiérarchique. Il utilise un système de navigation par translation et zoom efficace associé à un algorithme de mise en page qui occupe bien l'espace. La sémiotique est bonne et met en évidence les caractéristiques de l'ontologie ainsi que les relations de proximité entre les nœuds. Toutefois, l'animation continuelle des nœuds et le fait de ne pas pouvoir les masquer pénalisent le rendu d'ensemble.

**Treebolic**

Treebolic [Bou, 1996] génère un arbre hyperbolique de très grande qualité possédant une sémiotique et des méta-données personnalisables. La navigation est bonne et permet de se déplacer rapidement dans une structure conséquente. Fonctionnalité singulière de cet outil : il est capable de charger progressivement les branches avec des fragments Xml et autorise ainsi un chargement progressif d'arbres de grande taille. On peut regretter que cet outil demande de convertir un graphe en arbre.

**Wandora**

Wandora [Team, 2014] est un éditeur de Topic Maps proposant un graphe hyperbolique et une vue arborescente de la classification qui sont de bonne facture. La navigation et les interactions sont fluides et de nombreux filtres sont disponibles pour personnaliser le graphe. Toutefois, il n'implémente pas toutes les spécificités du standard ISO 13250 et en particulier la réification. De plus, bien que la sémiotique et les méta-données soient appréciables, les couleurs utilisées pour distinguer les éléments changent à chaque chargement ce qui peut être quelque peu déroutant.

## A.2 Comparaison et synthèse des outils de visualisation

La majorité des études sur les techniques et outils existants concluent que la technique ou l'application « universelle » n'existe pas [Katifori et al., 2007], [Boinski et al., 2010] et [Dmitrieva and Verbeek, 2009].

Un bon système de visualisation doit plutôt viser l'intégration de plusieurs techniques, afin d'offrir de multiples points de vue aux utilisateurs. [Shneiderman, 1996] a établit une liste, toujours d'actualité, de sept fonctionnalités à considérer pour concevoir un logiciel de visualisation efficient et adapté aux ontologies et que nous prenons en considération dans les Topic Maps :

- Affichage d'une vue d'ensemble pour avoir une idée de la structure globale ;
- Affichage de la vue détaillée d'un élément en conservant les nœuds environ-

nants ;

- Filtrage de l'ontologie selon sa classification ;
- Affichage de différents niveaux de détails par regroupement de concepts ;
- Visualisation des relations entre éléments ;
- Historique des actions de navigation pour annuler, rejouer ou affiner les cheminements dans l'ontologie ;
- Exportation de l'ontologie visualisée et des paramètres ayant permis le filtrage.

Afin de choisir l'outil de visualisation adéquat pour notre Topic Maps contextualisée, nous enrichissons les critères de [Shneiderman, 1996] avec celles dédiées à la visualisation de Topic Maps contextualisées. Ces critères sont :

- **Navigation** : qui fait référence à la qualité du système de navigation et des possibilités de déplacement au sein d'un graphe sémantique. Elle prend notamment en compte la possibilité d'effectuer des zooms, des glissements avec la souris ou encore des changements de perspective pour naviguer via les différents concepts.
- **Interaction** : qui prend en considération les actions d'affichage ou de masquage de nœuds, de branches ou de menus contextuels relatifs aux nœuds et arêtes.
- **Visibilité des nœuds** : qui évalue la qualité visuelle des éléments et de l'ensemble de la Topic Map, lors de l'affichage d'un graphe comprenant de nombreux nœuds.
- **Facilité d'adaptation** : qui caractérise le degré d'adaptabilité des logiciels à prendre en charge les spécificités de notre modèle de Topic Maps contextualisées tel que la gestion de la contextualisation des termes.
- **Dédié aux Topic Maps** : qui identifie les outils dédiés aux Topic Maps même si la plupart de ces outils ne le sont pas. Il reste appréciable d'avoir à disposition des fonctionnalités dédiées à cette technologie.

De ce fait, nous présentons dans le Tableau A.1 une comparaison entre les outils de

visualisation selon les critères présentés ci-dessus en mettant en avant les principales fonctionnalités, les techniques utilisées ainsi que la qualité de la navigation et des interactions.

| | Techniques | Vue d'ensemble | Vue détaillée | Navigation | Interaction | Visibilité des nœuds | Facilité d'adaptation | Filtres | Historique | Dédié aux Topic Maps |
|---|---|---|---|---|---|---|---|---|---|---|
| **Wandora** | Graphe hyperbolique Liste indentée | Oui | Oui | + | + | ~ | + | + | Oui | Oui |
| **Treebolic** | Arbre hyperbolique | Oui | Oui | + | ~ | + | + | x | Non | Non |
| **Cytoscope** | Graphe | Oui | Oui | + | - | + | ~ | ~ | Non | Non |
| **Vizdesktop** | Graphe | Non | Oui | - | - | - | ~ | + | Non | Oui |
| **TMNav** | Graphe hyperbolique Liste indentée | Oui | Oui | ~ | ~ | ~ | + | + | Oui | Oui |
| **Ontograph** | Graphe Liste indentée | Non | Oui | ~ | ~ | ~ | ~ | ~ | Non | Oui |
| **Wilmascope** | Graphe 3D | Oui | Oui | ~ | ~ | - | + | x | Non | Non |
| **Owl2Prefuse** | Graphe arbre | Oui | Oui | ~ | ~ | - | + | - | Non | Oui |
| **H3 viewer** | Graphe 3D hyperbolique | Oui | Oui | - | - | ~ | ~ | ~ | Non | Non |

Evaluation: + bon ~ moyen - mauvais x inexistant

FIGURE A.1 – Tableau comparatif des outils de visualisation : Wandora et Treebolic

**Synthèse des outils de visualisation**

De cette comparaison, nous constatons que la visualisation en 3D n'apporte que peu de bénéfices, voire des inconvénients comme indiqué dans [Katifori et al., 2007]. L'espace n'est pas optimisé comme attendu avec une dimension supplémentaire. L'interactivité et la navigation se révèlent, dans le même temps, plus compliquées. Ces observations se recoupent avec les tests réalisés et amènent à écarter d'emblée les solutions reposant sur la 3D telles que H3Viewer et Wilmascope. Notons toutefois, que des prototypes en 2.5D à mi-chemin entre la 2D et la 3D, telles que [da Silva and Freitas, 2011], semblent porter leur fruit.

Seules les solutions Tmnav, Cytoscape, Treeblic et Wandora se sont avérées convaincantes. Elles satisfont à cet égard la majorité des critères en termes de navigabilité, d'ergonomie et de visibilité tout en considérant leur facilité d'adaptation. Pour conclure, le comparatif différentiel de ces quatre outils laisse entrevoir une adéquation bien plus prononcée de Treebolic et Wandora pour visualiser convenablement des Topic Maps contextualisées.

**Etude comparative de Wandora et Treebolic**

Du tableau présenté ci-dessus, il en ressort que peu d'approches offrent des caractéristiques satisfaisantes pour visualiser de façon appropriée un graphe sémantique. Toutefois, les solutions Treebolic et Wandora sortent du lot et concordent de façon satisfaisante avec la majorité des critères proposés. Ces outils proposent en particulier une vue hyperbolique de très bonne facture où la navigation et l'interactivité y sont intuitives et efficaces. Pour permettre leur étude plus approfondie et arrêter un choix, le tableau a été établit en répertoriant les principaux avantages et inconvénients de ces deux applications (voir figure B.1).

| | Avantages | Inconvénients |
|---|---|---|
| *Treebolic* | Arbre hyperbolique et méta-données de qualité | Adaptationdu graphe en arbre |
| | Chargement dynamique des branches de l'arbre | Adaptationpour matérialiser la réification des terme |
| | Configuration du graphe avec le format natif | Pas de vue en liste indentée pour la classification |
| | Déploiement sur serveur avec son applet Java | Pas de méta-données propres au*Topic Maps* |
| *Wandora* | Graphe hyperbolique et méta-données de qualités | Adaptation nécessaire pour gérer le concept de réification |
| | Vue de la classification avec plusieurslistesindentées | Demande un moyen de détecter la structure contextuelle |
| | Vue de la classification sous forme de graphe | Le déploiement en ligne demande de développer un |
| | Facilité d'adaptation (système d'extensionet Api) | |
| | Accès direct à la structurecontextuelle | |
| | Fonctionnalités dédiéesau*Topic Maps* | |

FIGURE A.2 – Avantages et inconvénients de Treebolic et Wandora

Treebolic utilise un arbre (et non un graphe) hyperbolique qui requiert une adaptation pour visualiser une Topic Map. Il possède un format natif particulier et exige de pré-formater toutes les données contextuelles pour afficher le détail de la vue synthétique. Il faut mentionner, de plus, qu'il ne permet pas de visualiser la classification.

Wandora est un outil dédié aux Topic Maps qui peut charger une sérialisation au format XTM et utiliser directement les données de la structure contextuelle. La Topic Map est visualisée au moyen d'un graphe hyperbolique qui ne réclame pas de transformation. Cependant, il nécessite la configuration du filtrage des éléments à afficher. Son modèle de données ne prend cependant pas en compte toutes les spécificités du TMDM et, en particulier, la réification. Son architecture modulaire et sa simplicité d'adaptation permettent, malgré tout, d'envisager le dépassement de ces limitations.

Pour trancher, Treebolic utilise les arbres, le chargement dynamique des branches. Toutefois, Wandora autorise une réelle exploration de la structure contextuelle et son

architecture orientée modules extensibles permet de le personnaliser presque entièrement. De plus, Wandora étant dédié aux Topic Maps, il possède des fonctionnalités très utiles à l'image de l'affichage de multiples vues de la classification. Wandora est, par conséquent, la solution retenue pour assurer la représentation graphique des Topic Maps contextualisées.

# Annexe B

# Ingénierie Dirigée par les Modèles et cadriciel Eclipse

Cette annexe présente de façon très succincte, dans sa première section, l'IDM (Ingénierie Dirigée par les Modèles) et, dans sa seconde section, le cadriciel Eclipse ( Eclipse modeling framework).

## B.1 Bref aperçu sur l'ingénierie dirigée par les modèles

L'IDM applique les concepts de l'architecture dirigée par les modèles MDA (Model driven Architecture) [OMG, 2014a] dont le but principal est de pérenniser les savoir-faire métiers et d'augmenter la productivité en exploitant plus concrètement les modèles. La figure B.1 illustre son architecture qui repose sur des méta-modèles et des modèles exploitant les technologies de l'OMG. Dans cette architecture MOF (Meta Object Facility) [OMG, 2014b] est une spécification définissant les concepts de base pour créer et manipuler des méta-modèles. C'est un méta-méta-modèle permettant de décrire des langages de modélisation tels que Uml.

MDA se concentre sur la séparation des préoccupations ce qui consiste à isoler la partie métier de toute problématique technologique. Elle est concrétisée par

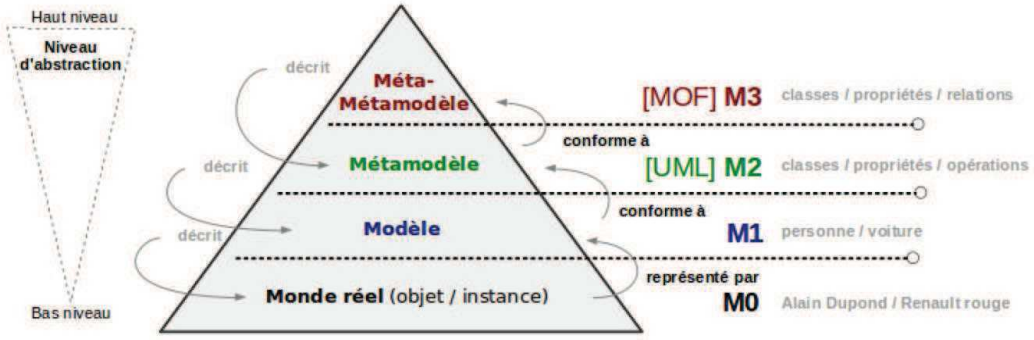


FIGURE B.1 – Architecture MDA sur quatre niveaux

le découplage des fonctionnalités considérées comme plus pérennes dans le temps de leurs implémentations techniques. Il en résulte des modèles indépendants technologiquement. Les déclinaisons sur diverses plate-formes sont ensuite automatisées à l'aide de générateurs de code qui assurent la transformation des modèles vers la technologie visée. Ces processus s'articulent sur des modèles utilisant trois niveaux d'abstraction dominants : CIM (Computation independant model), PIM (Plateform independant model) et PSM (Plateform specific model). Le CIM est le niveau d'abstraction le plus élevé qui précise les modèles d'analyse relatifs aux besoins métiers. Les exigences du domaine y sont modélisées dans leurs grandes lignes. Les modèles PIM sont indépendants technologiquement et s'attachent à la structuration du système. Ils permettent de concevoir les fonctionnalités nécessaires qui ont été identifiées dans les modèles CIM. Les modèles PSM adjoignent des considérations techniques aux PIM et possèdent donc le niveau d'abstraction le moins élevé. Les générateurs de code peuvent, dans un dernier temps, s'appuyer sur ces modèles pour cibler une plate-forme spécifique et automatiser une partie de la production du logiciel. Les transitions entre les différents niveaux d'abstraction impliquent des techniques de transformation de modèles. Ces procédés font partie des aspects essentiels de MDA et participent également à l'automatisation de la production.

L'IDM applique les concepts du MDA à des espaces technologiques variés et ne

considère plus la séparation des préoccupations comme un aspect primordial. La figure B.2 illustre cette généralisation à divers espaces afin de répondre à davantage de problématiques. Le passage d'un contexte à un autre est assuré par des transfor-

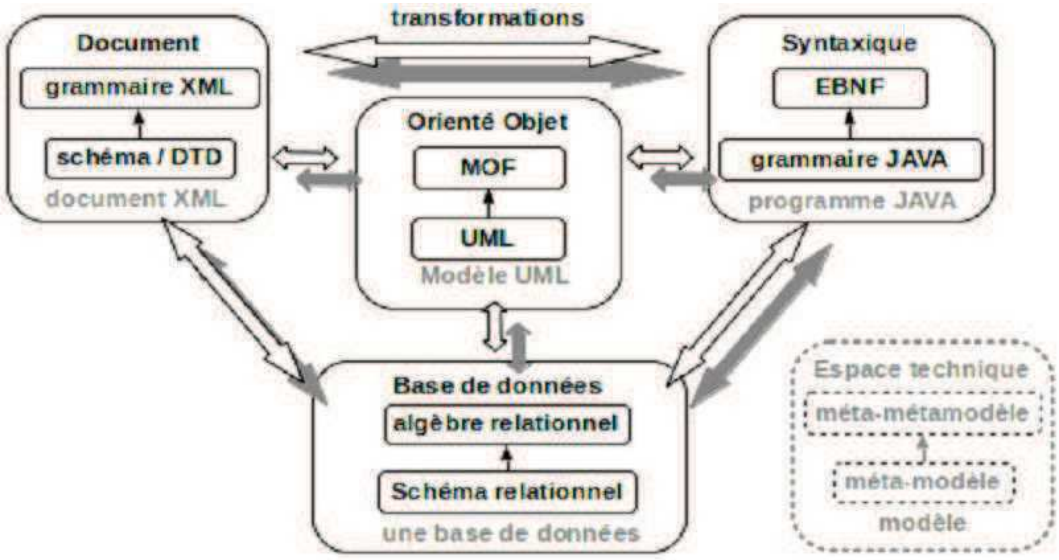


FIGURE B.2 – Exemple de relations entre divers espaces technologiques

mations de modèles qui, comme dans l'approche MDA, jouent un rôle essentiel. La section suivante présentes ces transformations.

### B.1.1 Les transformations de modèles dans l'IDM

L'IDM étend les concepts MDA à de multiples espaces technologiques et fait appel à des transformations de modèles pour passer de l'un à l'autre. Ces techniques sont capitales pour mettre à profit des formes variées de méta-modèles au sein d'un même processus de développement. A l'image des modèles traités, une transformation se conforme, elle aussi, à un méta-modèle spécifique comme le montre la figure B.3. C'est une opération qui prend un ou plusieurs modèles sources en entrée et produit un ou plusieurs modèles cibles en sortie. Elle peut être endogène si les modèles transformés se conforment au même méta-modèle, ou exogènes, si les méta-modèles sont différents. Une transformation peut, dans le même temps, être dite horizontale

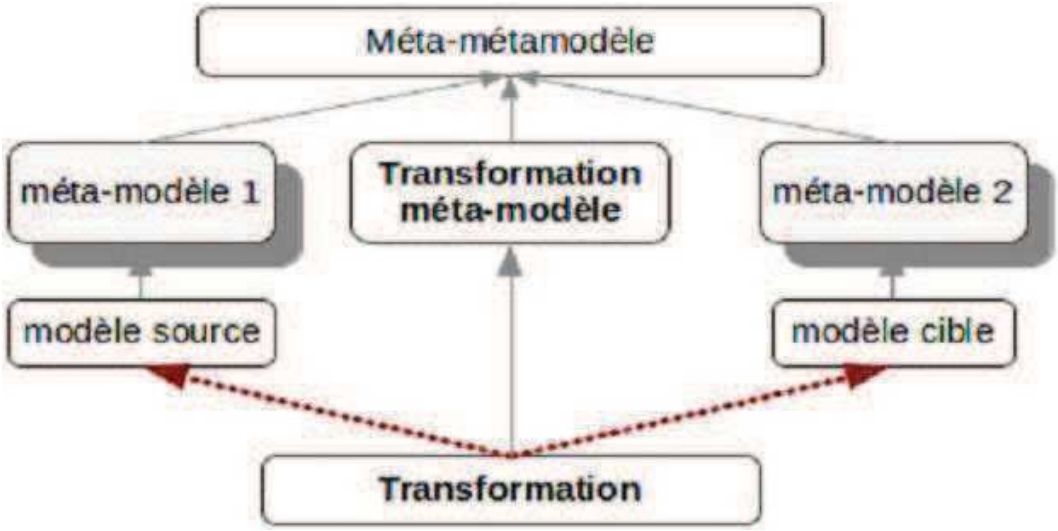


FIGURE B.3 – Méta-modèle d'une transformation

ou verticale. Elle est horizontale, si les modèles sources et cibles sont de même niveau d'abstraction. Réciproquement, elle est dite verticale, si les niveaux sont différents. Ces quatre caractéristiques se combinent et adressent divers besoins comme l'illustre le tableau B.1.

| | Horizontale | Verticale |
|---|---|---|
| Endogène | restructuration, optimisation, normalisation intégration de patrons de conception | raffinement |
| Exogène | migration de logiciels fusion de modèles | changement du niveau d'abstraction génération de code et de texte rétro-conception |

TABLE B.1 – Contextes d'utilisation des transformations

Une transformation possède, par ailleurs, des propriétés telles que sa traçabilité, sa réutilisabilité ou sa réversibilité. La traçabilité fait allusion aux mécanismes de sauvegarde des modifications effectuées. En conservant les « traces » des éléments transformés et leurs sources, des tests et autres validations des transformations sont par exemple envisageables. La réversibilité est aujourd'hui encore rarement prise en compte. Cette propriété liée à l'aspect bidirectionnel d'une transformation soulève, en particulier, des problèmes de perte d'information [Stevens, 2007]. Le recours à deux transformations distinctes permet, néanmoins, de contourner le problème.

Notons que la méthodologie IDM encourage, dans la mesure du possible, l'utilisation de multiples modèles intermédiaires pour passer progressivement d'une représentation à une autre. Cette bonne pratique cherche à limiter la complexité des transformations et simplifier la maintenance des chaînes de transformation.

Plusieurs techniques de transformation cohabitent dans une approche IDM. Le choix de la meilleure solution se fait en fonction des espaces technologiques concernés et des problématiques rencontrées. Elles sont classées en les trois grandes catégories suivantes [Vara Mesa, 2009] :

**Modèle vers Texte (M2T)**

Cette catégorie concerne les procédés de conversion d'un modèle dans une forme textuelle. Les deux techniques, mises en avant ci-après, permettent communément d'échanger des instances de méta-modèle ou de générer du code. Cette transformation peut être Manuelle ou Basée sur des gabarits textuels. La transformation Manuelle est programmée directement dans un langage généraliste tel que Java. Ces langages n'étant pas dédiés à ce genre de tâches, certaines transformations peuvent vite s'avérer complexes à réaliser et difficilement maintenables. La transformation Basée sur des gabarits textuels propose d'intégrer des expressions dans des gabarits qui fournissent la structure du texte à générer. Cette approche, très répandue, s'associe généralement à des technologies comme OCL pour parcourir les modèles. Les procédés de génération automatique de code sont fréquemment basés sur cette

technique.

**Modèle vers Modèle (M2M)**

Ce groupe se rapporte aux transformations d'un modèle vers un autre modèle. Il existe de nombreuses techniques M2M. Parmi lesquelles on peut citer : la transformation par manipulation directe, la transformation déclarative ou relationnelle, la transformation impérative.

La première transformation est une transformation effectuée programmatiquement. Elle exige que le code correspondant à la structure des méta-modèles soit disponible. Le modèle de données est alors utilisé directement pour rechercher les éléments sources et instancier les objets du modèle cible.

La transformation déclarative définit des relations entre les modèles sources et cibles. Elle établit des règles de correspondance entre les classes et attributs de chaque modèle. Si une relation est satisfaite, alors l'outil applique automatiquement les modifications. Cette approche est recommandée, et souvent privilégiée, car faisant partie des techniques les plus naturelles pour réaliser une transformation. Par ailleurs, la lisibilité du code et sa clarté facilitent le développement et la maintenance. Cette technique trouve, cependant, ses limites lors du traitement de transformations complexes.

La transformation impérative se base sur des mécanismes proches d'une programmation de type impérative, mais dédiée aux transformations de modèles. Elle propose des séquences d'instructions pour décrire les actions à réaliser. Ce genre d'approche est communément recommandé pour traiter des transformations complexes.

La transformation hybride combine les deux précédentes pour profiter de leurs avantages respectifs. Le choix d'une des options se fait alors selon les besoins et la complexité de la transformation. Comme précisé plus haut, le mode déclaratif reste à privilégier dans la mesure du possible.

**Texte vers Modèle (T2M)**

L'approche T2M regroupe les techniques permettant d'interpréter un contenu textuel afin d'instancier le modèle correspondant. Le texte analysé doit être structuré par une grammaire faisant office de méta-modèle.

Les outils T2M s'attachent principalement à la création d'analyseurs syntaxiques qui autorisent la transformation des contenus textuels en modèle. Leur domaine d'application est large et comprend, entre autres, l'interprétation des langages de programmation ou les techniques de rétro-ingénierie permettant d'extraire les méta-modèles de programmes existants.

## B.2 Bref aperçu sur le cadriciel ECLIPSE

Eclipse modeling framework (Emf) fournit un ensemble de composants dédiés à la création et la manipulation de modèles, ainsi que des générateurs de code basés sur ceux-ci. Il offre un environnement de développement basé sur la plate-forme Eclipse et des facilités pour construire des outils s'appuyant sur des méta-modèles.

Les méta-modèles utilisés se basent sur un méta-méta-modèle, nommé Ecore, implémentant la composante Emof de la spécification Mof 2.0. Contrairement à Uml, Ecore se limite à la description des aspects structurels tels que ceux illustrés dans un diagramme de classe. Sa structure se résume principalement aux concepts centraux des « Eclass », « EdataType », « EAttribute » et Ereference pour décrire respectivement les classes, les types de données, les propriétés et les associations entre classes.

De par sa large diffusion, Emf constitue le socle technologique de nombreux outils dédiés à l'IDM. Son méta-méta-modèle Ecore apparaît ainsi comme un standard de facto qui facilite la mise en œuvre d'une telle approche. On peut, par ailleurs, remarquer que les concepteurs de Emf ont pris le parti d'une approche médiane, entre le tout modélisation et le tout programmation.

La création des méta-modèles peut être effectuée de façon manuelle ou à l'aide de diagramme Ecore. Cependant, l'outil de création graphique est décevant et ne

propose que peu de fonctionnalités. Ecore étant une extension de Emof, il existe tout de même une certaine interopérabilité entre les outils Emf et UML. Des modèles UML peuvent, par conséquent, être convertis vers le format Ecore.

C'est dans ce cadre que l'éditeur graphique de modèles Uml, Papyrus, prend toute son utilité. Il offre une implémentation solide de la spécification de l'Omg et permet de créer, entre autres, des diagrammes de classes, diagrammes de cas d'utilisation ou diagrammes de profils de très bonne qualité. Il gère en outre efficacement les interdépendances entre de multiples méta-modèles.

De par son support du standard UML et sa richesse fonctionnelle, Papyrus s'est donc avéré être un bon candidat pour produire les méta-modèles du prototype. Notons, du reste, que tous les diagrammes de classe présentés dans ce document proviennent de cette application.

**Génération automatisée du code**

La génération du code correspondant à l'ossature des méta-modèles est prise en charge par des générateurs Emf. Ceux-ci produisent exclusivement du code Java qui traduit les concepts de classes, d'attributs, d'opérations et de références des modèles Ecore.

Le code généré est de bonne qualité et exploite, notamment, un patron de conception réalisant la séparation des interfaces et de leurs implémentations. D'autres patrons de conception y sont promus tels que la fabrique « Factory » qui facilite l'instanciation des classes du modèle ou le patron d'observation « Observer » qui gère la notification d'événements sur les objets.

La génération automatique du contenu des opérations est une chose bien plus complexe et difficilement automatisable, à moins d'utiliser des Dsml. Emf ne peut générer ces aspects dynamiques et comportementaux des modèles. Une fois le squelette produit, l'implémentation des opérations doit donc être réalisée programmatiquement.

Notons toutefois que des initiatives de génération des opérations existent. Ces

techniques tendent vers le tout modélisation, mais restent encore du domaine de la recherche ou du prototypage. Notre approche, qui se veut médiane, se contente, par conséquent, de la génération de la structure des méta-modèles.

Pour assurer la synchronisation entre les méta-modèles et le code modifié des opérations, Emf utilise diverses annotations. Celles-ci indiquent aux générateurs l'emplacement du code personnalisé et leur permettent de ne mettre à jour que les zones qui doivent l'être. Ce mécanisme évite, de cette façon, de perdre les modifications effectuées et maintient une cohérence entre les méta-modèles et le code auto-généré.

# Annexe C

# Algorithmes référencés

Dans cette annexe nous présentons les deux algorithmes qui ont été cités dans les chapitres 5 et 6 de cette thèse. Le premier algorithme concerne la dérivation de la structure contextuelle d'une Topic Map à partir de sa descriptionsous le format standard.

## C.1 Algorithme pour l'identification des objets de la structure contextuelle

```
Algorithme IdentifierStructureContextuelle
{détermine le type contextuel des topics et associations, puis initialise la propriété transfoType des objets du modèle pivot}
Entrées / modelPivot : TopicMap, enumTerminologicType : Tableau de chaînes
Sorties  / void

classe Liste
        attributs entête: ListElement {référence vers premier élément} , curseur: ListElement {réf. vers élément en cours}
        fonctions suivant(), premier(), dernier(), nombreElement(), horsListe() {sous algorithmes non détaillés}

type ListElement  = agrégat valeur: Info {type de l'élement} , suivant: ListElement {référence le suivant} fin

{les structures de données reprennent le méta-modèle standard (modèle pivot)}
type Topic        = agrégat
                      transfoType : chaîne {stocke les résultats}
                      name : chaîne   reified : Association   roles: Liste de Role fin  {rôles joués dans les associations}
type Association  = agrégat
                      transfoType : chaîne {stocke les résultats}
                      reifier:Topic  type:Topic   roles: Liste de Role fin {rôle (membre) de l'association}
type Role         = agrégat  player : Topic, parent : Association fin
type TopicMap     = agrégat topics : Liste de Topic,  associations : Liste de Association fin

variables
        definitionLink, descriptionLink, assocCtxTerm : Association
        ctxTerm, parametre, contexte, terme,membre1, membre2  : Topic
début
  tant que non modelPivot.topics.horsListe() faire {vérification de l'ensemble des topics du modèle pivot}
          contextuelTerm ← modelPivot.topics.suivant() ;
          contexte, terme, membre1, membre2 ← null ;
          {vérifie que le topic n'a pas déjà été identifié et qu'il réifie une association binaire}
          si ctxTerm.transfoType==null et ctxTerm.reified !=null et
            ctxTerm.reified.getClass()== "Association" et ctxTerm.reified.roles.nombreElement()==2 alors
              definitionLink ← ctxTerm.refier ;
              membre1        ← definitionLink.roles.suivant().player ;
              membre2        ← definitionLink.roles.suivant().player ;
              {vérifie qu'il y a un terme sans réifié et un contexte qui réifie une autre association}
              si membre1.reified.getClass() =="Association" et membre2.reified==null alors
                  contexte ←membre1 ; terme ← membre2 ;
              sinon si membre1.reified==null et membre2.reified.getClass() =="Association" alors
                  contexte ←membre2 ; terme ← membre1 ;
fsi
```

```
            si contexte != null alors {si assoc definitionLink ok alors vérifier qu'aucun paramètre ne réifie un élément}
               descriptionLink ← contexte.refied ; aucunParameterReifiant ← vrai;
               tant que non descriptionLink.roles.horsListe() faire
                  si descriptionLink.roles.suivant().player.reified != null alors
                     aucunParameterReifiant ← faux ; quitterBoucle() ;
                  fsi
               ftq
               {si la structure est validée, alors on tague les objets rencontrés}
               si aucunParameterReifiant alors
                  ctxTerm.transfoType ← "ContextualTerm"; definitionLink.transfoType ← "DefinitionLink";
                  terme.transfoType   ← "Term";
                  si context.transfoType == null alors {si le contexte n'a pas déjà été identifié alors on tague}
                     contexte.transfoType ← "Context"; desciptionLink.transfoType ← "DescriptionLink";
                     tant que non descriptionLink.roles.horsListe() faire
                        descriptionLink.roles.suivant().player.transfoType ← "ContextParameter";
                     ftq
                  fsi
                  {identification des associations entre termes contextualisés (terminologique ou ontologique)}
                  tant que non ctxTerm.roles.horsListe() faire
                     assocCtxTerm ← ctxTerm.roles.suivant().parent ;
                     si assocCtxTerm.type.name in enumTerminologicType alors {sous-algorithme non traité}
                           assocCtxTerm.transfoType ← "TerminologicalLink";
                     sinon assocCtxTerm.transfoType ← "OntologicalLink"; fsi
                  ftq
               fsi {test : les paramètres ne sont pas des topics réifiants}
            fsi {test : membre de l'association binaire (terme + contexte)}
         fsi {test : association binaire}
      ftq
fin
```

## C.2 Algorithme de création de vues synthétiques

```
Algorithme creerVueSynthetique
{création des liens entre termes de base et de leur réification pour obtenir différents niveaux de détails}
Entrées / topicmap : ContextualTopicMap
Sorties  / void

classe Liste
        attributs  entête: ListElement {premier élément} ,  curseur: ListElement {élément en cours}
        fonctions suivant(), premier(), dernier(), nombreElement(), horsListe(), ajouter()  {sous-algorithmes non détaillés}
type ListElement =  agrégat
                        valeur: Info {type de base ou complexe} ,
                        suivant: ListElement {référence à un autre élément} fin
type Association = agrégat
                        transfoType : chaine , reifier:Topic fin
type ContextualTopicMap = agrégat {ContextualTopicMap est aussi une TopicMap car elle la spécialise}
                            associations : Liste d'Association fin
type ContextualTerm = agrégat {terme contextualisé}
                            termLinks : Liste de TermLink , roles: Liste de Role fin
type DefinitionLink = agrégat {association entre un terme et un contexte (spécialise le type Association)}
                            term : Term fin
type Term = agrégat {terme non contextualisé}
                        id : Chaine {subjectLocator ou sujectIdentifier (pour comparer les termes des termLink)}
                        termLinks : Liste de TermLink , definitionLinks : Liste de DefinitionLink fin
type TermLink = agrégat {association entre termes  (spécialise le type Association)}
                        terms : Liste de Term fin
variables
        association : Association , terms : Liste de Term , termLink : TermLink , term, testedTerm : Term ,
        termFound , termLinkExistant : Booléen , nbTermFound : Entier

début
  tant que non topicmap.associations.horsList() faire
       association ← topicmap.associations.suivant(); {on ne traite que les associations entre termes contextualisés}
       si association.transfoType="TerminologicalLink" ou association.transfoType="OntologicalLink' alors
        termLinkExistant  ← Faux ;
        terms             ← nouvelle Liste de Term ; {remarque : on aurait aussi pu cheminer par les RelatedTermLink ...}

        tant que non association.contextualTerms.horsList() faire {récupération des termes de base concernés}
           terms.ajouter( association.contextualTerms.suivant().reified.term ) ;
        ftq
```

```
{vérifier si un TermLink existe déjà pour lui ajouter la relation entre termes contextualisés}
      term ← terms.suivant(); {terme choisi au hasard car tous ont une référence vers le termLink, s'il existe}
      tant que non term.termLinks.horsList() faire {examine les membres de chaque termLink du terme}
          termLink       ← term.termLinks.suivant() ;
          nbTermFound ← 0 ;
          si termLink.terms.nombreElement() = = terms.nombreElement() alors {nécessite le même nombre de termes}
             tant que non termLink.terms.horsList() faire {vérifie chacun des membres}
                testedTerm ← termLink.terms.suivant() ;
                terms.premier(); {réinitialise la liste pour la vérification}
                termFound ← Faux ;
                tant que non terms.horsList() faire {vérifie que le terme fait partie du termLink}
                  si testedTerm.id== terms.suivant() .id alors
                     termFound ← Vrai  ;
                     nbTermFound ← nbTermFound +1 ;
                     quitterBoucle();
                  fsi
               ftq {terme ko, alors arrêt de la boucle pour tester un autre termLink}
                si non termFound  alors termLink.terms.dernier() ;  fsi
             ftq
             si nbTermFound  == termLink.terms.nombreElement() alors
               termLinkExistant ← Vrai;
               quitterBoucle();
             fsi
        fsi {test : nombres de termes}
      ftq {boucle de test des termLink}
      si non termLinkExistant  alors {S'il'association n'existe pas, alors créer une relation entre les termes + topic réifiant}
         termLink ← creerNouveauTermLink(terms) ; {fonction : sous-algorithme non détaillé}
      fsi {créer une association naire RelatedTermLink entre les termes contextualisés et le réifiant de termLink}
      termLink  ← creerNouveauRelatedTermLink(termLink , association);  {fonction : sous-algorithme non détaillé}
    fsi {test : association entre termes contextualisés}
  ftq {boucle de test des associations de la topic map}
fin
```

# Annexe D

# Règles de transformation M2M

Cette annexe présente quelques règles de transformation M2M spécifiées dans le cadre du prototypage des modules de la platforme de gestion de Topic Maps Contextualisées. Ces règles sont spécifiées avec un pseudo code inspiré de la syntaxe de l'outil de transformation ATL. Le principe de présentation des règles est décrit dans la figure D.1 .

**[nom]**
Nom de la règle
**[extension]**
Nom de la règle étendue
**[condition d'application]**
Condition d'applicabilité de la règle
**[transformation]**
**source** méta-modèle source : nom de la classe
**cible** méta-modèle cible : nom de la classe
**(**
*Initialisation des propriétés de la classe cible:*
Nom de l'attribut de la classe cible ← nom de l'attribut de la classe source
**)**
**[post-traitement]**
Traitements effectués une fois l'objet cible créé

FIGURE D.1 – Principe de présentation des règles ATL

La première section de cette annexe présente quelques règles de transformation modèle standard vers le modèle contextuel. La seconde section présente quelques règles de transformation spécifiées dans le cadre de la construction d'une Topic Map contextualisée.

## D.1 Règles de transformation du module Importation du modèle standard vers le modèle contextuel

**Règles générales**

Cette série de règles établit des correspondances un à un entre les éléments des modèles source et cible. Il s'agit essentiellement de règles simples qui reproduisent tels quels les objets du modèle pivot dans le modèle contextuel.

**Règle *TopicMap()***

**Règle spécifique transposant le conteneur global des topics et associations.**
**[nom]**
TopicMap
**[transformation]**
**source** Standard:TopicMap
**cible** Contextual: ContextualTopicMap
**(**
topics ← topics
associations ← associations
parameters ← toutes les instances crée dans le modèle cible
contextualTerms ← toutes les instances crée dans le modèle cible
**)**

**Règles spécialisées** Cet ensemble de règles surchargent les règles Topic() et Association() afin de créer les objets spécialisés du méta-modèle contextuel.

Règle ***Topic()***
Duplique l'ensemble des topics dans le modèle contextuel.**[nom]**
Topic**[transformation]**
**source** Standard: Topic
**cible** Contextual: Topic
**(**
id ← id
itemIdentifiers ← itemIdentifiers
subjectIdentifiers ← subjectIdentifiers
subjectLocators ← subjectLocators
occurrences ← occurrences
reified ← reified
types ← types
names ← names
**)**

Règle ***Occurrence()***
Reporte les occurrences dans le modèle cible.
**[nom]**
Occurrence
**[transformation]**
**source** Standard:Occurrence
**cible** Contextual:Occurrence
**(**
parent ← parent
value ← value
datatype ← datatype
type ← type
**)**

## D.2 Règles de transformation du module Construction d'une Topic Map contextualisée

Règle ***Name()***
Reproduit les noms des *topics* dans le modèle.
**[nom]**
Name
**[transformation]**
**source** Standard:Name
**cible** Contextual:Name
**(**
parent ← parent
value ← value
**)**

Règle ***Association()***
Calque les associations du méta-modèle standard dans la structure contextuelle.
**[nom]**
Association
**[transformation]**
**source** Standard:Association
**cible** Contextual:Association
**(**
type ← type
roles ← roles
reifier ← reifier
**)**

Règle ***Role2Role()***
Dédiée aux rôles des associations et établie des correspondances simples.
**[nom]**
Role2Role
**[transformation]**
**source** Standard: Role
**cible** Contextual: Role
**(**
parent ← parent
player ← player
type ← type
**)**

Règles ***Term(), Context(), ContextualTerm()***

Ces règles, qui spécialisent la règle *Topic()*, partagent la même structure.
**[nom]**
Term | Context | ContextualTerm
**[extension]**
Règle *topic*
**[condition d'application]**
source.transfoType = 'Term' | 'Context' | 'ContextualTerm'
**[transformation]**
**source** Standard:Topic
**cible** Contextual:Term | Contextual: Context | Contextual:ContextualTerm

Règle ***ContextParameter()***

Les paramètres de contexte demandent l'initialisation de leur valeur et de leur catégorie.
**[nom]**
ContextParameter
**[extension]**
Règle *topic*
**[condition d'application]**
source.transfoType = 'ContextParameter'
**[transformation]**
**source**Standard:*topic*C3
**cible** Contextual:ContextParameter
**(**
category ← nom du premier topic représentant son type
value ← nom du *topic*
**)**

Règle ***DescriptionLink()***

Les associations entre paramètres de contexte sont transformées avec la règle qui suit.
**[nom]**
DescriptionLink
**[extension]**
Règle Association
**[condition d'application]**
Méta-donnée transfoType = 'DescriptionLink'
**[transformation]**
**source**Standard:Association
**cible** Contextual:DescriptionLink
**(**
context ← reifier
contextParameter ← tous les membres de chaque rôle
**)**
**[post-traitement]**
context.descriptionLink ← cible
pour chaque contextParameter ajouter descriptionLinks ← cible

Règle ***DescriptionLink()***

Les associations entre paramètres de contexte sont transformées avec la règle qui suit.
**[nom]**
DescriptionLink
**[extension]**
Règle Association
**[condition d'application]**
Méta-donnée transfoType = 'DescriptionLink'
**[transformation]**
**source**Standard:Association
**cible** Contextual:DescriptionLink
**(**
context ← reifier
contextParameter ← tous les membres de chaque rôle
**)**
**[post-traitement]**
context.descriptionLink ← cible
pour chaque contextParameter ajouter descriptionLinks ← cible

Règle ***TerminologicalLink()***

Cette règle s'appuie sur la détection préalable des associations terminologiques pour établir les correspondances.
**[nom]**
TerminologicalLink
**[extension]**
Règle Association
**[condition d'application]**
Méta-donnée transfoType = 'TerminologicalLink'
**[transformation]**
**source**Standard:Association
**cible** Contextual:TerminologicalLink
**(**
linkType ← nom du topic représentant le type
contextualTerms ← tous les membres des rôles
**)**
**[post traitement]**
pour chaque contextualTerm ajouter terminologicalLinks ← cible

Règle ***OntologicalLink()***

Cette transformation s'appuie sur la détection préalable des associations ontologiques pour établir les correspondances.
**[nom]**
OntologicalLink
**[extension]**
Règle Association
**[condition d'application]**
Méta-donnée transfoType = 'OntologicalLink'
**[transformation]**
**source**Standard:Association
**cible** Contextual: OntologicalLink
**(**
contextualTerms ← tous les membres des rôles
**)**
**[post traitement]**
pour chaque contextualTerm ajouter ontologicalLinks ← cible

Règles ***Term(), Context(), ContextualTerm()***

Ces règles, qui spécialisent la règle *Topic()*, partagent la même structure.
**[nom]**
Term | Context | ContextualTerm
**[extension]**
Règle *topic*
**[condition d'application]**
source.transfoType = 'Term' | 'Context' | 'LigneDeDEclarationRDF'
**[transformation]**
**source** annotation:t
**cible** Contextual:Term | Contextual: Context | Contextual:ContextualTerm

Règle ***ContextParameter()***

Les paramètres de contexte demandent l'initialisation de leur valeur et de leur catégorie.
**[nom]**
ContextParameter
**[extension]**
Règle *topic*
**[condition d'application]**
source.transfoType = 'ContextParameter'
**[transformation]**
**source** Annotation:*EnscontexteParameter*
**cible** Contextual:ContextParameter
**(**
category ← nom du premier topic représentant son type
value ← nom du *topic*
**)**

Règle ***OntologicalLink()***

Cette transformation s'appuie sur la détection préalable des associations ontologiques pour établir les correspondances.
**[nom]**
OntologicalLink
**[extension]**
Règle Association
**[condition d'application]**
Méta-donnée transfoType = 'OntologicalLink'
**[transformation]**
**source**Standard:Association
**cible** Contextual: OntologicalLink
**(**
contextualTerms ← tous les membres des rôles
**)**
**[post traitement]**
pour chaque contextualTerm ajouter ontologicalLinks ← cible

Règles ***Term(), Context(), ContextualTerm()***

Ces règles, qui spécialisent la règle *Topic()*, partagent la même structure.
**[nom]**
Term | Context | ContextualTerm
**[extension]**
Règle *topic*
**[condition d'application]**
source.transfoType = 'Term' | 'Context' | 'LigneDeDEclarationRDF'
**[transformation]**
**source** annotation:t
**cible** Contextual:Term | Contextual: Context | Contextual:ContextualTerm

Règle ***ContextParameter()***

Les paramètres de contexte demandent l'initialisation de leur valeur et de leur catégorie.
**[nom]**
ContextParameter
**[extension]**
Règle *topic*
**[condition d'application]**
source.transfoType = 'ContextParameter'
**[transformation]**
**source** Annotation:*EnscontexteParameter*
**cible** Contextual:ContextParameter
**(**
category ← nom du premier topic représentant son type
value ← nom du *topic*
**)**

Règle ***Occurrence()***
Reporte les prédicats (est annoté) dans le modèle cible.
**[nom]**
Occurrence
**[transformation]**
**source** Annotation:EstAnnotePar
**cible** Contextual:Occurrence
**(**
parent ← parent
value ← value
datatype ← datatype
**)**

Règle ***Name()***
Reproduit les sujets (termes) dans le modèle.
**[nom]**
Name
**[transformation]**
**source** Annotation:Name
**cible** Contextual:Name
**(**
parent ← parent
value ← value
**)**

# Bibliographie

[Afnor, 2013] Afnor, editor (2013). *ISO 25964-1 - Thésaurus pour la recherche documentaire*.

[Ahmed, 2003] Ahmed, K. (2003). Tmshare–topic map fragment exchange in a peer-to-peer-application. In *Proceedings of XML Europe*, volume 2003. Citeseer.

[Ampornaramveth and Aizawa, 2001] Ampornaramveth, V. and Aizawa, A. (2001). Saikam : Collaborative japanese-thai dictionary development on the internet. In *The Asian Association for Lexicography (ASIALEX) Biennial Conference, Korea*.

[Andrews and Heidegger, 1998] Andrews, K. and Heidegger, H. (1998). Information slices : Visualising and exploring large hierarchies using cascading, semi-circular discs. In *Proc of IEEE Infovis' 98 late breaking Hot Topics*, pages 9–11.

[Arndt et al., 2011] Arndt, H.-K., Jacob, S., and Tietz, S. (2011). Multi-layer topic maps to support management-systems by structured information. In Golinska, P., Fertsch, M., and Marx-Gómez, J., editors, *Information Technologies in Environmental Engineering*, Environmental Science and Engineering, pages 61–72. Springer Berlin Heidelberg.

[Bastien, 1992] Bastien, C. (1992). Le décalage entre logique et connaissances. *Le Courrier du CNRS*, (79).

[Batini et al., 1986] Batini, C., Lenzerini, M., and Navathe, S. B. (1986). A comparative analysis of methodologies for database schema integration. *ACM computing surveys (CSUR)*, 18(4) :323–364.

[Baumeister et al., 2011] Baumeister, J., Reutelshoefer, J., and Puppe, F. (2011). Knowwe : a semantic wiki for knowledge engineering. *Applied Intelligence*, 35(3) :323–344.

[Berners-Lee, 2005a] Berners-Lee, e. a. (2005a). Internationalized resource identifiers (iris).

[Berners-Lee, 2005b] Berners-Lee, e. a. (2005b). Uniform resource identifier (uri) : Generic syntax.

[Bernhard and Rudolf, 1999] Bernhard, G. and Rudolf, W. (1999). Formal concept analysis : Mathematical foundations.

[Bernotaityte et al., 2012] Bernotaityte, G., Nemuraite, L., and Stankeviciene, M. (2012). Methodology for developing topic maps based on principles of ontology engineering. In Skersys, T., Butleris, R., and Butkiene, R., editors, *Information and Software Technologies*, volume 319 of *Communications in Computer and Information Science*, pages 406–419. Springer Berlin Heidelberg.

[Boinski et al., 2010] Boinski, T., Jaworska, A., Kleczkowski, R., and Kunowski, P. (2010). Ontology visualization. In *Information Technology (ICIT), 2010 2nd International Conference on*, pages 17–20. IEEE.

[Bolaffi, 2003] Bolaffi, G. (2003). *Dictionary of race, ethnicity and culture*. Sage.

[Bold et al., 2010] Bold, N., Kim, W.-J., and Yang, J.-D. (2010). Converting object-based thesauri into xml topic maps. In *Education Technology and Computer (ICETC), 2010 2nd International Conference on*, volume 2, pages V2–102. IEEE.

[Bond, 2007] Bond, Francis, B. J. (2007). Semi-automatic refinement of the jmdict/edict japanese-english dictionary.

[Bou, 1996] Bou, B. (1996). Treebolic.

[Brémond and Thonnat, 1997] Brémond, F. and Thonnat, M. (1997). Issues in representing context illustrated by scene interpretation applications. In *proc. of the*

*Int'l and Interdisciplinary Conf. on Modeling and Using Context (CONTEXT-97), Rio de Janeiro*.

[Brézillon, 1999] Brézillon, P. (1999). Context in problem solving : a survey. *The Knowledge Engineering Review*, 14(01) :47–80.

[Buffa et al., 2008] Buffa, M., Gandon, F., Ereteo, G., Sander, P., and Faron, C. (2008). Sweetwiki : A semantic wiki. *Web Semantics : Science, Services and Agents on the World Wide Web*, 6(1) :84–97.

[Burita, 2014] Burita, L. (2014). Knowledge management system for military universities cooperation. In *Communications (COMM), 2014 10th International Conference on*, pages 1–4.

[Cahier, 2005] Cahier, J.-P. (2005). Ontologies sémiotique pour le web socio sémantique : étude de la gestion coopérative des connaissances avec des cartes hypertopiques.

[Cahier et al., 2006] Cahier, J.-P., Turner, W. A., and Zacklad, M. (2006). A conflictual co-building method with agoræ.

[Cahier et al., 2004] Cahier, J.-P., Zacklad, M., Monceaux, A., et al. (2004). Une application du web socio-sémantique à la définition d'un annuaire métier en ingénierie. In *Ingénierie des Connaissances*.

[Canals et al., 2008] Canals, G., Molli, P., Maire, J., Laurière, S., Pacitti, E., and Tlili, M. (2008). Xwiki concerto : un wiki sur réseau p2p supportant le nomadisme. In *Proceedings of the 4th French-speaking conference on Mobility and ubiquity computing*, pages 83–85. ACM.

[CNRTL, 2012] CNRTL (2012). Défintion dictionnaire, centre national de ressources textuelles et lexicales.

[da Silva and Freitas, 2011] da Silva, I. C. S. and Freitas, C. M. D. S. (2011). Using multiple views for visual exploration of ontologies. Citeseer.

[Damen and Van den Bulcke, 2012] Damen, D. and Van den Bulcke, T. (2012). Towards a Flexible Semantic Framework for Clinical Trial Eligibility using Topic Maps. In *Proceedings of the ACM SIGKDD Workshop on Health Informatics, HI-KDD '12*, New York, NY, USA, 2012. ACM.

[Descotte, 1999] Descotte, S., J. L. H. L. R.-M. V. C. e. N. V. (1999). From specialised lexicography to conceptual databases : which format for a multilingual maritime dictionary.

[Dey, 2001] Dey, A. K. (2001). Understanding and using context. *Personal and ubiquitous computing*, 5(1) :4–7.

[Dicheva and Dichev, 2006] Dicheva, D. and Dichev, C. (2006). Tm4l : Creating and browsing educational topic maps. *British Journal of Educational Technology*, 37(3) :391–404.

[Ditcheva and Dicheva, 2007] Ditcheva, B. and Dicheva, D. (2007). Visual browsing and editing of topic map-based learning repositories. In *Leveraging the Semantics of Topic Maps*, pages 44–55. Springer.

[Dmitrieva and Verbeek, 2009] Dmitrieva, J. and Verbeek, J. (2009). Multi-view ontology visualization. In *The 11th International Protégé Conference*.

[Doi and Bester, 1977] Doi, T. and Bester, J. (1977). *The anatomy of dependence*, volume 5. Kodansha International.

[Dragu et al., 2011] Dragu, D., Gomoi, V., and Stoicu-Tivadar, V. (2011). Topic maps as knowledge base to automatically generate medical recommendations. In *Intelligent Systems and Informatics (SISY), 2011 IEEE 9th International Symposium on*, pages 459–464. IEEE.

[Dudycz, 2011] Dudycz, H. (2011). Research on usability of visualization in searching economic information in topic maps based application for return on investment indicator. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu*, (206) :45–56.

[Dury, 2000] Dury, P. (2000). Les variations sémantiques en terminologie : étude diachronique et comparative appliquée à l'écologie. *Sémantique des termes spécialisés*, 275 :17.

[Dury, 2006] Dury, P. (2006). La dimension diachronique en terminologie et en traduction spécialisée : le cas de l'écologie. *F. Gaudin et D. Candel, Aspects diachroniques du vocabulaire, Rouen : Publications des Universités de Rouen et du Havre*, pages 109–124.

[Dury, 2013] Dury, P. (2013). Que montre l'etude de la variation d'une terminologie dans le temps. quelques pistes de reflexion appliquees au domaine medical. *Debate Terminológico*, 9 :2–10.

[Dwyer and Eckersley, 2004] Dwyer, T. and Eckersley, P. (2004). Wilmascope—a 3d graph visualization system. In *Graph Drawing Software*, pages 55–75. Springer.

[Ellouze et al., 2008] Ellouze, N., Ben Ahmed, M., and Métais, E. (2008). Overview of topic map construction approaches. In *Advanced Information Networking and Applications-Workshops, 2008. AINAW 2008. 22nd International Conference on*, pages 1642–1647. IEEE.

[Ellouze et al., 2012] Ellouze, N., Lammari, N., and Métais, E. (2012). Citom : An incremental construction of multilingual topic maps. *Data & Knowledge Engineering*.

[Ellouze et al., 2010] Ellouze, N., Lammari, N., Métais, E., and Ahmed, M. B. (2010). Citom : incremental construction of topic maps. In *Natural Language Processing and Information Systems*, pages 49–61. Springer.

[Eric Prud'hommeaux, 2002] Eric Prud'hommeaux, G. M. (2002). RDF-topic-maps. `http://www.w3.org/2002/06/09-RDF-topic-maps/`.

[Eslami and Nazami, 2011] Eslami, S. and Nazami, E. (2011). An automatic approach for topic maps development using relational databases. In *Computer Research and Development (ICCRD), 2011 3rd International Conference on*, volume 1, pages 304–308.

[Eslami and Nazemi, 2013] Eslami, S. and Nazemi, E. (2013). An application of topic map-based ontology generated from wikipedia for query expansion. *International Journal of Machine Learning and Computing*, 3(4).

[Falconer, 2010] Falconer, S. (2010). Ontograf.

[Flouris et al., 2008] Flouris, G., Manakanatas, D., Kondylakis, H., Plexousakis, D., and Antoniou, G. (2008). Ontology change : Classification and survey. *The Knowledge Engineering Review*, 23(02) :117–152.

[Fuchs et al., 2013] Fuchs, N. E., Kaljurand, K., and Kuhn, T. (2013). Multilingual semantic wiki.

[Garrido et al., 2013] Garrido, A., Buey, M., Escudero, S., Ilarri, S., Mena, E., and Silveira, S. (2013). Tm-gen : A topic map generator from text documents. In *Tools with Artificial Intelligence (ICTAI), 2013 IEEE 25th International Conference on*, pages 735–740.

[Garrido and Ilarri, 2014] Garrido, A. L. and Ilarri, S. (2014). Tmr : A semantic recommender system using topic maps on the items' descriptions.

[Garshol, 2004] Garshol, L. M. (2004). Metadata ? Thesauri ? Taxonomies ? Topic Maps !Making sense of it all. . `http://www.ontopia.net/topicmaps/materials/tm-vs-thesauri.html.2004`.

[Geisser et al., 2008] Geisser, M., Happel, H.-J., Hildenbrand, T., Korthaus, A., and Seedorf, S. (2008). New applications for wikis in software engineering. In *Proceedings of the PRIMIUM Subconference at the Multikonferenz Wirtschaftsinformatik*.

[Goldhammer, 2012] Goldhammer, L. Woyand, H. B. (2012). Using topic maps to describe engineering knowledge and geometry of cad-designed products. In *Intelligent Engineering Systems (INES), 2012 IEEE 16th International Conference on*, pages 265–270.

[Gomoi et al., 2012] Gomoi, V.-S., Dragu, D., and Stoicu-Tivadar, V. (2012). Clinical decision support based on topic maps and virtual medical record. In *INTELLI 2012, The First International Conference on Intelligent Systems and Applications*, pages 71–75.

[Gruber, 1993] Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowl. Acquis.*, 5(2) :199–220.

[Hall, 2004] Hall, L. E. (2004). *Dictionary of multicultural psychology : Issues, terms, and concepts*. Sage Publications.

[Hart and Emery, 2004] Hart, L. and Emery, P. (2004). Including topic maps in the ontology definition meta-model.

[Herring, 2007] Herring, C. S. (2007). A faceted classification scheme for computer-mediated discourse. *Language@Internet*, 4(1).

[Hudon and Hadi, 2011] Hudon, M. and Hadi, W. M. e. (2011). Organisation des connaissances et des ressources documentaires. *Les Cahiers du numérique*, 6(3) :9–38.

[ISO1951, 2007] ISO1951 (2007). *ISO 1951 : 2007(E), Presentation/representation of Entries in Dictionaries : Requeriments, Recommendations and Information*. ISO.

[ISO/IEC, 2003] ISO/IEC (2003). Iso iec 13250 2003 information technology sgml applications topic maps.

[ISO/IEC, 2008] ISO/IEC (2008). Information technology — topic maps — part 2 : Data model.

[Jethro and Giles, 2007] Jethro, B. and Giles, J. (2007). Welcome to the owl2prefuse project page.

[Jose-Garcia, 2012] Jose-Garcia, I. Lopez-Arevalo, V. S.-S. (2012). Building Topic Maps from Relational Databases. *International Conference on Electrical Engineering, Computing Science and Automatic Control*.

[Jung et al., 2008] Jung, E.-H., Cho, K.-M., Song, K.-H., Nam, S.-H., and Lee, S.-W. (2008). Methodology of topic maps creation and semantic web for technological information search regarding injection-mold based on collaboration hub. In *Smart Manufacturing Application, 2008. ICSMA 2008. International Conference on*, pages 78–83. IEEE.

[Kásler et al., 2006] Kásler, L., Venczel, Z., and Varga, L. Z. (2006). Framework for semi automatically generating topic maps. In *T1R-06, Proceedings of the 3rd international workshop on text-based information retrieval. Riva del Grada*, pages 24–30.

[Katifori et al., 2007] Katifori, A., Halatsis, C., Lepouras, G., Vassilakis, C., and Giannopoulou, E. (2007). Ontology visualization methods&mdash ;a survey. *ACM Comput. Surv.*, 39(4).

[Kawamoto et al., 2006] Kawamoto, K., Kitamura, Y., and Tijerino, Y. (2006). Kawawiki : A semantic wiki based on rdf templates. In *Web Intelligence and Intelligent Agent Technology Workshops, 2006. WI-IAT 2006 Workshops. 2006 IEEE/WIC/ACM International Conference on*, pages 425–432. IEEE.

[Khattak et al., 2009] Khattak, A. M., Latif, K., Lee, S., and Lee, Y.-K. (2009). Ontology evolution : a survey and future challenges. In *U-and E-service, science and technology*, pages 68–75. Springer.

[Kiesel et al., 2006] Kiesel, M. et al. (2006). Kaukolu : Hub of the semantic corporate intranet. In *In Völkel et*. Citeseer.

[Kim et al., 2007] Kim, J.-M., Shin, H., and Kim, H.-J. (2007). Schema and constraints-based matching and merging of topic maps. *Information processing & management*, 43(4) :930–945.

[Korthaus et al., 2006] Korthaus, A., Henke, S., Aleksy, M., and Schader, M. (2006). A distributed topic map architecture for enterprise knowledge management. In *Proceedings of the 5th IEEE/ACIS International Conference on Computer and*

*Information Science and 1st IEEE/ACIS International Workshop on Component-Based Software Engineering,Software Architecture and Reuse*, ICIS-COMSAR '06, pages 96–102, Washington, DC, USA. IEEE Computer Society.

[Korthaus and Hildenbrand, 2003] Korthaus, A. and Hildenbrand, T. (2003). Creating a java- and corba-based enterprise knowledge grid using topic maps.

[Korthaus and Schader, 2006] Korthaus, A. and Schader, M. (2006). Using a topic grid and semantic wikis for ontology-based distributed knowledge management in enterprise software development processes. In *Proceedings of the 10th IEEE on International Enterprise Distributed Object Computing Conference Workshops*, EDOCW '06, pages 4–, Washington, DC, USA. IEEE Computer Society.

[Krötzsch et al., 2006] Krötzsch, M., Vrandečić, D., and Völkel, M. (2006). Semantic mediawiki. In *The Semantic Web-ISWC 2006*, pages 935–942. Springer.

[Küçü and Arslan, 2014] Küçü, D. and Arslan, Y. (2014). Semi-automatic construction of a domain ontology for wind energy using wikipedia articles. *Renewable Energy*, 62(0) :484 – 489.

[Kumar et al., 2012] Kumar, S., Schiffer, P. H., and Blaxter, M. (2012). 959 nematode genomes : a semantic wiki for coordinating sequencing projects. *Nucleic Acids Research*, 40(D1) :D1295–D1300.

[Lacher and Decker, 2001] Lacher, M. S. and Decker, S. (2001). On the integration of topic maps and rdf data. In *Proc. Extreme Markup Languages*, volume 2001.

[Lambe, 2014] Lambe, P. (2014). *Organising knowledge : taxonomies, knowledge and organisational effectiveness*. Elsevier.

[Lammari et al., 2003] Lammari, N., Akoka, J., and Comyn-Wattiau, I. (2003). Supporting database evolution : Using ontologies matching. In *Object-Oriented Information Systems, 9th International Conference, OOIS 2003, Geneva, Switzerland, September 2-5, 2003, Proceedings*, pages 284–288.

[Le Grand and Soto, 2003] Le Grand, B. and Soto, M. (2003). Topic maps visualization. In *Visualizing the Semantic Web*, pages 49–62. Springer.

[Librelotto et al., 2004] Librelotto, G. R., Ramalho, J. C., and Henriques, P. R. (2004). Tm-builder : An ontology builder based on xml topic maps. *Clei electronic journal*, 7(2).

[Lu and Feng, 2009] Lu, H. and Feng, B. (2009). An intelligent topic map-based approach to detecting and resolving conflicts for multi-resource knowledge fusion. *Information Technology Journal*, 8(8) :1242–1248.

[Lu et al., 2008a] Lu, H., Feng, B., Zhao, Y., Zheng, Q., and Liu, J. (2008a). Distributed knowledge management based on extended topic maps. In *Computer Science and Software Engineering, 2008 International Conference on*, volume 1, pages 649–652. IEEE.

[Lu et al., 2008b] Lu, H., Feng, B., Zhao, Y., Zheng, Q., and Liu, J. (2008b). A new model for distributed knowledge organization management. In *Grid and Cooperative Computing, 2008. GCC'08. Seventh International Conference on*, pages 261–265. IEEE.

[Maicher et al., 2009] Maicher, L., Ahmed, K., Isolani, A., Kivela, A., Oh, S., Pitts, A., and Vassallo, S. (2009). Topic maps in the ehumanities. In *e-Science, 2009. e-Science'09. Fifth IEEE International Conference on*, pages 6–13. IEEE.

[Maicher and Witschel, 2004] Maicher, L. and Witschel, H. F. (2004). Merging of distributed topic maps based on the subject identity measure (sim) approach. *Proceedings of Berliner XML tags*, 4 :301–307.

[Mangeot, 2006] Mangeot, M. (2006). Papillon project : Retrospective and perspectives. In *Acquiring and Representing Multilingual, Specialized Lexicons : the Case of Biomedicine, LREC workshop*.

[Mio, 1999] Mio, J. (1999). *Key Words in Multicultural Interventions : A Dictionary*. ABC-Clio ebook. GREENWOOD Publishing Group Incorporated.

[Morel-Pair, 2007] Morel-Pair, C. (2007). Métadonnées et xml. des standards efficients de l'environnement numérique. *Ingénierie des systèmes d'information*, 12(2) :9–39.

[Morris, 2007] Morris, J. C. (2007). Distriwiki : : a distributed peer-to-peer wiki network. In *Proceedings of the 2007 international symposium on Wikis*, pages 69–74. ACM.

[Munzner, 1998] Munzner, T. (1998). Drawing large graphs with h3viewer and site manager. In *Graph Drawing*, pages 384–393. Springer.

[NISO, 2004] NISO (2004). Understanding metadata. Technical report.

[Noy et al., 2001] Noy, N. F., Sintek, M., Decker, S., Crubézy, M., Fergerson, R. W., and Musen, M. A. (2001). Creating semantic web contents with protege-2000. *IEEE intelligent systems*, 16(2) :60–71.

[OmegaWiki, 2009] OmegaWiki (2009).

[OMG, 2009] OMG (2009). Ontology definition metamodel (omg) version 1.0. Technical Report formal/2009-05-01, Object Management Group.

[OMG, 2014a] OMG (2014a). Object management group model driven architecture (mda)mda guide rev. 2.0. Technical report.

[OMG, 2014b] OMG (2014b). Omg-mof, meta object facility. Technical report.

[ontopia, 2001] ontopia (2001). Ontopia.

[Oster et al., 2005] Oster, G., Urso, P., Molli, P., Imine, A., et al. (2005). Edition collaborative sur réseau pair-à-pair à large échelle. In *Journées Francophones sur la Cohérence des Données en Univers Réparti-CDUR 2005*, pages 42–47.

[Parsia et al., 2005] Parsia, B., Wang, T., and Golbeck, J. (2005). Visualizing web ontologies with cropcircles. In *Proceedings of the 4th International Semantic Web Conference*, pages 6–10.

[Paul-Jacques and Patrice, 1995] Paul-Jacques, L. and Patrice, M. (1995). *Le Referis : droit des affaires, comptable, gestion financière, fiscal, droit communautaire, social, gestion budgétaire. A à H, Volume 1.* Maxima.

[Psyché et al., 2003] Psyché, V., Mendes, O., Bourdeau, J., et al. (2003). Apport de l'ingénierie ontologique aux environnements de formation à distance. *Revue des Sciences et Technologies de l'Information et de la Communication pour l'Education et la Formation (STICEF)*, 10 :89–126.

[Rahhal et al., 2008] Rahhal, C., Skaf-Molli, H., Molli, P., et al. (2008). Swooki : A peer-to-peer semantic wiki.

[Roy et al., 2013] Roy, S., Sawant, K. P., and Charvin, O. M. (2013). Generating topic maps from xml/xsd documents. In *Proceedings of the 6th ACM India Computing Convention*, page 13. ACM.

[Sasha and Rudan, 2008] Sasha, R. and Rudan, S. (2008). SocioTM -Relevancies, Collaboration, and Socio-knowledge in Topic Maps. *TMRA Leipzig*.

[Schaffert et al., 2006] Schaffert, S., Westenthaler, R., and Gruber, A. (2006). Ikewiki : A user-friendly semantic wiki. In *ESWC*.

[Schwotzer, 2008] Schwotzer, T. (2008). Building Context Aware P2P Systems with the Shark Framework. *TMRA Leipzig*.

[Sergieh et al., 2013] Sergieh, H. M., Egyed-Zsigmond, E., Döller, M., Gianini, G., Kosch, H., and Pinon, J. (2013). Tag similarity in folksonomies. In *Actes du XXXIème Congrès INFORSID, Paris, France, 29-31 Mai 2013.*, pages 319–334.

[Shneiderman, 1992] Shneiderman, B. (1992). Tree visualization with tree-maps : 2-d space-filling approach. *ACM Transactions on graphics (TOG)*, 11(1) :92–99.

[Shneiderman, 1996] Shneiderman, B. (1996). The eyes have it : A task by data type taxonomy for information visualizations. In *Visual Languages, 1996. Proceedings., IEEE Symposium on*, pages 336–343. IEEE.

[Sigel, 2004] Sigel, A. (2004). kPeer as a context-aware topic map P2P application for the distributed integration of knowledge.

[Smoot et al., 2011] Smoot, M. E., Ono, K., Ruscheinski, J., Wang, P.-L., and Ideker, T. (2011). Cytoscape 2.8 : new features for data integration and network visualization. *Bioinformatics*, 27(3) :431–432.

[Smoot et al., 2013] Smoot, M. E., Ono, K., Ruscheinski, J., Wang, P.-L., and Ideker, T. (2013). Cytoscape : An open source platform for complex network analysis and visualization.

[Sowa, 2000] Sowa, J. (2000). *Knowledge Representation : Logical, Philosophical and Computational Foundations*. Brooks/Cole, Pacific Grove, CA.

[Steele, 2009] Steele, T. (2009). The new cooperative cataloging. *Library Hi Tech*, 27(1) :68–77.

[Steinberg et al., 2008] Steinberg, D., Budinsky, F., Merks, E., and Paternostro, M. (2008). *EMF : eclipse modeling framework*. Pearson Education.

[Stevens, 2007] Stevens, P. (2007). Bidirectional model transformations in qvt : Semantic issues and open questions. In *Model Driven Engineering Languages and Systems*, pages 1–15. Springer.

[Storey et al., 1994] Storey, V., Goldstein, R., Chiang, R., and Dey, D. (1994). A commonsense reasoning facility based on the entity-relationship model. In Elmasri, R., Kouramajian, V., and Thalheim, B., editors, *Entity-Relationship Approach — ER '93*, volume 823 of *Lecture Notes in Computer Science*, pages 218–229. Springer Berlin Heidelberg.

[Studer et al., 1998] Studer, R., Benjamins, V. R., and Fensel, D. (1998). Knowledge engineering : Principles and methods. *Data and Knowledge Engineering*, 25(1-2) :161–197.

[Sure et al., 2003] Sure, Y., Angele, J., and Staab, S. (2003). Ontoedit : Multifaceted inferencing for ontology engineering. In *Journal on Data Semantics I*, pages 128–152. Springer.

[Svensson, 2009] Svensson, M. (2009). Contextual metadata in practice. In *Advances in Multimedia, 2009. MMEDIA'09. First International Conference on*, pages 12–17. IEEE.

[Team, 2014] Team, W. (2014). wondora.

[tm4j, 2006] tm4j (2006).

[TMdesigner, 2006] TMdesigner (2006).

[Vara Mesa, 2009] Vara Mesa, J. M. (2009). M2dat : a technical solution for model-driven development of web information systems.

[W3C, 2005] W3C (2005). Editing 'internationalized resource identifiers (iris)'.

[Wang and Guo, 2007] Wang, H.-C. and Guo, J.-L. (2007). Automatic construction of topic maps based on lexical chain—using avia-tion safety knowledge base as a case.

[Weber et al., 2010a] Weber, G. E., Eilbracht, R., and Kesberg, S. (2010a). Topic maps for improved access to and use of content in relational databases – a case study on the descriptive variety lists of germany's bundessortenamt. In *TMRA2010*, pages 191–198.

[Weber et al., 2010b] Weber, G. E., Eilbracht, R., and Kesberg, S. (2010b). Topic maps for subject-centric publishing from document-centric content management systems – a case study on a website of a regional cluster of companies. In *TMRA2010*, pages 191–198.

[Weller, 2010] Weller, K. (2010). *Knowledge Representation in the Social Semantic Web*. K G Saur Verlag, 1 edition.

[Weller et al., 2010] Weller, K., Peters, I., and Stock, W. G. (2010). Folksonomy. the collaborative knowledge organization system. *Handbook of research on social*

*interaction technologies and collaborative software : Concepts and trends*, pages 132–146.

[wikipédia, 2002] wikipédia (2002).

[Will, 2012] Will, L. (2012). The iso 25964 data model for the structure of an information retrieval thesaurus. 38.

[Xue et al., 2010] Xue, Y., Liu, W., Feng, B., and Cao, W. (2010). Merging of topic maps based on corpus. In *Electrical and Control Engineering (ICECE), 2010 International Conference on*, pages 2840–2843. IEEE.

[Zablith et al., 2013] Zablith, F., Antoniou, G., d'Aquin, M., Flouris, G., Kondylakis, H., Motta, E., Plexousakis, D., and Sabou, M. (2013). Ontology evolution : a process-centric survey. *The Knowledge Engineering Review*, pages 1–31.

[Zaher et al., 2007] Zaher, L., Cahier, J.-P., Lejeune, C., and Zacklad, M. (2007). Construction coopérative de carte de thèmes : vers une modélisation de l'activité socio-sémantique. *Revue des Nouvelles Technologies de l'Information*, 1(E-9).

[Zaher et al., 2006] Zaher, L., Cahier, J.-P., and Zacklad, M. (2006). The agoræ/hypertopic approach.

[Zhang et al., 2014] Zhang, Z., Sang, J., Ma, L., Wu, G., Wu, H., Huang, D., Zou, D., Liu, S., Li, A., Hao, L., Tian, M., Xu, C., Wang, X., Wu, J., Xiao, J., Dai, L., Chen, L.-L., Hu, S., and Yu, J. (2014). Ricewiki : a wiki-based database for community curation of rice genes. *Nucleic Acids Research*, 42(D1) :D1222–D1228.

[Zheng et al., 2009] Zheng, Q., Wu, Z., Jiang, L., and Liu, J. (2009). A collaborative knowledge construction system design for massive knowledge resources. In *Computer Supported Cooperative Work in Design, 2009. CSCWD 2009. 13th International Conference on*, pages 137–142.

**Résumé**

Cette thèse s'inscrit dans le cadre de la construction et de l'évolution de Topic Maps en tant que ressources sémantiques servant à l'organisation de contenus pluridisciplinaires et multilingues. Cette Topic Map vise à prendre en charge la variation du sens des termes afin d'assurer une meilleure recherche d'informations au sein d'un contenu. Une approche de visualisation de cette Topic Map a également été proposée. La problématique de cette thèse a découlé du programme FSP-Maghreb qui est un projet franco-maghrébin initié par la FMSH (Fondation Maison des Sciences Humaines et Sociales). Ce projet vise à promouvoir l'échange et le partage de connaissances dans le domaine des sciences humaines et sociales. Ce projet consiste en la construction et la mise en œuvre d'un Wiktionnaire (dictionnaire électronique implémenté sous la technologie wiki sémantique) multilingue et multiculturel pour les sciences humaines et sociales.

**Mots-clés** : Topic Maps contextualisées, Wiktionnaire multilingue et multiculturel, variation du sens des termes, construction et évolution de Topic Maps.

**Abtsract:**

This thesis concerns the construction and the evolution of Topic Maps as semantic ressource used to describe and organise multidisciplinary and multilingual contents. This Topic Map aims to support variation of meaning to ensure a better information retrieval in content.

The problematic of this research work is resulted from a project initiated by the FMSH called FSP-Maghreb. This project allows exchanges between Maghrebi and French researchers. It also allows the sharing of knowledge related to the two cultures and to the two societies in the human and social sciences. This project consists of the construction of an on-line multicultural and multilingual dictionary based on wiki technology.

**Keywords:** Contextualized Topic Maps, Multicultural and multilingual Wiktionary, variation of meaning, building Topic Maps, evolution of Topic Maps.