
Optimisation des relations multidimensionnelles pour une ontologie dédiée à la conception de systèmes décisionnels

Ahlem Nabli*** — Jamel Feki***

* Laboratoire MIRACL

Route de Tunis km 9, 3000 Sfax, Tunisie

{Ahlem.nabli,Jamel.feki}@fsegs.rnu.tn

** Faculté des sciences de Sfax

*** Faculté des sciences Economiques et de Gestion de Sfax

RÉSUMÉ. Les systèmes décisionnels (SD) sont dédiés au management de l'entreprise pour l'aider à la prise de décision. La construction de ces systèmes est une tâche complexe et difficile qui exige un effort intellectuel considérable. Parallèlement, les ontologies sont à l'heure actuelle au cœur des travaux menés en ingénierie des connaissances et ont fait preuve dans plusieurs domaines. Une ontologie dédiée à la conception de SD est un modèle d'organisation des connaissances pour un domaine donné. Elle représente les concepts multidimensionnels d'un domaine ainsi que leurs relations multidimensionnelles (R_{MD}) et sémantiques. Après avoir rappelé notre méthode de construction d'ontologie décisionnelle, nous détaillerons dans ce papier la phase d'optimisation des R_{MD} suite à l'enrichissement de l'ontologie. Cette phase est régie par des règles d'optimisation que nous définissons.

ABSTRACT. Decisional Support systems (DSS) are dedicated to the management of the enterprise to help decision makers in the decision process. The construction of these systems is a complex and difficult task that requires a considerable intellectual effort. From the other hand, ontologies are at the present time at the main of works led in knowledge engineering and gave proof in several domains. An ontology dedicated to the design of DSS is a model of organization of the knowledge for a given domain. It represents the multidimensional concepts of a domain, as well as their multidimensional and semantic relations. After recalling our method of construction of decisional ontology, we will detail in this paper the phase of multidimensional relations optimization at the time of the ontology enrichment. This phase is accomplished by optimization rules which we define.

MOTS-CLÉS: Système décisionnel, ontologie décisionnelle, conception, règles d'optimisation.

KEYWORDS: Decisional Support System, decisional ontology, design, optimization rules.

1. Introduction

Les Systèmes d'information Décisionnels (SD) constituent un domaine d'expertise de plus en plus répandue dans les entreprises soucieuses d'extraire l'information pertinente et cachée dans leurs systèmes opérationnels (Kimball *et al.*, 2000). Ces SD sont dédiés au pilotage des activités de l'entreprise et, constituent une synthèse des informations opérationnelles, internes ou externes, choisies pour leur pertinence et leur transversalité fonctionnelles. Les SD sont basés sur des structures particulières de stockage volumineux appelées entrepôt de données.

Les Entrepôts de Données (ED) (Bonifati *et al.*, 2001) (Kimball, 1996) sont construits par extraction de données à partir de sources d'information et, par intégration de ces données dans un espace de stockage commun à tous les utilisateurs décisionnels. L'ED est important pour plusieurs applications, spécialement pour celles évoluant dans des environnements à grande échelle et faisant appel à des sources d'information distribuées (Moody *et al.*, 2000) (Ruzzi, 2004). Les problèmes posés par la construction d'un entrepôt à partir de plusieurs sources de données concernent la définition de son schéma, son chargement et sa mise à jour, en fonction des différentes sources de données à partir desquelles il est construit (Vassiliadis *et al.*, 2000).

Une bonne conception d'un SD doit tenir compte de certains aspects essentiels pour la réussite du projet décisionnel, à l'instar de la spécification des besoins décisionnels, la modélisation multidimensionnelle et les contraintes de qualité. En effet, le couplage entre la conception et l'ingénierie des besoins est crucial et doit privilégier les points suivants (Paim *et al.*, 2002) :

- Représenter les faits et leurs mesures : les faits représentent les centres d'intérêt des SD.
- Etablir les liens entre les dimensions et les faits : les dimensions représentent les axes d'analyse selon lesquels un fait peut être analysé.
- Assurer l'additivité des mesures : vérifier la contrainte de *Summarizability*.
- Intégrer les données à partir de sources hétérogènes.
- Faciliter et accélérer la prise en charge des besoins des décideurs.

Ces différents points nécessitent une réflexion approfondie sur une méthode de conception consistante de SD dans laquelle la spécification des besoins décisionnels doit être bien fondée. En effet, la méthode doit assister les futurs utilisateurs d'un ED qui sont des décideurs généralement incapables d'exprimer avec précision et à l'avance leurs besoins analytiques.

Parallèlement, les ontologies sont à l'heure actuelle au cœur des travaux menés en Ingénierie des Connaissances (IC). Ces travaux visent l'établissement de représentations à travers lesquelles les machines puissent manipuler la sémantique des informations. Les ontologies apparaissent comme des composants logiciels

s'insérant dans les SI en leur apportant une dimension sémantique qui leur faisait défaut jusqu'ici. Le champ d'application des ontologies ne cesse de s'élargir et couvre les systèmes conseillers (systèmes d'aide à la décision, systèmes d'enseignement assisté par ordinateur), les systèmes de résolution de problèmes ou les systèmes de gestion de connaissances. Plus précisément, les ontologies ont donné une grande satisfaction pour résoudre les conflits sémantiques et structurels entre les différentes sources de données lors d'un processus d'intégration de BD (Ruzzi, 2004) (Bellatrache *et al.*, 2006).

Récemment, quelques méthodologies ont été proposées pour soutenir le développement des ontologies et les mettre en place dans un système d'intégration (Ruzzi, 2004)(Nguyen Xuan *et al.*, 2006).

Dans un travail antérieur (Nabli *et al.*, 2005), nous avons proposé une approche de conception de système décisionnel basée sur les besoins OLAP («On Line Analytical Processing») tabulaires et dans laquelle nous avons introduit le concept d'ontologie décisionnelle comme un moyen d'assistance de l'utilisateur décisionnel à l'expression de ses besoins analytiques (Nabli *et al.*, 2006). Le recours à une ontologie est motivé par sa capacité à résoudre des ambiguïtés sémantiques et syntaxiques. En effet, elle représente un référentiel contenant un ensemble de concepts ainsi que leurs liens caractérisant un domaine donné.

Nous avons proposé une architecture de construction d'ontologie décisionnelle à base de concepts multidimensionnels (Nabli *et al.*, 2006). Cette architecture se prête à construire l'ontologie d'une manière incrémentale et progressive et se base sur quatre phases (c.f. Figure 1) : i) l'*extraction* des concepts multidimensionnels à partir de sources de données hétérogènes, ii) la *comparaison* des concepts extraits de la phase précédente avec celle de l'ontologie afin de déduire les relations sémantiques, iii) l'*alimentation* de l'ontologie par les concepts et les relations extraits et, enfin, iv) l'*optimisation* des relations multidimensionnelles.

Dans cet article nous nous intéressons principalement à la quatrième phase, c'est-à-dire, à l'optimisation des relations multidimensionnelles. Cet article est organisé comme suit. La section 2 introduit le concept d'Ontologie Décisionnelle (OD) pour la conception de SD et rappelle notre démarche de sa construction. La section 3 détaille les règles d'optimisation des relations multidimensionnelles suite à la déduction des relations sémantiques. La section 4, synthétise la proposition de l'article et présente nos perspectives.

2. Ontologie décisionnelle

Le terme 'Ontologie' a plusieurs définitions. Une des plus simples et populaires est celle de Gruber (Gruber, 1993) «*Une ontologie est une spécification formelle explicite d'une conceptualisation partagée*». Les qualificatifs "*formelle*" et "*explicite*" signifient qu'une ontologie permet une interprétation automatisée de la

conceptualisation par la machine. Autrement dit, une ontologie définit un vocabulaire commun pour les utilisateurs qui ont besoin de se partager l'information sur un domaine particulier. Ces utilisateurs peuvent être des personnes, des bases de données et/ou des applications. Les ontologies incluent des définitions, des informations exploitables, des concepts élémentaires ainsi que leurs relations. Elles codent une connaissance dans un domaine particulier (pouvant s'étendre à plusieurs domaines). De cette façon, les ontologies permettent la réutilisation des connaissances. Entre autres, elles permettent de (Pierra, 2005) :

— *Partager la compréhension commune de la structure de l'information.*

— *Permettre la réutilisation du savoir sur un domaine.*

— *Expliciter ce qui est considéré comme implicite sur un domaine.*

— *Distinguer le savoir sur un domaine du savoir opérationnel.*

Dans nos travaux antérieurs (Nabli *et al.*, 2006a) nous avons justifié le recours à une ontologie pour la conception d'un système décisionnel.

Une telle ontologie décisionnelle est une représentation de connaissances dédiées aux systèmes décisionnels. Elle peut être définie comme un référentiel des concepts multidimensionnels d'un domaine, ainsi que leurs relations multidimensionnelles et sémantiques. Son utilisation couvre les différents niveaux du cycle de vie d'un SD, c'est-à-dire, depuis la spécification des besoins, la conception du schéma de l'ED et des magasins de données jusqu'aux phases d'exploitation et d'évolution du SD. Durant ces phases, elle aide à résoudre les problèmes d'hétérogénéité des sources de données par la résolution des ambiguïtés sémantiques. En phase de spécification des besoins OLAP, elle permet, d'une part, de proposer des éléments multidimensionnels et d'autre part, de prévenir des associations '*illégales*' entre concepts telle que, par exemple, l'association d'une dimension *Fournisseur* à un fait *Vente*, ou d'une mesure *quantité vendue* à un fait *achat*, etc. En plus l'ontologie décisionnelle évite les associations entre concepts non associables, telle qu'une association fait-fait, dimension-dimension, hiérarchie-fait, etc. Cette prévention est assurée par la présence des relations sémantiques dans l'ontologie décisionnelle, alors que les relations multidimensionnelles assurent des associations valides entres concepts multidimensionnels.

Nous avons proposé dans (Nabli *et al.*, 2006a) une démarche de construction d'OD, basée sur les quatre étapes suivantes : i) l'extraction, ii) la comparaison, iii) l'optimisation et iv) l'alimentation. La Figure 1 schématise ces étapes.

Notre démarche de construction d'une ontologie décisionnelle est progressive et itérative. Des interventions du concepteur, tout au long de cette construction, sont nécessaires, voire obligatoires. En effet, elles permettent l'approbation des résultats obtenus et la résolution de quelques ambiguïtés.

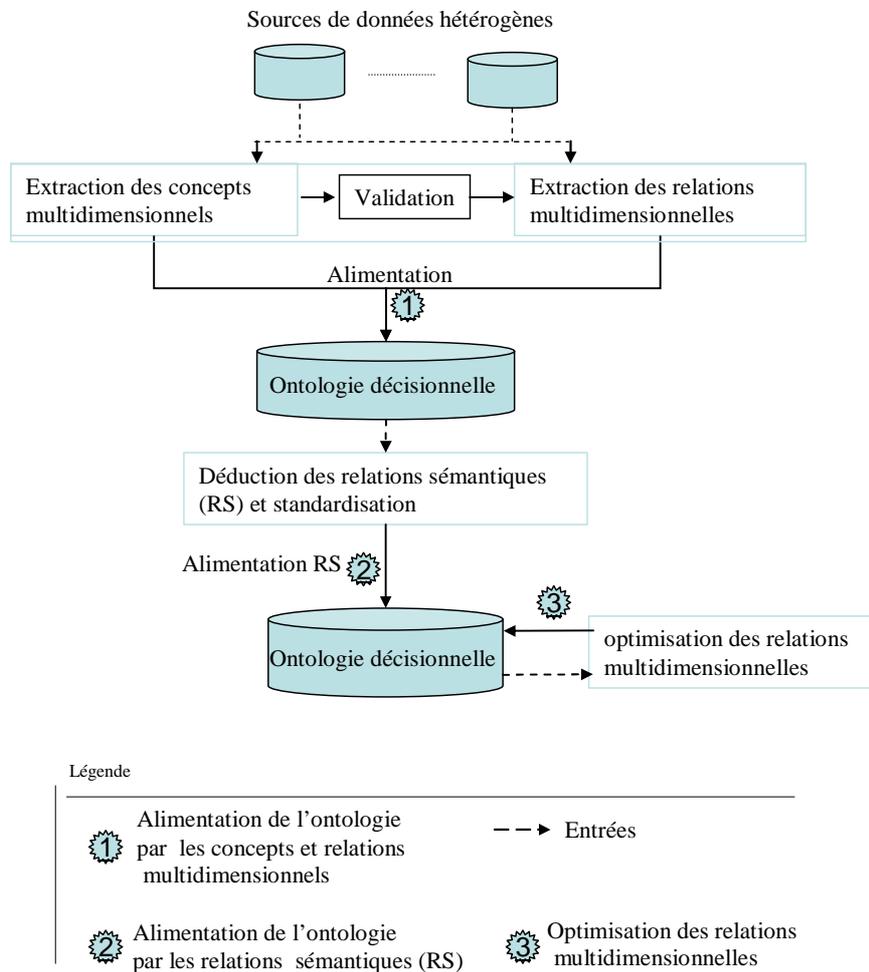


Figure 1. Architecture générale d'un système de construction d'ontologies décisionnelles

La suite de cette section introduit les concepts d'une ontologie décisionnelle puis, explique brièvement les phases d'extraction des éléments multidimensionnels, la déduction des relations sémantiques et la standardisation.

2.1. Concepts de base pour l'OD

L'ontologie décisionnelle est une représentation de connaissances dédiées aux systèmes décisionnels. Elle peut être définie comme un référentiel des éléments

multidimensionnels, à savoir les concepts et relations d'un domaine ainsi que leurs relations sémantiques (Nabli *et al.*, 2006a) (Nabli *et al.*, 2006b).

2.1.1. *Éléments multidimensionnels*

La modélisation multidimensionnelle s'appuie sur les concepts multidimensionnels (C_{MD}) de : *Fait*, *Mesure*, *Dimension*, *Paramètre*, *Attribut faible* et *Hiérarchie*. Les C_{MD} sont reliés entre eux par des *relations* qui définissent les différentes associations de la sémantique multidimensionnelle. Nous avons identifié les relations multidimensionnelles (R_{MD}) suivantes :

- *Est_Dimension* (F, D) indique que D est une dimension possible du fait F .
- *Est_Mesure* (F, M) indique que M est une mesure possible du fait F .
- *Est_Identifiant* (D, P) indique que P est l'identifiant (paramètre de plus faible granularité) de la dimension D .
- *Est_hierarchie* (D, H) indique que H est une hiérarchie de la dimension D .
- *Est_Parametre* (H, P) indique l'appartenance d'un paramètre P à une hiérarchie H .
- *Est_Suivie* (P_r, P_j) indique que le paramètre P_r précède le paramètre P_j .
- *Est_faible* (P, AF) indique que AF est un attribut faible pour le paramètre P .

2.1.2. *Relations sémantiques*

Les Relations Sémantiques (RS) relient les occurrences des concepts multidimensionnels de différentes sources de données traitant le même domaine. Les RS supportées par notre ontologie sont : *l'identité*, *la synonymie*, *l'équivalence* et *l'homonymie*.

La déduction des relations sémantiques entre occurrences de concepts est réalisée dans le module 3 de la figure 1. Suite à l'alimentation de l'ontologie par ces éléments, des redondances peuvent avoir lieu. L'optimisation élimine les liens multidimensionnels redondants et, en conséquence, améliore les résultats des interrogations et minimise le temps d'accès.

2.2. *Extraction des éléments multidimensionnels*

L'ensemble des éléments multidimensionnels potentiels et susceptible d'alimenter notre ontologie décisionnelle est composé par : les faits à analyser, les mesures d'analyse, les dimensions d'analyse de faits et leurs attributs et également des relations multidimensionnelles entre concepts.

L'extraction des éléments multidimensionnels est composée de trois phases itératives (cf. figure 1) : une première phase d'extraction manuelle ou semi-

automatique des concepts, une deuxième phase de validation effectuée par le concepteur de l'ontologie, et enfin, une troisième phase de détermination des relations multidimensionnelles entre les concepts extraits.

Ces trois phases se répètent pour chaque concept multidimensionnel (Fait, Mesure, Dimension, Hiérarchie, Paramètre, Attributs faibles). Dans (Nabli *et al.*, 2006a) une définition complète des heuristiques d'extraction de ces éléments multidimensionnels à partir d'une source E/A est présentée.

Le résultat de cette étape est un document XML contenant l'ensemble des occurrences d'éléments multidimensionnels extrait à savoir les concepts et les relations multidimensionnelles.

2.3. Déduction des relations sémantiques et standardisation

La deuxième étape de construction de l'ontologie décisionnelle consiste à comparer chaque C_{MD} valide issu de l'étape précédente avec l'ontologie. Cette comparaison permet de déterminer les RS entre un sous ensemble de concepts ayant le même type. Les RS permettent de relier les C_{MD} des différentes sources de données traitant le même domaine. Elles apportent plus de sémantique aux concepts extraits.

Une fois les relations sémantiques entre occurrences de concepts multidimensionnels sont déduites, une phase de standardisation est réalisée. Cette dernière consiste à proposer, d'une part, des noms génériques pour tout ensemble de concepts similaires et d'autre part, des noms distincts pour tout couple d'occurrences ayant des noms homonymes. Cette phase de standardisation est réalisée moyennant des heuristiques détaillées dans (Nabli *et al.*, 2006b).

Tout au long de la phase de standardisation et d'optimisation les relations sémantiques sont classées en deux catégories. Une première catégorie, dite *homogène*, est composée des relations d'identité, d'équivalence et de synonymie. La deuxième catégorie, dite *hétérogène*, est associée à la relation d'homonymie.

La phase de standardisation permet d'affecter :

- Un nom générique pour l'ensemble des occurrences des concepts ayant une relation sémantique *homogène*.
- Un nom différent pour l'une des occurrences de concept intervenant dans cette relation sémantique *hétérogène*.

Le résultat de cette phase est un ensemble d'occurrences de concepts multidimensionnels reliés par les relations sémantiques (cf. 2.1.2.) et standardisées. Ensuite, cet ensemble est utilisé comme un vocabulaire commun lors de la phase de conception de SD.

3. Optimisation des relations multidimensionnelles

Suite à des alimentations répétées, la taille de l'ontologie augmente et son interrogation devient de plus en plus lente. Par exemple, pour deux faits synonymes les ensembles de leurs mesures peuvent être redondants ce qui accroît le nombre de leurs relations sémantiques et multidimensionnelles. Pour pallier à ce dilemme, nous envisageons optimiser le contenu de l'ontologie.

Cette optimisation permet de découvrir l'impact des relations sémantiques déduites entre occurrences de concepts multidimensionnels sur les R_{MD} extraites. Pour cela, nous définissons un ensemble d'heuristiques ; elles touchent toute occurrence de concept ayant fait l'objet soit d'une généralité de concepts pour la catégorie *homogène* soit par l'attribution de nom différent pour la catégorie *hétérogène*. L'optimisation de la catégorie *homogène* transfère vers le nom générique et/ou élimine les R_{MD} appliquées aux occurrences de C_{MD} sources de standardisation. Alors que pour la catégorie *hétérogène*, il y a juste transfert des R_{MD} du concept, avant standardisation, vers le concept après standardisation.

Dans ce qui suit nous définissons les règles d'optimisation des relations multidimensionnelles pour chaque type de concept.

3.1. Optimisation des relations multidimensionnelles associées au concept *Fait*

Chaque occurrence du concept *Fait* est associée, d'une part à au moins une occurrence du concept *Mesure* à travers la $R_{MD} Est_Mesure$ et, d'autre part, à au moins une occurrence du concept *Dimension* à travers la $R_{MD} Est_Dimension$. L'optimisation des R_{MD} associées au concept *Fait*, porte sur les $R_{MD} Est_Mesure$ et $Est_Dimension$. Pour cela nous définissons des règles pour chacune des deux catégories et pour chaque type de relations faisant intervenir le concept *Fait*.

3.1.1. Optimisation dans la catégorie *homogène*

Soit $EO_F = \{O_{F1}, O_{F2}, \dots, O_{Fn}\}$ l'ensemble d'occurrences homogènes du concept *Fait* pour lequel la standardisation a produit le terme générique O_{FG} . Du fait que chaque O_F est associée à ses mesures et ses dimensions, nous définissons les règles d'optimisation liées aux mesures puis celles liées aux dimensions.

Optimisation liée aux Mesures

Soit MO_{Fi} un ensemble d'occurrences du concept *Mesure* associé à une occurrence de *Fait* $O_{Fi} \in EO_F$.

$$\forall O_{Fi} \in EO_{Fi} \wedge \forall O_m \in MO_{Fi} \wedge Est_Mesure(O_{Fi}, O_m) \wedge \neg Est_Mesure(O_{FG}, O_m) \Rightarrow Est_Mesure(O_{FG}, O_m) \wedge \neg Est_Mesure(O_{Fi}, O_m) \quad [1]$$

$$\forall O_{Fi} \in EO \wedge \forall O_m \in MO_{Fi} \wedge Est_Mesure(O_{Fi}, O_m) \wedge Est_Mesure(O_{FG}, O_m) \Rightarrow \neg Est_Mesure(O_{Fi}, O_m) \quad [2]$$

L'optimisation liée aux mesures nécessite, d'une part, le transfert des mesures MO_{fi} d'un fait O_{fi} non générique à son fait générique correspondant O_{FG} sans redondance. Ce transfert est assuré par l'application de la règle 1. D'autre part, l'élimination des mesures appartenant simultanément à un fait non générique et à son fait générique correspondant. Cette élimination est assurée par la règle 2.

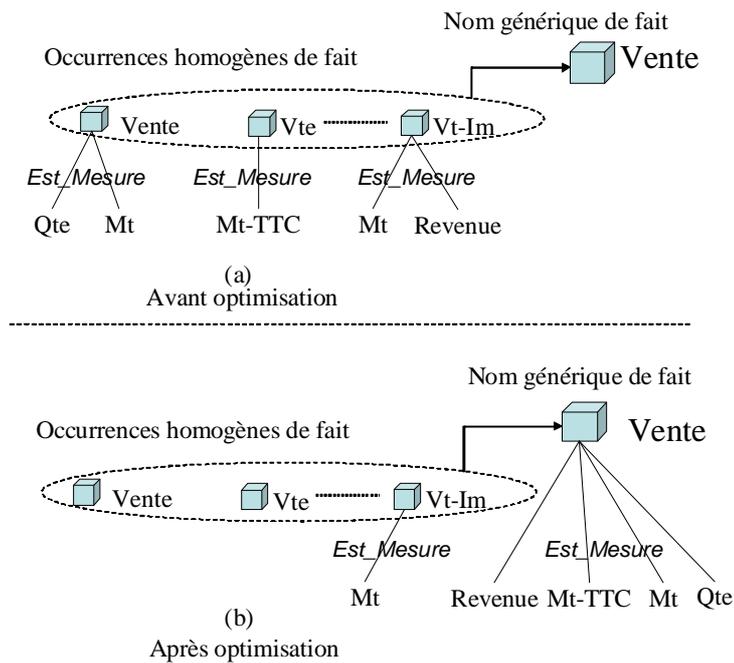


Figure 2. Optimisation de R_{MD} Est_Mesure pour le concept Fait dans la catégorie homogène.

Exemple : La figure 2.a présente un ensemble d'occurrences homogène ('Vente', 'Vte', 'Vt-Im') du fait générique 'Vente' et les mesures de chaque élément de cet ensemble. L'application de la règle 1 transfère les mesures des faits 'Vente', 'Vte' et 'Vt-Im' vers le nom générique 'Vente' comme le montre la Figure 2.b. L'application de la règle 2 élimine le lien entre le fait Vt-Im et sa mesure Mt du fait que celle-ci est devenue une mesure du fait générique 'Vente'.

Optimisation liée aux Dimensions

Soit DO_{fi} un ensemble d'occurrences du concept *Dimension* associé à une occurrence de fait $O_{fi} \in EO_F$.

L'optimisation liée aux dimensions nécessite, dans un premier temps, le transfert sans redondance des dimensions DO_{fi} d'un fait O_{fi} non générique à son fait générique

correspondant O_{FG} ; ce transfert est assuré par la règle 3. Dans un deuxième temps, l'optimisation élimine, s'il existe, les dimensions appartenant simultanément à un fait non générique et à son fait générique correspondant. Cette élimination est assurée par la règle 4.

$$\forall O_{Fi} \in EO_F \wedge \forall O_d \in DO_{Fi} \wedge Est_Dimension(O_{Fi}, O_d) \wedge \neg Est_Dimension(O_{FG}, O_d) \Rightarrow Est_Dimension(O_{FG}, O_d) \wedge \neg Est_Dimension(O_{Fi}, O_d) \quad [3]$$

$$\forall O_{Fi} \in EO_F \wedge \forall O_d \in DO_{Fi} \wedge Est_Dimension(O_{Fi}, O_d) \wedge Est_Dimension(O_{FG}, O_d) \Rightarrow \neg Est_Dimension(O_{Fi}, O_d) \quad [4]$$

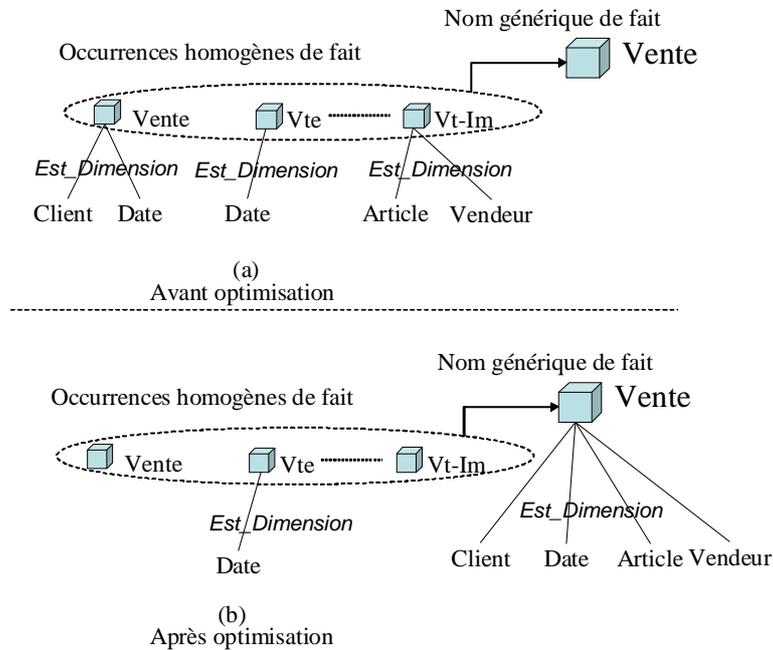


Figure 3. Optimisation de $R_{MD} Est_Dimension$ pour le concept Fait dans la catégorie homogène.

Exemple : La figure 3.a présente un ensemble d'occurrences homogène du fait générique Vente et les dimensions de chacune. L'application de la règle 3 transfère les dimensions des faits 'Vente', 'Vte' et 'Vt-Im' vers le nom générique du fait 'Vente' comme le montre la Figure 3.b. L'application de la règle 4 élimine le lien entre le fait Vt-Im et sa dimension Date devenue une dimension du fait générique 'Vente'.

3.1.2. Optimisation dans la catégorie hétérogène

Soit O_F une occurrence du concept *Fait* pour laquelle la phase de standardisation a attribué un nouveau nom O_{FV} . Cette attribution nécessite le transfert des mesures et des dimensions de O_F à O_{FV} . Nous définissons les règles d'optimisation liées aux mesures et celle liées aux dimensions.

Optimisation liée aux Mesures

Soit MO_{Fi} un ensemble d'occurrences du concept *Mesure* associé à O_F .

L'optimisation liée aux mesures transfère toutes les mesures du fait O_F à son fait O_{FV} standard. La règle 5 assure ce transfert.

$$\begin{aligned} \forall O_m \in M_{OF} \wedge Est_Mesure(O_F, O_m) \Rightarrow \\ Est_Mesure(O_{FV}, O_m) \wedge \neg Est_Mesure(O_F, O_m) \end{aligned} \quad [5]$$

Prenons l'exemple de la figure 4.a présentant deux occurrences de fait et possédant la même syntaxe ; c'est-à-dire que lors de la phase de comparaison la relation d'homonymie ($homonymie(Vte, Vte)$) à été appliquée sur ces occurrences, ce qui a nécessité l'attribution d'un nom différent 'Vt_Etrang' à l'une des deux occurrence en phase de standardisation. L'optimisation pour la catégorie hétérogène transfère les relations multidimensionnelles utilisant le fait 'Vte' par 'Vt_Etrang'. Cette optimisation est assurée par l'application de la règle 5 permettant le transfert des mesures de 'Vte' à 'Vt_Etrang' comme présentée dans la figure 4.b.

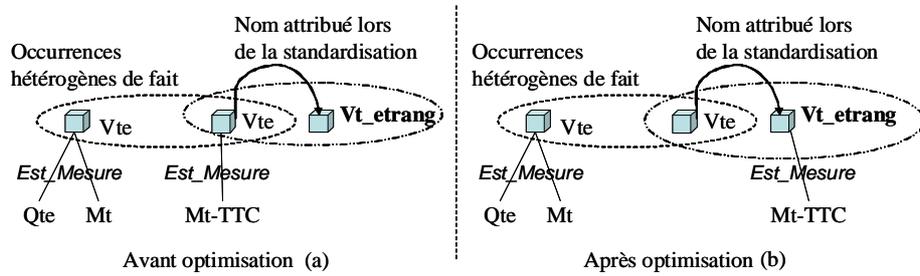


Figure 4. Optimisation de $R_{MD} Est_Mesure$ pour le concept *Fait* dans la catégorie hétérogène.

Optimisation liée aux Dimensions

Soit DO_{Fi} , un ensemble d'occurrences du concept *Dimension* associé à O_F .

L'optimisation liée aux dimensions transfère toutes les dimensions du fait O_F à son fait O_{FV} correspondant. Ce transfert est assuré par la règle 6.

$$\begin{aligned} \forall O_d \in DO_F \wedge Est_Dimension(O_F, O_d) \Rightarrow \\ Est_Dimension(O_{FV}, O_d) \wedge \neg Est_Dimension(O_F, O_d) \end{aligned} \quad [6]$$

3.2. Optimisation des relations multidimensionnelles associées au concept Dimension

Notons que chaque occurrence du concept *Dimension* opère sur les quatre R_{MD} *Est_Hierarchie*, *Est_Dimension*, *Est_Identifiant* et *Est_Suivie*. En effet, chaque occurrence du concept *Dimension* est associée, d'une part à au moins une occurrence du concept *Fait* à travers la R_{MD} *Est_Dimension* et d'autre part, à au moins une occurrence du concept *Hierarchie* à travers la R_{MD} *Est_Hierarchie*. Chaque dimension possède un seul paramètre vérifiant la R_{MD} *Est_Identifiant*. De plus, chaque paramètre peut être suivie d'un ou plusieurs paramètres à travers la R_{MD} *Est_Suivie*. L'optimisation de ces relations est assurée par un ensemble de règles pour chaque catégorie.

3.2.1. Optimisation dans la catégorie homogène

Soit $EO_D = \{O_{D1}, O_{D2}, \dots, O_{Dk}\}$ un ensemble d'occurrences du concept *Dimension* pour lequel la phase de standardisation a produit le terme générique O_{DG} .

Optimisation liée aux Faits

Soit FO_D un ensemble d'occurrences du concept *Fait* associé à une occurrence de dimension $O_D \in EO_D$.

Chaque dimension appartenant à EO_D est associé à au moins une occurrence du concept *fait* utilisant la R_{MD} *Est_Dimension*. En conséquence, l'optimisation liée aux faits transfère les faits FO_D d'une dimension non générique O_D à sa dimension générique correspondante O_{DG} en éliminant la redondance. La règle 7 assure ce transfert.

$$\forall O_D \in EO_D \wedge \forall O_F \in FO_D \wedge Est_Dimension(O_F, O_D) \wedge \neg Est_Dimension(O_F, O_{DG}) \Rightarrow Est_Dimension(O_F, O_{DG}) \wedge \neg Est_Dimension(O_F, O_{Di}) \quad [7]$$

Optimisation liée aux Hiérarchies

Soit HO_D un ensemble d'occurrences du concept *Hiérarchie* associé à une occurrence de dimension $O_D \in EO_D$.

Chaque dimension appartenant à EO_D est associé à au moins une occurrence du concept *hiérarchie* utilisant la R_{MD} *Est_Hierarchie*. Alors, l'optimisation liée aux hiérarchies transfère les hiérarchies HO_D d'une dimension non générique O_D à sa dimension générique correspondante O_{DG} sans redondance. Ensuite, elle élimine les hiérarchies appartenant simultanément à une dimension non générique O_D et à sa dimension générique O_{DG} correspondante. Cette optimisation est assurée par les règles 8 et 9.

$$\forall O_D \in E_{OD} \wedge \forall O_h \in HO_D \wedge Est_Hierarchie(O_D, O_h) \wedge \neg Est_Hierarchie(O_{DG}, O_h) \Rightarrow Est_Hierarchie(O_{DG}, O_h) \wedge \neg Est_Hierarchie(O_D, O_h) \quad [8]$$

$$\forall O_D \in E_{OD} \wedge \forall O_h \in HO_D \wedge Est_Hierarchie(O_D, O_h) \wedge Est_Hierarchie(O_{DG}, O_h) \Rightarrow \neg Est_Hierarchie(O_D, O_h) \quad [9]$$

Optimisation liée aux identifiants

Soient :

- $EO_{id} = \{O_{id1}, O_{id2}, \dots, O_{idk}\}$ un ensemble des k occurrences du concept *Paramètre* représentant les identifiants des k dimensions de EO_D .

- O_{idG} le nom générique des identifiants de EO_{id} déterminé lors de la phase de standardisation.

La standardisation de l'ensemble des occurrences de dimensions (EO_D) implique la standardisation de leurs identifiants EO_{id} . En conséquence, l'optimisation liée aux identifiants des dimensions est assurée, d'une part, par l'affectation de l'identifiant générique O_{idG} à sa dimension générique correspondante O_{DG} et, d'autre part, par l'élimination des identifiants des dimensions non génériques appartenant à EO_D .

Pour cela nous définissons les règles 10 et 11 d'optimisation des R_{MD} *Est_Identifiant*.

$$Est_Identifiant(O_{DG}, O_{IdG}) \quad [10]$$

$$\forall k \in [1..K] \wedge Est_Identifiant(O_{Dk}, O_{IdG}) \Rightarrow \neg Est_Identifiant(O_{Dk}, O_{Idk}) \quad [11]$$

Pour illustrer l'optimisation des relations multidimensionnelles associées au concept *Dimension*, nous travaillons sur la Figure 5.a présentant les occurrences de dimension 'Produit', 'Article', 'Marchandise' et leur identifiant 'IdProd', 'IdArt' et 'IdMar' respectivement et, où chaque dimension est associée à ses hiérarchies à travers la R_{MD} *Est_hierarchie*. Supposons que le nom générique attribué à cet ensemble homogène de dimensions (respectivement identifiant) est 'Article' (respectivement 'IdArt'). L'application de la règle 8 associe toutes les hiérarchies des trois dimensions homogènes à la dimension générique nommée 'Article'. Alors que la règle 10 affecte l'identifiant générique 'IdArt' à la dimension générique 'Article'. Finalement, la règle 11 élimine toutes les relations *Est_Identifiant* entre les trois dimensions et leur identifiant non générique. Le résultat de l'application de ces règles est présenté dans la Figure 5.b.

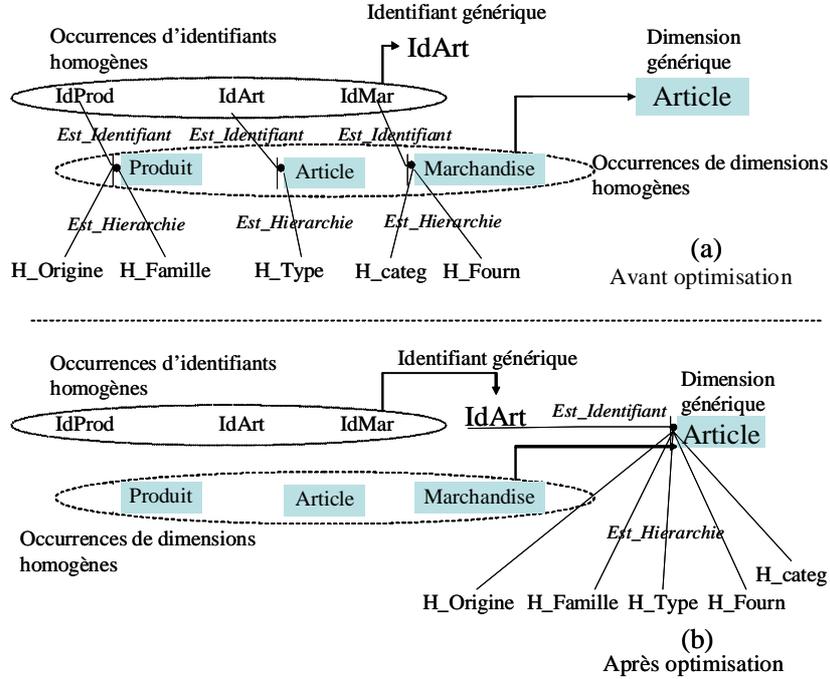


Figure 5. Optimisation des R_{MD} $Est_Hierrarchie$ et $Est_Identifier$ pour le concept $Dimension$.

Optimisation liée aux paramètres

Soit $PO_{id} = \{O_{p1}, O_{p2}, \dots, O_{ps}\}$ l'ensemble d'occurrences du concept $Paramètre$ reliées directement à une occurrence d'identifiant O_{id} .

Une occurrence du concept paramètre jouant le rôle d'identifiant pour une dimension peut être suivie (relation Est_Suivie) par d'autres paramètres pour former une hiérarchie. L'optimisation de cette relation nécessite le transfert des paramètres successeurs PO_{id} d'un identifiant non générique O_{id} à son identifiant générique correspondant O_{idG} . Ce transfert est assuré par la règle 12.

$$\forall O_{id} \in EO_{id} \wedge \exists O_P \in PO_{id} \wedge Est_Suivie(O_{id}, O_P) \Rightarrow Est_Suivie(O_{idG}, O_P) \wedge \neg Est_Suivie(O_{id}, O_P) \quad [12]$$

La règle 13 élimine les paramètres successeurs appartenant simultanément à un identifiant non générique O_{id} et à son identifiant générique correspondant O_{idG} .

$$\forall O_{id} \in EO_{id} \wedge \forall O_P \in EOP \wedge Est_Suivie(O_{id}, O_P) \wedge Est_Suivie(O_{idG}, O_P) \Rightarrow \neg Est_Suivie(O_{idG}, O_P) \quad [13]$$

Poursuivons l'illustration de l'optimisation des relations multidimensionnelles

associées aux concept *Dimension* ; La Figure 6.a présente les occurrences des identifiants 'IdProd', 'IdArt' et 'IdMar' où chaque identifiant est associée à ses paramètres successeurs à travers la $R_{MD} Est_Suivie$.

Comme présenté dans la figure 5.a le nom générique attribué à cet ensemble homogène d'identifiants est 'IdArt'. L'application de la règle 12 associe tous les successeurs des trois identifiants homogènes à l'identifiant générique nommée 'IdArt'. Le résultat est présenté dans la Figure 6.b.

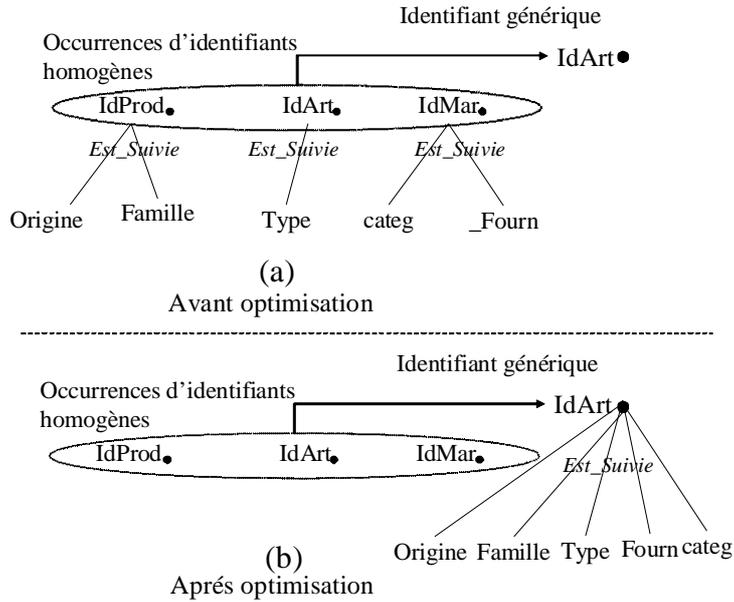


Figure 6. Optimisation de la $R_{MD} Est_Suivie$ pour le concept *Dimension*

3.2.2. Optimisation dans la catégorie hétérogène

Soit O_{D1} une occurrence du concept *Dimension* pour laquelle la phase de standardisation a attribué un nouveau nom O_{DV} . Cette attribution nécessite le transfert de l'identifiant, des faits et des hiérarchies associés O_{D1} vers O_{DV} . Ceci est assuré par les règles 10, 14 et 15.

Optimisation liée aux Faits

Soit FO_D un ensemble d'occurrences du concept *Fait* associé à une occurrence de dimension O_D .

L'optimisation liée aux faits transfère les faits de la dimension O_D à sa dimension correspondante O_{DV} . Cette opération est assurée par la règle 14.

$$\forall O_F \in FO_D \wedge Est_Dimension(O_F, O_D) \Rightarrow Est_Dimension(O_F, O_{DV}) \wedge \neg Est_Dimension(O_F, O_D) \quad [14]$$

Optimisation liée aux Hiérarchies

Soit HO_D , un ensemble d'occurrences du concept *Hiérarchie* associé à une occurrence de dimension O_D .

L'optimisation liée aux hiérarchies transfère les hiérarchies de la dimension O_D vers sa dimension O_{DV} . Cette opération est réalisée par la règle 15.

$$\begin{aligned} \forall O_h \in HO_D \wedge Est_Hierarchie(O_D, O_h) \Rightarrow \\ Est_Hierarchie(O_{DV}, O_h) \wedge \neg Est_Hierarchie(O_D, O_h) \end{aligned} \quad [15]$$

Optimisation liée aux identifiants

Chaque dimension possède un identifiant. La modification de l'occurrence O_D par O_{DV} nécessite le transfert de l'identifiant de O_D à O_{DV} . La règle 16 effectue ce transfert.

$$Est_Identifiant(O_D, O_{Id}) \Rightarrow Est_Identifiant(O_{DG}, O_{IdG}) \wedge \neg Est_Identifiant(O_D, O_{Id}) \quad [16]$$

3.3. Optimisation des relations multidimensionnelles associées au concept *Mesure*

Notons que chaque concept *Mesure* est associée à au moins une occurrence du concept *Fait* à travers la R_{MD} *Est_Mesure*. L'optimisation de la R_{MD} associée au concept *Mesure* est régie par deux règles pour chacune des deux catégories.

3.3.1. Optimisation dans la catégorie homogène

Soit $EO_M = \{O_{M1}, O_{M2}, \dots, O_{Mn}\}$ un ensemble d'occurrences du concept *Mesure* pour lequel la phase de standardisation a produit le terme générique O_{MG} .

Optimisation liée aux Faits

Soit FO_M un ensemble d'occurrences du concept *Mesure* associé à une occurrence de mesure $O_M \in EO_M$.

L'optimisation liée aux faits transfère les faits FO_D d'une mesure non générique O_M à sa mesure générique correspondante O_{MG} . La règle 17 permet ce transfert.

$$\begin{aligned} \forall O_M \in EO_M \wedge \forall O_F \in FO_M \wedge Est_Mesure(O_F, O_M) \Rightarrow \\ Est_Mesure(O_F, O_{MG}) \wedge \neg Est_Mesure(O_F, O_M) \end{aligned} \quad [17]$$

Prenons l'exemple de la figure 7.a présentant les occurrences de mesure homogènes 'Mt' 'MT-TTC' et 'Mt' associée respectivement à 'Vente', 'Vte', 'Vt-Im' ; ces mesures sont standardisées par le nom générique 'Montant'. L'application de la règle 17 transfère les trois mesures 'Mt' 'MT-TTC' et 'Mt' associée respectivement à 'Vente', 'Vte', 'Vt-Im' par le nom générique 'Montant' associé à chacun des faits 'Vente', 'Vte', 'Vt-Im' comme le montre la figure 7.b. Poursuivons

l'optimisation, la règle 2 transfère les mesures des faits 'Vente', 'Vte', 'Vt-Im' (cf. figure 7.c).

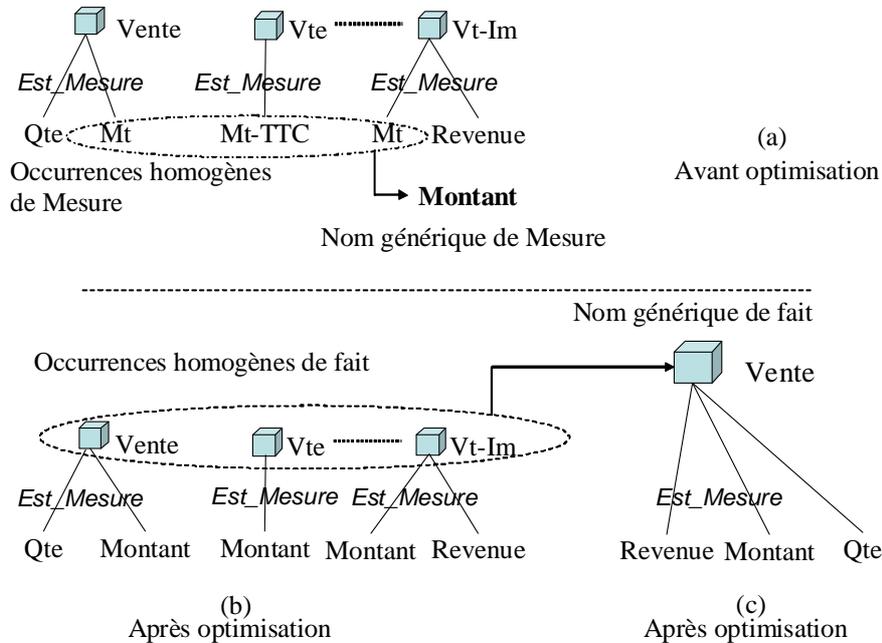


Figure 7. Optimisation de la R_{MD} Est_Mesure pour le concept Mesure

L'exemple de la figure 2.a est composé de trois faits, cinq mesures et cinq occurrences de la relation Est_Mesure. Une première optimisation de cet exemple suite à l'application des règles 2 et 3 présente un seul fait, quatre mesures et quatre relations Est_Mesure (cf. Figure 2.b). Nous avons repris dans l'exemple précédent le même exemple celui de la figure 2.a sur lequel nous avons appliqué successivement les règles 17, 2 et 3. Le résultat de cette deuxième optimisation a donné un seul fait, trois mesures et trois relations Est_Mesure. Donc l'optimisation permet de minimiser la taille de l'ontologie et cela à travers le transfert et l'élimination des relations multidimensionnels et l'élimination de la redondance. En conséquence l'optimisation améliore les résultats d'interrogations et minimise le temps d'accès lors d'un processus de conception de système décisionnelle.

3.3.2. Optimisation dans la catégorie hétérogène

Soit O_M une occurrence du concept Mesure pour laquelle la phase de standardisation a attribué un nouveau nom O_{MV} . Cette attribution nécessite le transfert des faits de O_M à O_{MV} .

Optimisation liée aux Faits

Soit FO_M un ensemble d'occurrences du concept *Fait* associé à une occurrence de dimension O_M .

L'optimisation liée aux faits pour la catégorie hétérogène, transfère les faits de la mesure O_M à sa correspondante O_{MV} . Cette opération est assurée par la règle 18.

$$\begin{aligned} \forall O_F \in FO_M \wedge Est_Mesur(O_F, O_M) \Rightarrow \\ Est_Mesur(O_F, O_{MV}) \wedge \neg Est_Mesur(O_F, O_M) \end{aligned} \quad [18]$$

3.4. Optimisation des relations multidimensionnelles associées au concept hiérarchie

Chaque occurrence du concept *Hiérarchie* est associée à au moins une dimension à travers la $R_{MD} Est_Hierarchie$. Chaque hiérarchie est composée de paramètre. Cette composition est assurée par la $R_{MD} Est_Parametre$. L'optimisation des R_{MD} associées au concept *Hierarchie* porte sur les relations multidimensionnelles *Est_Hierarchie* et *Est_Parametre*. Elle est assurée par cinq règles permettant le transfert et/ou l'élimination de ces R_{MD} .

3.4.1. Optimisation dans la catégorie homogène

Soit $EO_H = \{O_{H1}, O_{H2}, \dots, O_{Hn}\}$ un ensemble d'occurrences du concept *Hiérarchie* pour lequel la phase de standardisation à produit le terme générique O_{HG} .

Optimisation liée aux Dimensions

Soit DO_{Hi} , l'ensemble d'occurrences du concept *Dimension* associé à une occurrence de hiérarchie $O_{Hi} \in EO_H$.

Chaque occurrence du concept *Hiérarchie* de EO_H est relié à une occurrence du concept *Dimension* par la relation *Est_Hierarchie*. L'optimisation liée aux dimensions transfère les dimensions associées à chaque hiérarchie O_{Hi} par la hiérarchie générique correspondante O_{HG} . Cette opération est assurée par la règle 19.

$$\begin{aligned} \forall O_{Hi} \in EO_H \wedge \forall O_d \in DO_{Hi} \wedge est_hierarchie(O_d, O_{Hi}) \Rightarrow \\ est_hierarchie(O_d, O_{HG}) \wedge \neg est_hierarchie(O_d, O_{Hi}) \end{aligned} \quad [19]$$

Optimisation liée aux Paramètres

Soit PO_{Hi} un ensemble d'occurrences du concept *Paramètre* associé à $O_{Hi} \in EO_H$. Chaque hiérarchie est composée d'un ensemble de paramètre. Cette composition est assurée par la $R_{MD} Est_Parametre$. Par conséquent, chaque hiérarchie O_{Hi} qui participe dans une standardisation nécessite le transfert de ses paramètres à la hiérarchie générique correspondante O_{HG} . Ce traitement est assuré par la règle [20].

$$\begin{aligned} & \forall O_{Hi} \in EO_H \wedge \forall O_p \in PO_{Hi} \wedge est_parametre(O_{Hi}, O_p) \wedge \neg est_parametre(O_{HG}, O_p) \quad [20] \\ & \Rightarrow est_parametre(O_{HG}, O_p) \wedge \neg est_parametre(O_{Hi}, O_p) \end{aligned}$$

Suite au transfert, l'optimisation élimine les paramètres appartenant simultanément à une hiérarchie non générique O_{Hi} et à sa hiérarchie générique correspondante O_{HG} .

$$\begin{aligned} & \forall O_{Hi} \in EO_H \wedge \forall O_p \in PO_{Hi} \wedge est_parametre(O_{Hi}, O_p) \wedge est_parametre(O_{HG}, O_p) \Rightarrow \quad [21] \\ & \neg est_parametre(O_{Hi}, O_p) \end{aligned}$$

4. Conclusion

Dans ce papier nous avons proposé un ensemble de règles permettant l'optimisation des relations multidimensionnelles dans une ontologie dédiée à la conception de systèmes décisionnels. Cette optimisation se base sur les relations sémantiques entre les concepts multidimensionnels déduites lors de la phase de comparaison. Suite à cette comparaison, une standardisation est réalisée pour l'attribution de termes génériques avec classification des relations sémantiques en deux catégories *homogène* et *hétérogène*. Nos règles d'optimisation sont définies pour chaque catégorie et pour chaque concept multidimensionnel. Elles permettent le compactage des relations multidimensionnelles utilisées pour chaque type de concept ainsi que l'élimination des liens multidimensionnels redondants. En conséquence, cette optimisation produit des résultats concis et améliore les temps de réponse lors du processus de conception de systèmes décisionnels (SD). Comme perspectives, nous visons l'implémentation des règles d'optimisation proposées et l'intégration de notre ontologie décisionnelle dans notre outil d'aide à la conception de SD.

5. Bibliographie

- Bellatrache L., Nguyen Xuan D., Pierra G., Hondjack D., «Contribution of Ontology-based Data Modeling to Automatic Integration of Electronic Catalogues within Engineering Databases» *Computers in Industry Journal*, 2006.
- Bonifati A. F., Cattaneo S., Fuggetta A., araboschi S., «Designing Data Marts for Data Warehouse», *ACM Transaction on Software Engineering and Methodology*, ACM, vol. 10, Octobre 2001, pp. 452-483.
- Moody D., M. Kortnik, «From Enterprise Models to Dimensional Models: A Methodology for Data Warehouse and Data Mart Design», *DMDW'00*, Suède, 2000.
- Gruber T.R., «Toward Principles for the Design of Ontologies, Used for Knowledge Sharing», *Stanford Knowledge Systems Laboratory*, 1993.
- Grüniger M., Fox M., «Methodology for the Design and Evaluation of Ontologies»,

- Workshop on Basic Ontological Issues in Knowledge Sharing, IJCAI, 1995, Montreal.
- Kimball R., «The Data Warehouse Toolkit», *John Wiley and Sons, Inc.*, New York, 1997.
- Kimball R, Reeves L, Thornthwaite M., *Concevoir et déployer un data warehouse Guide de conduite de projet*, Eyrolles, 2000.
- Lehner W., Albrecht J., Edekind H., «Normal Forms for Multidimensional Database». 8th International Conference on Statistical and Scientific Database Management (SSDBM), IEEE Computer Society, 1998.
- Mhiri M., Mtibaa A., Gargouri F., «Towards an approach for building information systems'ontologies», *1st workshop "FOMI'2005» Formal Ontologies Meet Industry*, Verona, Italy, June, 9-10, 2005.
- Nabli A., Soussi A., Feki J., Ben Abdallah H., Gargouri F., « Towards an automatic Data Mart Design » ICEIS, 2005.
- Nabli A., Feki J., Gargouri F., «An ontology based method for normalisation of multidimensional terminology». SITIS'06, 21-24 Décembre, Tunis, 2006.
- Nabli A., Feki J., Mhiri M., Gargouri F., «Vers une approche de construction d'ontologie décisionnelle : Extraction des Éléments Multidimensionnels». MCSEAI'06, 6-9 Décembre, Maroc, 2006.
- Nguyen Xuan D., Bellatreche L., PIERRA G., «Un modèle à base ontologique pour la gestion de l'évolution asynchrone des entrepôts de données», 6eConférence Francophone de MOdélisation et SIMulation - MOSIM'06 - du 3 au 5 avril 2006 - Rabat – Maroc.
- Noy N., McGuinness D., «*Ontology Development 101: A Guide to Creating Your First Ontology*», Stanford Medical Informatics Report, SMI-2001-0880. 2001.
- Vassiliadis P., «Gulliver in the Land of Data Warehousing: Practical Experiences and Observations of a Researcher», 2nd International Workshop on Design and Management of Data Warehouses (DMDW), Stockholm, Sweden, 2000, pp. 12.1-12.16.
- Paim F. R. S. and Castro J. B., «Enhancing Data Warehouse Design with the NFR Framework», 5th Workshop on Requirements Engineering (WER2002), Valencia, Spain, November 11-12, 2002, pp. 40-57.
- Paim F. R. S., Carvalho A. E., Castro J. B., «Towards a Methodology for Requirements Analysis of Data Warehouse Systems», XVI Simpósio Brasileiro de Engenharia de Software (SBES'2002), Gramado, Rio Grande do Sul, Brasil, October 16-18, 2002, pp. 146-161.
- Pierra G., «Context-explication in conceptual ontologies : PLIB ontologies and their use for industrial data», *Journal of Advanced Manufacturing Systems*, 2005.
- Ruzzi M., «*Data Integration : state of the art, new issues and research plan*», 2004.