

Vers la gestion des bases de données fondée sur les logiques de descriptions

H. Kherbachi (), Z. Bouzidi (*), A. Hocine (**)*

() Laboratoire Economie du Développement,
Université Abderrahmane Mira,*

Targa Ouzemmour 06000 Béjaïa (Algérie)

E-mail : zair_bouzidi(a)hotmail.com

*(**) Laboratoire LICIAF, Université de Pau et des Pays
de l'Adour, route de l'Université 64000 Pau (France)*

E-mail : Amrane.Hocine(q).crisv2.univ-pau.fr

1. Introduction :

De nombreuses bases de données, qui **sont** implantées dans des systèmes différents, évoluent depuis de longues années. Mais devant l'évolution croissante des systèmes de gestion de bases de données et la prolifération rapide des réseaux d'interconnexion, l'on est amené à faire partager les informations relevant d'un intérêt commun entre ces bases de données existantes et autonomes pour les faire coopérer les unes avec les autres tout en maintenant, en même temps, leurs autonomies. Pour y accéder d'une façon intégrée, l'utilisateur n'a aucune idée sur les hétérogénéités syntaxiques et sémantiques qui existent parmi les bases de données.

Les hétérogénéités syntaxiques qui sont des conséquences d'utilisation des modèles de données différents, sont résolues d'une façon commune par l'utilisation d'un modèle canonique dans lequel les schémas locaux seront convertis.

Cependant, même si tous les schémas sont exprimés dans le même modèle de données, un concept unique du monde réel peut être représenté de différentes façons dans les bases de données en aboutissant à des hétérogénéités sémantiques qui seront identifiées pour accéder d'une façon intégrée.

En plus, non seulement les concepts équivalents doivent être détectés mais aussi, tout type de relation sémantique parmi les concepts de bases de données doit être relevé.

La détection des relations sémantiques entre bases de données exige une connaissance approfondie de la sémantique de la base de données.

Cependant, ces connaissances ne sont pas connues notamment lorsque l'on se retrouve avec plusieurs bases de données indépendantes et différentes comme dans le contexte d'interopérabilité. En plus, les schémas de bases de données ne permettent pas d'acquérir ces connaissances du fait de la pauvreté de leur sémantique due à l'expression limitée des modèles de données traditionnels dans lesquels ces schémas sont exprimés.

Une solution pour surmonter cette limitation (des schémas) consiste à élever leur niveau sémantique par un processus d'enrichissement sémantique où les connaissances sont découvertes et rendues explicites en convertissant les schémas dans un modèle canonique plus riche.

Ce processus d'enrichissement est fondé sur les logiques de descriptions qui consistent à comprendre les aspects structurels et sémantiques des bases existantes dans le but de découvrir de nouvelles connaissances permettant le passage à des modèles plus évolués.

2. Enrichissement Sémantique des Bases de Données

Le sujet de passage d'un schéma de base de données d'un modèle à un autre est étudié par plusieurs auteurs. Beaucoup de travaux se réfèrent au passage d'un modèle plus sémantique à un modèle moins sémantique, particulièrement du modèle Entité-Association ou une extension au modèle relationnel [6].

Cependant, le sujet de passage d'un modèle moins sémantique à un modèle plus sémantique tel qu'il est nécessaire pour l'interopérabilité est moins étudié. En effet, il est plus qu'un passage, du moment que des connaissances doivent être acquises d'une certaine manière dans le but d'élever le schéma à un niveau sémantique plus élevé.

Un bref aperçu de tous les travaux sur la conversion en un modèle plus riche utilisant le modèle relationnel comme modèle de données source et le modèle Entité-Association ou une extension du modèle Entité-Association comme le modèle de données final est présenté dans [5]. Ces travaux ne sont pas considérés comme adéquat au contexte d'interopérabilité pour les raisons suivantes:

- la dichotomie Entités, Associations de la famille Entité-Association crée plusieurs problèmes aussi bien dans le processus de conversion que dans celui de l'intégration ;
- pour acquérir les connaissances nécessaires afin d'élever le niveau sémantique des schémas, elles relient entièrement le DBA pour obtenir

toutes les connaissances nécessaires interactivement, ou assument qu'elles sont spécifiées d'avance.

Ces travaux considèrent uniquement les connaissances implicites partielles, en considérant seulement les clés primaires, en effectuant des substitutions de clés candidates, ou en assumant que les schémas sont bien désignés et considèrent uniquement les dépendances d'inclusion à base de clés.

Le modèle BLOOM [5], qui est un modèle orienté objet, inclut une phase d'acquisition des connaissances où le schéma et l'extension de la base de données sont analysés pour découvrir toute la sémantique implicite. Cependant, il focalise, comme les autres travaux, le nombre d'opérations et le nombre d'accès disque, comme facteur primaire de coût dans les algorithmes qui analysent les extensions de bases de données.

3. Approche fondée sur les logiques de descriptions

Notre approche, par contre, pour enrichir sémantiquement les bases de données consiste, tout en préservant les investissements antérieurs, à comprendre les aspects structurels et sémantiques des bases existantes dans le but de découvrir de nouvelles connaissances permettant le passage à des modèles plus évolués fondés sur les Logiques de Descriptions du langage de concepts. Ces Logiques de Descriptions sont :

- non seulement des outils de représentation et de raisonnement en intelligence artificielle;
- mais aussi adaptées à la représentation de la sémantique des situations de l'univers réel à modéliser. Les modèles de données sémantiques tels que Entité Association, peuvent être exprimés en utilisant ces logiques de descriptions. Ces dernières permettent également d'exprimer des requêtes et définir des vues organisées dans un graphe d'héritage.

Le raisonnement déductif au travers d'une terminologie est réalisé par le mécanisme de classification [2], [3], [8], fondé sur la relation de subsomption, qui permet de classer des concepts dans un graphe. Ce mécanisme de classification permet:

- de vérifier la cohérence des descriptions ;
- d'éliminer certaines redondances en utilisant la notion d'héritage ;
- et de rendre plus explicites les contraintes.

La relation de subsomption représente le mécanisme le plus important dans les Logiques de Descriptions. Avec l'adoption d'une sémantique ensembliste,

l'on permet de donner une définition formelle à la relation de subsomption et de définir la subsomption comme une fonction logique notée "Subsume" appliquée à une paire de concepts.

Ainsi, notre approche consiste à :

- identifier le contenu du dictionnaire (noms de relations, noms d'attributs, et leurs liens) ;
- mettre en évidence la sémantique des domaines qui n'est pas explicitement présente dans le système étudié (clés étrangères, dépendances diverses, ...) ;
- proposer des spécifications possibles correspondant à cette sémantique ;
- et représenter le résultat dans un modèle canonique.

Ceci nous permet de promouvoir le niveau sémantique des schémas existants en utilisant des modèles plus riches du moment que le modèle canonique est défini dans le langage de concepts, qui est développé au sein de notre laboratoire comme outil de représentation et de raisonnement.

L'hétérogénéité à différents niveaux (sémantiques, structurels, descriptifs) implique la traduction des bases de données locales dans un modèle canonique plus riche dont le choix est fondamental car il joue un rôle important dans la comparaison des schémas locaux et dans la détection des conflits inter schémas.

3.1 Implémentation de cette approche

Ce travail implémenté dans le langage de concepts, qui est développé au sein de notre laboratoire comme outil de représentation et de raisonnement, est une extension du langage centré objet RECOS [2], [3]. Il utilise la notion de concepts et de rôles pour produire des descriptions qui sont, en plus, utilisées dans le raisonnement taxinomique des Logiques de Descriptions [1], [4], [7].

3.2 Les Logiques de Descriptions :

Les logiques de descriptions décrivent la structure des objets, à un niveau terminologique, en termes de concepts et de rôles [1], [7], [9]. Une sémantique ensembliste est ensuite donnée pour associer un sens aux termes utilisés dans la description.

Les Concepts et les Rôles :

Un concept décrit un ensemble d'objets et correspond à la notion de classe dans les modèles centrés objets. Nous distinguons deux types de concepts : les concepts primitifs dont la structure des concepts est interprétée comme un ensemble de conditions nécessaires et les concepts définis dont la structure

du concept est interprétée comme étant définie, c'est-à-dire les conditions nécessaires et suffisantes qui doivent être satisfaites par les instances. Un rôle décrit une relation entre les ensembles d'objets. Il est utilisé pour spécifier les propriétés des objets. Il peut être perçu comme la notion d'attribut dans le modèle centré objet.

Les Descriptions

Les descriptions sont construites, à partir d'une composition de termes, en utilisant un sous-ensemble de constructeurs de termes. Ces termes représentent des objets tels que les concepts et les rôles.

Les Constructeurs de concepts

Ces logiques de descriptions offrent un sous-ensemble de constructeurs de concepts qui sont riches sémantiquement en donnant une grande capacité d'expression:

- **constructeur AND**: qui est une conjonction de concepts, permet d'exprimer l'héritage multiple dans le langage centré objet;
- **constructeur ALL** : qui définit un concept sur la base d'une restriction de valeur sur le co-domaine d'un rôle;

Exemple

" Module **AND ALL** (Apprenant , Etudiant)"

C'est une composition qui correspond à la description d'un module suivi par des étudiants.

- **constructeur AT-LEAST**: introduit le nombre minimal de restrictions d'un rôle dans la composition d'un concept;

Exemple

" Module **AND ALL**(Enseigné, Enseignant)
AND AT-LEAST(Enseigné,2)"

C'est une composition qui correspond à la description d'un module enseigné par au moins 2 enseignants.

- **constructeur AT-MOST**: introduit le nombre maximal de restrictions d'un rôle dans la composition d'un concept;

Exemple

" Module **AND ALL** (Apprenant, Etudiant) **AND AT-LEAST** (Apprenant,4)
AND AT-MOST (Apprenant , 30) "

C'est une composition qui correspond à la description d'un module suivi par au moins 4 étudiants et au plus 30 étudiants.

4. Applications

Ce langage, fondé sur les logiques terminologiques, est défini comme un environnement de manipulation de concepts dans le sens où le processus de classification se présente naturellement comme un outil d'aide à l'acquisition des connaissances en opérant sur un ensemble d'objets, en construisant et en gérant une hiérarchie d'objets. Le raisonnement déductif au travers d'une terminologie, fondé sur la relation de subsomption, permet de classer des concepts dans un graphe, de vérifier la cohérence des descriptions, d'éliminer certaines redondances en utilisant la notion d'héritage et de rendre plus explicites les contraintes d'intégrité sémantiques et/ou structurelles.

De nombreuses bases de données, qui sont implantées dans des systèmes différents, évoluent depuis de longues années. Mais devant l'évolution croissante des systèmes de gestion de bases de données et la prolifération rapide des réseaux d'interconnexion, l'on est amené à enrichir sémantiquement ces bases existantes pour les faire coopérer les unes avec les autres afin de faire partager les informations relevant d'un intérêt commun. Ceci est très crucial pour l'interopérabilité des bases de données hétérogènes notamment quand il s'agit d'intégrer ou de fédérer des bases de données autonomes ayant des caractéristiques hétéroclites.

Ce langage peut aussi faire office de méthode de représentation de la sémantique des situations de l'univers réel à modéliser pour enrichir sémantiquement les bases de données relationnelles [5], [7].

4.1 Enrichissement Sémantiques des Bases de Données

De nombreuses bases de données, qui sont implantées dans des systèmes différents, évoluent depuis de longues années. Mais devant l'évolution croissante des systèmes de gestion de bases de données et la prolifération rapide des réseaux d'interconnexion, l'on est amené à enrichir sémantiquement ces bases existantes pour les faire coopérer les unes avec les autres afin de faire partager les informations relevant d'un intérêt commun. Ceci est très crucial pour l'interopérabilité des bases de données hétérogènes notamment quand il s'agit d'intégrer ou de fédérer des bases de données autonomes ayant des caractéristiques hétéroclites.

Cet enrichissement sémantique consiste, en préservant, avant tout, les investissements antérieurs, à comprendre les aspects structurels et

sémantiques des bases existantes dans le but de découvrir de nouvelles connaissances permettant le passage à des modèles plus évolués. Ceci revient à identifier le contenu du dictionnaire (noms de relations, noms d'attributs, et leurs liens), à mettre en évidence la sémantique des domaines qui n'est pas explicitement présente dans le système étudié (clés étrangères, dépendances diverses, ...), à proposer des spécifications possibles correspondant à cette sémantique, à représenter le résultat dans un modèle canonique. Ceci nous permet de promouvoir le niveau sémantique des schémas existants en utilisant des modèles plus riches.

L'hétérogénéité à différents niveaux (sémantiques, structurels, descriptifs) implique la traduction des bases de données locales dans un modèle canonique plus riche dont le choix est fondamental car il joue un rôle important dans la comparaison des schémas locaux et dans la détection des conflits inter-schémas.

Le modèle canonique est défini dans le langage de concepts. Ce langage utilise la notion de classes pour produire des descriptions qui sont, en plus, utilisées dans le raisonnement. Les logiques terminologiques décrivent la structure des objets, à un niveau terminologique, en termes de concepts et de rôles. Une sémantique ensembliste est ensuite donnée pour associer un sens aux termes utilisés dans la description [5].

4.2 Modélisation des Données

Des formalismes tels que les modèles de données sémantiques, les bases de données orientées objet et les réseaux sémantiques permettent d'obtenir un modèle symbolique de l'univers réel à modéliser en se basant sur l'intuition. Celle-ci consiste à décrire une situation en s'intéressant aux divers types d'objets reliés par des relations et regroupés dans des classes.

Les langages terminologiques, en utilisant la notion de classes pour produire des descriptions utilisées dans le raisonnement, outils de représentation et de raisonnement en intelligence artificielle, sont d'autant plus adaptées à la représentation de la sémantique des situations de l'univers réel à modéliser. Les modèles de données sémantiques tels que Entité-Association, peuvent être exprimés en utilisant ces logiques terminologiques. Ces dernières permettent également d'exprimer des requêtes et définir des vues organisées dans un graphe d'héritage [7].

5. Conclusion

L'intérêt de cette approche consiste à tirer profit de l'important pouvoir descriptif ainsi que du raisonnement taxinomique offerts par les logiques de descriptions. Le raisonnement déductif au travers d'une terminologie, fondé sur la relation de subsumption, permet de classifier des concepts dans un graphe, de vérifier la cohérence des descriptions, d'éliminer certaines redondances en utilisant la notion d'héritage et de rendre plus explicites les contraintes d'intégrité sémantiques et/ou structurelles.

Ainsi, les logiques de descriptions, qui sont des outils de représentation et de raisonnement en intelligence artificielle, sont d'autant plus adaptées à la représentation de la sémantique des situations de l'univers réel à modéliser pour enrichir sémantiquement les bases de données relationnelles. Elles peuvent aussi être utilisées tout simplement dans le processus de rétro-conception des bases de données: ce qui est l'objet de recherche dans le cadre de la découverte des connaissances dans les bases de données.

Le langage de description *ALC* est un langage de description qui permet de représenter et de raisonner sur des concepts et des individus. Il est basé sur la logique de descriptions *ALC* qui est une logique de description qui permet de représenter et de raisonner sur des concepts et des individus. Il est basé sur la logique de descriptions *ALC* qui est une logique de description qui permet de représenter et de raisonner sur des concepts et des individus.

Le langage de description *ALC* est un langage de description qui permet de représenter et de raisonner sur des concepts et des individus. Il est basé sur la logique de descriptions *ALC* qui est une logique de description qui permet de représenter et de raisonner sur des concepts et des individus. Il est basé sur la logique de descriptions *ALC* qui est une logique de description qui permet de représenter et de raisonner sur des concepts et des individus.

Le langage de description *ALC* est un langage de description qui permet de représenter et de raisonner sur des concepts et des individus. Il est basé sur la logique de descriptions *ALC* qui est une logique de description qui permet de représenter et de raisonner sur des concepts et des individus. Il est basé sur la logique de descriptions *ALC* qui est une logique de description qui permet de représenter et de raisonner sur des concepts et des individus.

Bibliographie

- [1] A. Borgida, " Description Logics in Data Management ", IEEE Transactions on Knowledge and Data Engineering, vol.7, n.5, Octobre 1995, pp. 671-682.
- [2] Z. Bouzidi, A. Hocine, " RECOS un langage de représentation et d'exploitation de connaissances par les schémas ", Rapport interne ICOG 3/95, Dépt d'Informatique, Université de Pau, Août 1995.
- [3] Z. Bouzidi, A. Hocine, H. Kherbachi, "Vers un Environnement de Développement Centré-objet ", IV èmes Rencontres de Recherche Opérationnelle, 6-8 Octobre 1996, USTHB, Alger.
- [4] Z. Bouzidi, H. Kherbachi, "Vers un Environnement de Manipulation de Concepts fondé sur les Logiques de Descriptions ", Congrès National de Mathématiques, Congrès National de Mathématiques "CNM'99 ", 17-19 Mai 1999, Annaba / Algérie.
- [5] Z. Bouzidi, A. Hocine, H. Kherbachi, " Enrichissement Sémantique des Bases de Données : Une approche fondée sur les Logiques de Descriptions ", III èmes Rencontres Francophones de Recherche Opérationnelle (FRANCORO III), 9-12 Mai 2001, Québec (Canada).
- [6] E. F. Codd, " Extending the database relational model to capture more meaning ", ACM TODS volume 4, number 4, december 1979.
- [7] H. Kherbachi, A. Hocine, Z. Bouzidi, " Une approche de Modélisation de Données fondée sur les Logiques Terminologiques ", II èmes Journées Francophones de Recherche Opérationnelles FRANCORO II, 6-8 Avril 1998, Sousse (Tunisie).
- [8] A. Napoli, C. Laurenço, " Représentation à Objets et Classification : Conception d'un système d'aide à la planification de systèmes organiques ", Revue I.A., vol.7, N.2, 1993, pp. 175-221.
- [9] A. Saoudi, G. Nachouki, " Utilisation d'un Système Terminologique pour la Rétro Conception des Bases de Données Relationnelles ", INFORSID'96, 1996, pp. 199-214.