

**A Formal Ontology for Data Mining : Principles, Design, and Evolution
(Une ontologie formelle pour le forage de données : principes, conception
et évolution)**

Yanfen Shen
30004656

RÉSUMÉ

Le forage de données (DM) et les systèmes d'aide à la décision (DS) sont deux domaines relativement indépendants qui sont largement appliqués dans la recherche scientifique et la gestion. L'intégration réussie des technologies associées à ces deux domaines pourra mener à la réalisation d'un système intelligent puissant pour soutenir la prise de décision. L'application d'un système intelligent d'aide au forage de données, qui peut fournir une aide intelligente au delà de nombre de méthodes et d'outils de DM, est une étape essentielle vers une meilleure intégration de DM et de DS. Cependant, le développement du système doit relever actuellement deux défis principaux : le support de foreur de données non expert et la définition de la connaissance de DM. Les ontologies formelles et informatisées, en tant que nouveau secteur de recherche pour la conceptualisation de la connaissance, possèdent un grand potentiel pour aider à résoudre les problèmes ci-dessus. En raison de son puissant formalisme de représentation de la connaissance et du mécanisme d'entretien associé, intégrer une ontologie dans le système d'aide au forage de données sera une manière efficace de rendre le système plus intelligent et utile pour des décideurs.

L'objectif de cette recherche est de développer une approche basée sur l'utilisation d'une ontologie dans le domaine du forage de données. Cette dernière constitue une base de connaissance du domaine de forage de données ; ainsi qu'une méthode d'évolution d'ontologie, laquelle fournit un mécanisme et un outil pour soutenir le développement et la mise à jour de l'ontologie. Cette recherche est une partie fondamentale d'un plus grand projet de développement d'un système intelligent d'aide au forage de données.

Basé sur Protégé (Université de Stanford) et le langage OWL, nous avons établi avec succès une fine ontologie de DM. Le rôle de l'ontologie de DM est de représenter la connaissance de forage de données nécessaire au système. Deux types de connaissance sont représentés : la connaissance du domaine de forage de données et la connaissance produite par le système. La connaissance du domaine de forage de données se compose de la méthodologie et de la connaissance détaillée applicable au processus de forage de données, alors que la connaissance produite par le système se compose de la représentation d'annotations de données et des cas du système à base de cas. Pour fournir un support plus intelligent pour des activités de forage de données, notre ontologie de DM reste intégrée avec les deux autres composants du système : l'entrepôt de données et le système de raisonnement basé sur les cas.

Nous avons aussi proposé une nouvelle méthodologie d'évolution d'ontologie, cette méthodologie est mise en application comme un plug-in de Protégé. Cette approche est basée sur les tâches de l'évolution de l'ontologie et la conséquence des opérations de changement. Cette approche définit avec grande précision les différentes opérations de changement et les tâches de l'évolution. Le plug-in groupe et arrange les étapes nécessaires des tâches de l'évolution les plus généralement utilisées. Il peut être employé comme un *wizard*, étape par étape, pour guider des décideurs pour exécuter les tâches de mise à jour de l'ontologie.

Les travaux accomplis dans cette recherche ont permis de construire la structure fondamentale pour notre système d'aide au forage de données et ont préparé le terrain pour une meilleure intégration du forage de données et des systèmes d'aide à la décision. En outre notre méthodologie d'évolution d'ontologie améliorera considérablement l'exactitude, l'uniformité et l'efficacité des tâches de l'évolution. À notre avis, cette approche évolutive possède un grand potentiel pour un développement ultérieur.