

Traitement et Fusion de Données dans le Cadre de l'Interopérabilité Sémantique des Systèmes d'Information Géographiques.

Atmane HADJI, Mohamed-Khireddine KHOLLADI
Dept. Mathématiques et Informatique Université d'Ouargla Ouargla, Algerie
Chef Dept. Informatique Université Mentouri Constantine Constantine, Algérie
h.atm1@yahoo.fr
Kholladi@yahoo.fr

Résumé

De nos jours, L'évolution exponentielle de la technologie dans le Web sémantique est orientée vers plusieurs domaines de recherches. Ceci signifie que la technologie comprenant celle des systèmes d'information géographiques SIG se déplace vers les solutions flexibles par les systèmes à base de connaissances (sémantiques) que la solution statique par les systèmes de base de données courants (schématique ou syntaxiques). L'interopérabilité est nécessaire dans plusieurs applications du SIG. Elle implique de partager et de réutiliser des données de divers systèmes d'information hétérogènes. Les Ontologies émergent comme une solution importante pour construire un ensemble de connaissances partagées et réutilisables qui supportent leur interaction.

Dans ce travail, nous avons essayé de décrire un cadre général d'un nouveau système de médiation capable de gérer l'interopérabilité sémantique entre différents systèmes d'information (SI) particulièrement les SIG basé sur les ontologies. Nous avons appliquées certaines techniques dans notre approche pouvant assurer la coopération entre les ontologies (données) pour réaliser l'interopérabilité sémantiques, tels que la fusion et le mapping, elles sont utilisées pour permettre le partage des données entre des bases de connaissance hétérogènes et aussi bien la réutilisation des informations de ces bases et de minimiser la perte d'informations afin d'assure une coopération d'un système d'information interopérable.

1. Introduction

Actuellement, il existe de nombreuses bases de données géographiques (BDG) qui couvrent la même région du monde réel à des données ou représentations géométriques différentes. La multiplicité et l'indépendance des bases de données géographiques posent aujourd'hui un certain nombre de problèmes. Le fait qu'il n'existe aucun lien entre les objets géographiques qu'il implique : une répétition des opérations de mises à jour, un manque de cohérence entre les différentes BD, et une impossibilité d'effectuer des

analyses à plusieurs niveaux de détail. Dans ce cas, l'interopérabilité entre ces systèmes d'information est compliquée à cause de l'hétérogénéité de différentes sources de données et de contexte. Alors le problème posé est : (1) Comment fournir un contexte globale de données hétérogènes et dans un espace distribués. Ou bien (2) Comment créer une coopération entre des ressources échangées à travers différents systèmes pour faciliter l'interopérabilité. (3) Comment faciliter et simplifier l'échange entre les données géographiques?

2. Vers la sémantique comme une solution d'interopérabilité

Il existe deux méthodes principales pour faciliter l'interopérabilité et l'échange d'information. La première, on a la solution de la normalisation globale des données, via des standards et des normes centralisés. D'autre part, Comme la sémantique est un aspect important dans la coopération des systèmes d'information, cette sémantique permet aux systèmes distribués de combiner les informations reçues avec des informations locales ou hétérogènes et de traiter l'ensemble de manière globale et fournir une modélisation de données indépendamment de leur degré de structuration.

Comme la sémantique est un aspect important dans la coopération des systèmes d'information, notre solution est basée sur l'ajout d'une couche sémantique à l'ensemble des objets géographiques indépendamment de leur degré de structuration. Cette couche permet de décrire le contexte de données dans un langage compréhensible par les utilisateurs (pour prend la décision) et les fournisseurs (mettre à jour les données géographiques) et aussi par la machine (pour automatiser les tâches). Cette couche est basée sur les ontologies. L'objectif est de concevoir des méthodes de traitement de données basées sur l'approche sémantique et sur une architecture de médiation qui permettre la coopération ou la fusion de l'information au profit des systèmes hétérogènes et distribués. De plus, les ontologies permettent des descriptions formelles et déclaratives des termes communs, prend en compte du

raisonnement automatique ou semi-automatique sur des données partagées d'un domaine.

2.1. L'objectif d'un médiateur de contexte dans notre système

Le médiateur de contextes est adapté à des environnements où les sources d'information sont probablement d'évoluer, d'apparaître et de disparaître, ces caractéristiques sont adaptées à l'évolution d'un environnement géographique. La médiation de contexte est également caractérisée par une prise en charge de la sémantique grâce aux mécanismes d'unification ou de réconciliation des contextes. Elle peut être considérée comme une approche orientée sémantique. Alors l'interopérabilité concerne la préservation du sens de l'information échangée, c'est ce qu'on appelle l'interopérabilité sémantique.

Une définition est couramment admise pour l'interopérabilité sémantique : "elle donne un sens aux informations échangées et s'assure que ce sens est commun dans tous les systèmes entre lesquels des échanges doivent être mis en œuvre" [1] [2] [3]. La prise en compte de cette sémantique permet aux systèmes distribués de combiner les informations reçues avec des informations locales et de traiter l'ensemble de manière cohérente.

3. Architecture conceptuelle de notre système proposé

Notre système d'interopération est divisé en trois niveaux (Figure 1.) :

Niveau d'application : représente les requêtes émises par les utilisateurs et aussi de différentes applications (couche consommateur).

Niveau médiation : c'est la couche intermédiaire entre les utilisateurs et les participants à la coopération des systèmes d'informations géographiques (couche intermédiaire).

Niveau SIG : L'approche est de type médiation de contexte alors chaque producteur du SIG participant à la coopération qui possède un contexte qui définit la sémantique des données partagées dans la coopération (couche fournisseur).

3.1. Définition de notre système proposé

Notre approche basée sur un système de médiation de contexte, permet d'adopter les sources d'informations sous forme structurées comme OWL (Web Ontology Language, informellement OWL) ou semi structurées que XML.

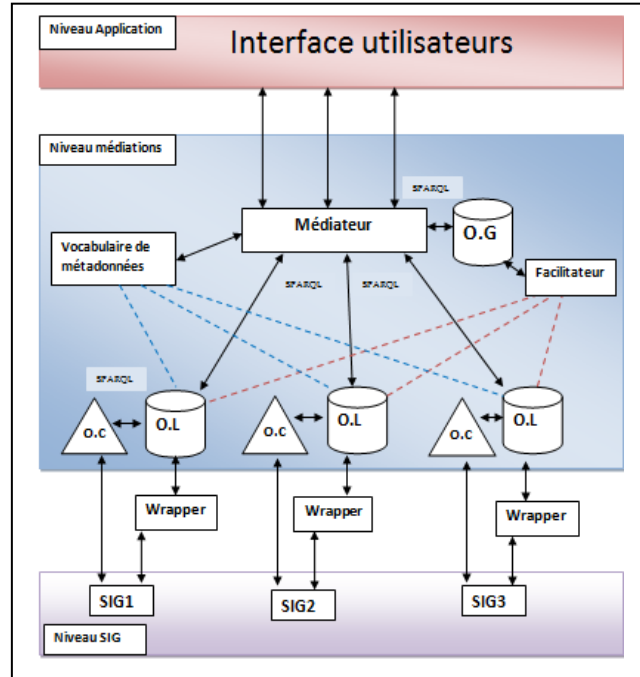


Figure 1. Architecture de médiation de concepts basée sur les ontologies.

Notre système est basé sur le modèle de données OWL car ce dernier permet d'établir des relations sémantiques complexes entre les différents contextes, et ainsi apportent les capacités de raisonnement nécessaires pour résoudre les hétérogénéités entre les différents contextes des fournisseurs, et aussi bien que notre système de médiation de contextes est basé sur les ontologies. Chaque source d'information locale est décrite au niveau d'un contexte à l'aide d'une ontologie locale qui permet d'avoir une vue sur la source. Dans le cas où les sources sont décrites à l'aide de schéma XML, le format OWL [6] sera généré semi automatiquement pour construire l'ontologie. À l'objectif de représenter un domaine géographique commun et compatible entre les systèmes (SIG) participants à la coopération, nous proposons une ontologie de domaine globale qui prend les connaissances et inclut un vocabulaire de concepts avec une spécification précise et formelle de leur signification à partir des ontologies locales. Une ontologie locale permet à un utilisateur (consommateur) d'exposer ses besoins par l'envoi des requêtes au médiateur pour reprendre les informations ou les données nécessaires à d'autres systèmes de la coopération. Un SIG participant fournit des informations adaptées aux besoins des autres systèmes et les met à jour à base d'une ontologie de contexte (application).

3.2. Niveau de médiation

Nous avons défini un cadre sémantique de médiation. Il se compose d'une description de contenu et de contextes à base d'ontologie, composants de médiation de l'information et un processeur de traitement des requêtes. Les composants de médiation sont employés pour chercher des réponses appropriées de requêtes et pour réconcilier des différences sémantiques parmi les sources de données hétérogènes. Les descriptions de contexte et d'ontologie fournissent la sémantique commune pour les composants de médiation. Le processeur QUERY reformule et soumet les requêtes sur les gisements de données locaux et combine les résultats des sous-requêtes soumise.

3.3. Types d'ontologies utilisées dans notre architecture

Ontologie de Domaine : dans notre architecture nous avons utilisé une ontologie globale pour décrire le domaine d'application global de la coopération et permet au médiateur, le traitement sémantique d'une requête. Elle décrit le vocabulaire ayant trait à un domaine générique (exemple : géographique, Spatial). elle garantit : (1) La capitalisation des connaissances. (2) Le partage sémantique et la communication entre les différents SIG participants à la coopération.

Ontologie Locale (générique) : est une ontologie locale dans notre approche exprime le contexte ou le domaine d'application local d'un SIG. Appelée, méta-ontologies, elle véhicule des connaissances génériques moins abstraites que celles véhiculées par l'ontologie de domaine, mais assez générales néanmoins pour être réutilisées à travers différents domaines. L'ontologie générique que nous construisons permet la mise en œuvre d'une terminologie propre au domaine concerné par l'interopérabilité afin de pouvoir éviter les conflits sémantiques.

Ontologie de contexte (ontologie d'application) : Une ontologie d'application offre le plus fin niveau de spécificité, c'est-à-dire elle est dédiée à un champ d'application précis à l'intérieur d'un domaine (géographique). Cette application a pour le but de décrire le contexte d'un concept de l'ontologie locale. Ainsi, à chaque concept d'une ontologie locale, on associe une ontologie de contexte. La mise en place d'une ontologie contextuelle est facile. Il s'agit d'une ontologie simple, qui modélise les différentes propriétés sémantiques d'un concept de l'ontologie locale, ainsi que leurs relations. (Voir figure 2.).

Cette séparation entre l'ontologie locale et l'ontologie de contexte permet de résoudre les conflits liés aux contextes des utilisateurs lors de ses requêtes et d'autre

part les conflits de la mise à jour de contexte par les fournisseurs ou les SIG participants à la coopération par l'ontologie de contexte.

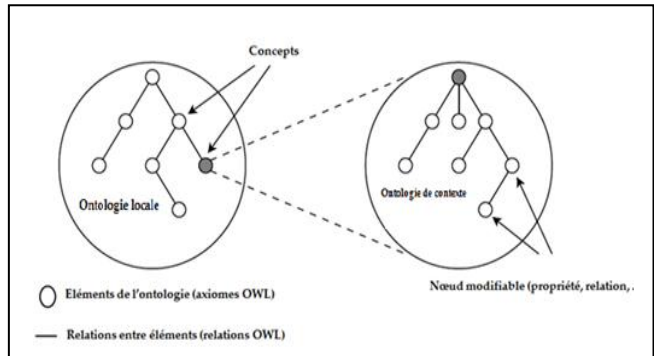


Figure 2. L'ontologie Locale et l'ontologie de contexte.

4. La correspondance entre les ontologies

Nous avons appliquées certaines techniques sur les ontologies dans notre approche pouvant assurer la coopération entre les ontologies pour réaliser l'interopérabilité sémantiques, tels que la fusion et le mapping d'ontologies (figure ci-dessous). Elles sont utilisées pour permettre le partage des données entre des bases de connaissance hétérogènes et pour la réutilisation des informations de ces bases et afin d'assurer la perte d'informations.

4.1. Le mapping entre les ontologies de contextes

La définition la plus pertinente est probablement celle de Noy pour le mapping d'ontologies est un processus qui spécifie une convergence sémantique entre différentes ontologies afin d'extraire les correspondances entre certaines entités [4]. Dans notre approche le mapping effectué entre les ontologies de contextes et de produit comme résultat une ontologie locale.

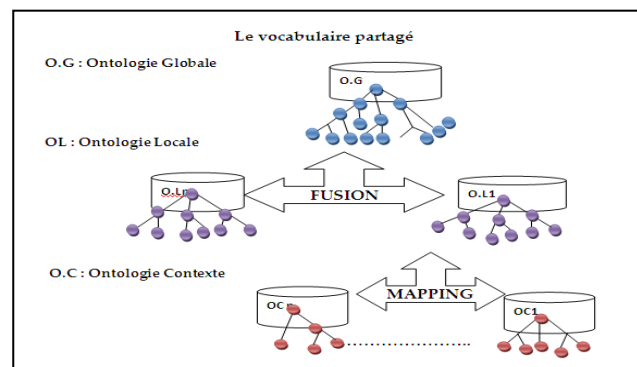


Figure 3. Techniques d'enrichissement entre les ontologies.

4.2. La fusion d'ontologies locales

La fusion d'ontologies permet de créer une nouvelle ontologie, appelée l'ontologie fusionnée (globale) prenant les connaissances des ontologies sources [5]. La fusion d'ontologies dans notre proposition concerne la création d'une nouvelle ontologie (ontologie globale) à partir de deux ontologies sources (ontologies locales). L'ontologie résultante unifie et remplace les ontologies d'origine (Voir Figure 3.). Les approches les plus courantes utilisent l'union ou l'intersection. Dans l'approche par union, l'ontologie résultante contient l'union des entités provenant des ontologies originales et résout les différences de représentation d'un même concept.

4.3. Processus de traitement des requêtes

La figure suivante (Figure 4.) représente le processus de traitement des requêtes des utilisateurs. Ce processus est constitué de deux phases :

Phase 1 le médiateur envoie la requête du demandeur vers les ontologies locales identifiées et de bénéficier le temps de recherche qu'il exploite les résultats d'un processus de mapping que ce dernier réalise la correspondance entre les différents sources de données et de résoudre le problème de la complexité des structures de données (ontologies de contextes) alors d'augmenter le temps de réponse de la requête demandeur.

Phase 2 : dans le cas où la réponse dans la phase précédente non octroyée alors le médiateur revoie la requête vers l'ontologie globale que cette dernière garantie la réponse demandeur par ce que cette dernière est résultante par les deux processus de mapping et de fusion d'ontologies qu'il implique de fournir une correspondance entre les différents sources de données et assurer le partage de données afin de créer un vocabulaire commun alors d'annuler le problème de perte d'informations.

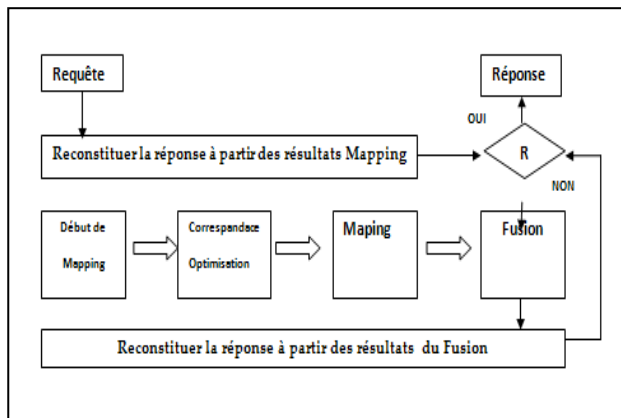


Figure 4. Processus de traitement des requêtes.

5. L'implémentation

L'explicitation des ontologies s'effectue au moyen de langages est une des principales décisions à prendre dans le procédé de développement d'ontologies consiste à choisir le langage dans lequel l'ontologie sera exprimée et utilisée. Comme l'éventail des choix possibles concernant les langages de représentation et de spécification d'ontologies étant très large, il convient de choisir celui qui respecte ces exigences et critères techniques: la lisibilité, la possibilité de faire des inférences c'est-à-dire permettre le traitement informatique des données en vue de calculer les déductions logiques possibles [7].

5.1. L'éditeur d'ontologie Protégé¹ et Prompt

Pour la spécification de l'ontologie, le langage OWL est le plus adéquat, pour l'implémentation, nous utilisons Protégé comme un éditeur d'ontologie qui génère un fichier en sortie sous le format OWL basé sur la syntaxe Xml compréhensible et traitable par la machine. Prompt est un outil de type interactif qui effectue les processus d'enrichissement d'ontologies (la fusion, le mapping, l'extraction, et aussi la comparaison entre les ontologies).

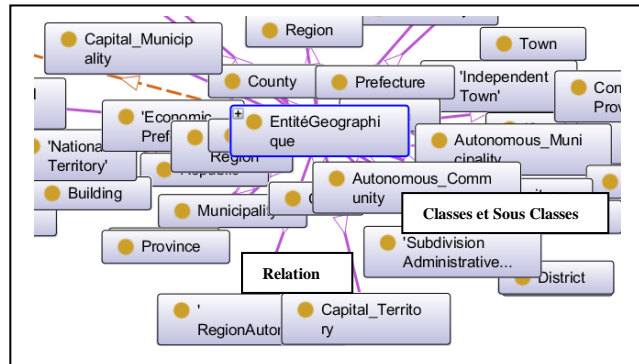


Figure 5. La visualisation d'une ontologie dans Protégé.

5.2 Processus d'enrichissement Mapping et Fusion

Le but de notre proposition basée sur une architecture de médiation sémantique à base des ontologies avec l'utilisation des techniques d'enrichissement entre cette dernière pour l'interopérabilité des systèmes d'information géographiques est de fournir aux utilisateurs des architectures d'intégration virtuelles et des outils interopérables.

¹ <http://protege.stanford.edu/download.html>

5.2.1. Le mapping d'ontologies :

Le mapping d'ontologies est un processus qui spécifie une convergence ou correspondance sémantique entre différentes ontologies (Figure 6.). Ces correspondances sont exprimées en introduisant des axiomes formulés dans un langage spécifique. Trois phases principales peuvent être distinguées dans ce processus : la découverte du mapping ; la représentation et l'exécution du Mapping ; l'exploitation du mapping.

La découverte du mapping : il faut identifier les ontologies (ontologies de contexte) de mapping , dans cette phase nous détenons de choisir les ontologies correspondantes pour le mappage selon notre besoins .

La représentation et l'exécution du mapping : dans cette phase il faut identifier le niveau de mapping (Classe, Sous classe, instance) et aussi de choisir le type de mapping (métadonnée, composants ou étape par étape) correspondantes pour le mappage selon notre besoins afin d'adapter les conditions de mapping et que le système effectue l'action demandée et exécute automatiquement les changements additionnels dérivés de cette opération.

L'exploitation du mapping : le composant d'analyse est une tâche importante appliqué dans cette phase. En conclusion, dans la phase de système, le composant logique est encore l'ordre pour vérifier l'uniformité et les inférences dans l'ontologie locale récemment créé dans la phase précédente.

5.2.2. L'Objectif de processus du mapping

Le mapping est une solution pour un grand volume et des conflits de données résultant de la diversité et de la complexité des structures de données spatiales afin de fournir une correspondance entre les différents sources de données alors de garantir le problème de l'hétérogénéité et le conflit entre les sources de données.

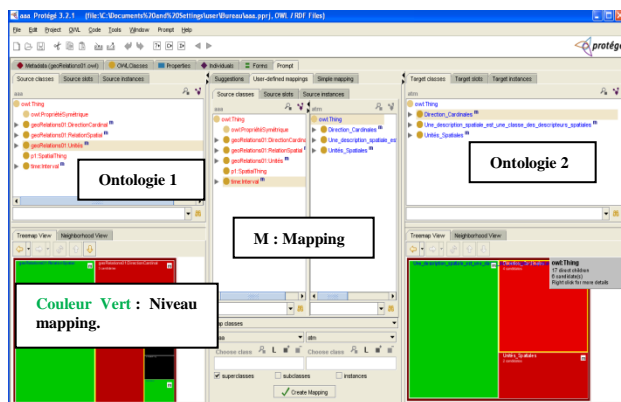


Figure 6. Le Processus de Mapping par Prompt.

5.2.3 Le Processus de Fusion

Le processus de fusionnement implique la tâche de fusionner les ontologies locales enrichies afin de créer un vocabulaire partagé ou une ontologie globale en définissant trois phases principales : La phase de départ (les deux ontologies locales fusionnées), la phase d'intégration, la phase finale (l'ontologie globale partagée).

La phase de départ (les deux ontologies locales à fusionner) : dans la phase de départ chaque ontologie spatiale (ontologie locale qui se résultante du processus de mapping entre les ontologies de contextes) ouverte séparément dans le processus sémantique d'enrichissement (fusionnement). La figure suivante représente les deux ontologies locales qui sont utilisées dans le processus du fusionnement. La fusion entre les classes est une tâche semi automatique il faut sélectionner les deux classes (sous classes ou les instances) et en suite on exécute le processus du fusionnement.

La phase d'intégration : Dans la deuxième phase d'intégration le processus est responsable de fusionner les deux ontologies locales normalisés afin de créer l'ontologie globale partagée est l'étape finale dans ce processus.

La phase finale : La figure ci-dessus exposer l'ontologie globale résultante par le processus du fusionnement à partir de deux ontologie locales et la suggestion sur le type du fusionnement.

5.2.4 L'objectif du processus du fusionnement

L'ensemble des collections existantes dans une coopération d'un SIG partage les données spatiales créés par différents établissements et organismes. Cependant, la fusion d'ontologies assure le partage de données spatiales afin de créer un vocabulaire commun partagé (l'ontologie globale). Alors de distribuer les coûts de gestion des données. C'est-à-dire que les diverses responsabilités et les coûts d'entretien sont partagés entre les différents acteurs qui fournissent des données.

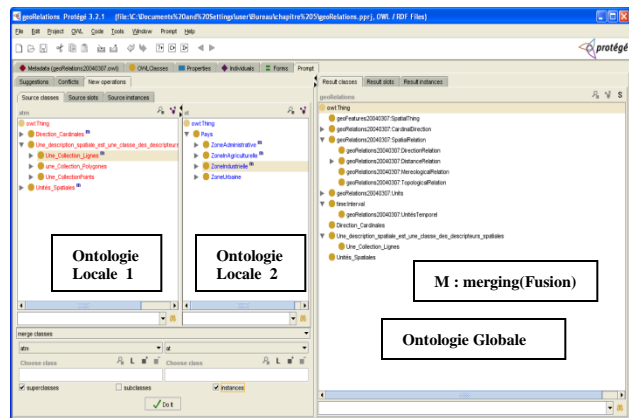


Figure 6. Le Processus de Fusionnement par Prompt..

6. SPARQL

SPARQL (acronyme dérivé de SQL) est un langage de requêtes destiné à interroger les bases de données (fichiers) RDF, standardisé par le W3C et implémenté sur les systèmes de base de données fournissent la récupération efficace des données par son langage d'interrogation sous la forme de langage d'interrogation structuré (SQL), l'ensemble de données dans un document RDF peut être interrogé par le langage d'interrogation appelé le SPARQL. C'est une composante clé de la technologie sémantique de Web. Comme langage d'interrogation, SPARQL « orienté-donnée » parce qu'il interroge seulement l'information tenue dans les modèles ; il n'y a aucune inférence dans le langage d'interrogation lui-même.

6.1. Les requêtes SPARQL

Le mouvement vers les Web sémantiques conduit la capacité de besoin des requêtes efficaces pour un grand ensemble de données comportent des relations supportées des ressources de Web. OWL est un ensemble des normes pour décrire et modéliser des données et est l'épine dorsale des technologies sémantiques de Web. Les ensembles de données RDF peuvent être très grands, et souvent sont sujets à des requêtes complexes avec l'intention de l'extraction et impliquent les raccordements autrement invisibles dans les données. La figure suivante (Figure 8.) représente un exemple d'exécution d'une requête SPARQL sur une ontologie géographique.

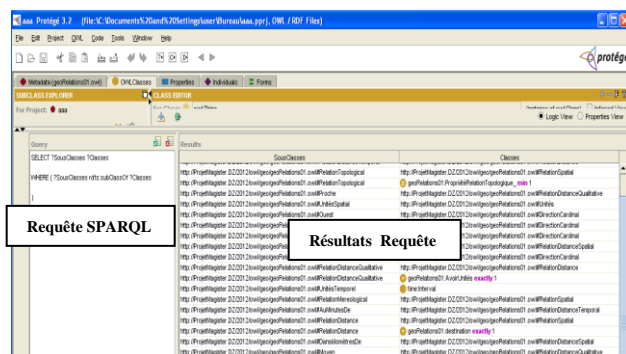


Figure 7. L'Exécution d'une requête SPARQL

7. Conclusion et perspectives

Dans ce travail, nous avons présenté notre approche fondée sur des techniques qui sont offertes pour réaliser l'interopérabilité sémantique. Les principales étapes de notre proposition sont : (1) la définition d'une architecture de médiation sémantique basée sur les ontologies. (2) la définition et la création de l'ontologie du contexte ou

d'application (3) la génération de l'ontologie locale par la technique de fusion d'ontologies (les ontologies d'applications contexte). (4) la génération de l'ontologie globale à partir des ontologies locales par la technique de mapping.(5) l'interrogation du modèle par un langage de requête SPARQL.

On distingue deux concepts principaux qui sont : (a) le mapping d'ontologies, pour l'objectif de la représentation et la correspondance entre les ontologies. (b) La fusion d'ontologies, leur objectif consiste à percevoir des fusions entre les ontologies afin d'assurer le partage de données spatiales et de créer un vocabulaire commun partagé (l'ontologie globale).

Dans un futur travail, nous comptons de réaliser une implémentation basé sur les techniques d'enrichissement des ontologies appartenant d'automatiser les différents catégories des opérations tel que la fusion, l'appariement, le mapping ...ect afin d'intégrer un système de raisonnement dans l'ontologie, comme Racer,Fact++, et HermiT,basé sur la logique des prédicats .

Nous tenterons aussi d'essayer d'autre utilisation de l'ontologie dans l'interopérabilité des systèmes comme les systèmes multi agents (SMA) dans lesquels chaque agent contient son ontologie pour augmenter l'interopérabilité entre la coopération des agents.

8. References

- [1] F. Vernadat., " Interoperable enterprise systems : Architectures, methods and metrics" ;LGIPM ;Université de Metz , France, 2007.
- [2] F. Jouanot , " Un modèle sémantique pour l'interopérabilité de systèmes d'information" ,In INFORSID, 2000.
- [3] Charlet ., Cordonnier., GIBaud ; "Interopérabilité en médecine : quand le contenu interroge le contenant et l'organisation", Information - Interaction - Intelligence, 2002.
- [4] N. Noy., "Semantic integration a survey of ontology-based approaches ", SIGMODRec,33(4):pp.65–70, 2004.
- [5] C. Namyoun , S. IL-YEOL ., H. HYOIL., " A survey on ontology mapping ", SIGMOD Rec,35(3):34–41, September 2006.
- [6] B. Audrey , "Construire une ontologie de la pneumologie : Aspects théoriques,Modèles et Expérimentations ", laboratoire INSERM UMR_S 872 – Santé Publique et Informatique Médicale, Thèse de doctorat de l'université PARIS 6, février 2007.
- [7] W3C., OWL Working Group,OWL 2 Web Ontology Language Document Overview; <http://www.w3.org/TR/2009/REC-owl2-overview-20091027>,W3C Recommendation , October 2009.