

Interopérabilité des ressources sémantiques pour la conception coopérative

P. Ghodous, L. Médini

Séminaire Ontologies Multiples – 9 septembre 2004

Laboratoire d'InfoRmatique en Image et Systèmes d'information

LIRIS FRE 2672 CNRS/INSA de Lyon/Université Claude Bernard Lyon 1/Université Lumière Lyon 2/Ecole Centrale de Lyon
Université Claude Bernard Lyon 1, bâtiment Nautibus
43, boulevard du 11 novembre 1918 — F-69622 Villeurbanne cedex
<http://liris.cnrs.fr/>



Plan de la présentation

- **Présentation de l'équipe**
 - Modélisation générique
 - Modélisation collaborative
 - Modélisation sémantique
- **Aspects « multi-ontologies »**
- **Application**
- **Perspectives**

Aspects « multi-ontologies »

☰ Présentation de l'équipe

☰ Aspects « multi-ontologies »

- Problématique de l'interopérabilité des ressources hétérogènes
- Approches existantes
- Démarche adoptée

☰ Application

☰ Conclusion et perspectives

☰ Hétérogénéité des Ressources Sémantiques (RS)

● Au niveau des modèles utilisés

● Différents types de ressources

- *Taxonomies, classifications*
- « *Linguistiques* » : *vocabulaires, thésaurus, dictionnaires*
- *Ontologies*

● Différents formats de représentation

- *Texte, MS-Word, (X)HTML...*
- *XML, RDF, DAML+OIL, OWL*
- *KIF, KL-One, UML+OCL*

● Différentes « utilisabilités »

- *Compréhensibilité*
- *Interprétabilité (computabilité, calculabilité, ...)*

☰ Hétérogénéité des Ressources Sémantiques

● Au niveau des données représentées

● Différences de contenu

Données utilisées, concepts définis (heureusement !)

● Différences de « points de vue »

○ *Cohérence des données*

○ *Structuration des connaissances*

● Différences de portée, véracité

Exemples : données physiques, terminologies techniques, réglementations locales, modèles d'activité

Problématique

Ce que l'on recherche

● Identités

Données identiques sans traitement

● Similarités

Données identiques après traitement(s)

● Interdépendances

Données reliées par des relations connues

● Correspondances

Relations permettant de mettre en évidence les similarités et les interdépendances

○ *Règles de mapping*

○ *Contraintes*

☰ Objectif à atteindre : l'interopérabilité

● Des données

- Identifier une valeur à partir de données identiques ou similaires extraites d'autres points de vue
- Déterminer une telle valeur à partir de correspondances

● Des sources d'information

Identifier les documents pertinents à travers les points de vue

● Des connaissances

Traiter les connaissances existantes pour générer des règles de correspondances non pré-établies et cohérentes

Problématique

☰ Contraintes (liées au domaine de l'ingénierie coopérative)

- Conserver les modèles et outils appropriés pour les tâches des co-opérateurs
- Permettre leur évolution
- ⇒ Ne pas modifier les RS existantes (C1)

- Identifier les interactions pertinentes entre les points de vue pour éliminer les informations inutiles
- Gérer l'ensemble des relations entre les données pour ne pas limiter les traitements
- ⇒ Disposer d'un référentiel de conception pour positionner ces modèles (C2)

Approches existantes

☰ Approche intégrée ou « médiation centralisée »

● Fusion d'ontologies

- Création d'une ontologie à partir d'ontologies existantes
 - Les ontologies d'origine cessent d'être utilisées
- ⇒ Incompatible avec C1

● Intégration d'ontologies

- Construction ou enrichissement d'une ontologie à partir d'ontologies existantes
 - Les ontologies d'origine peuvent continuer à être utilisées.
- ⇒ Lourde à réaliser avec des sources nombreuses et très hétérogènes
- ⇒ Nécessite une re-génération de l'ontologie de haut niveau à chaque modification d'une composante

Approches existantes

☰ Approche fédérée ou « ontologies multiples »

● Alignement

- Établissement de correspondances binaires entre concepts de deux ontologies

● Appariement (« matching »)

- Établissement de correspondances n-aires entre concepts de plusieurs ontologies

● Doivent aboutir à un ensemble « pertinent et cohérent » [Klein, 2001]

⇒ « Légères modifications » des ontologies acceptées pour conserver la cohérence

⇒ Stockage des correspondances dans une « ontologie d'articulation » si les RS ne doivent pas être modifiées

Approches existantes

Approche hybride ou unifiée

- Semblable à l'approche fédérée
- Utilise un vocabulaire partagé comme ontologie de plus haut niveau
- ⇒ Facilite l'établissement de correspondances
- ⇒ Facilite les traitements ultérieurs
- ⇒ Nécessite la mise au point d'un vocabulaire partagé

Remarque : ces trois approches sont issues du standard ISO 14258, présenté dans [Chen & Doumeingts 2003]

Démarche adoptée

☰ Approche fédérée « contextualisée »

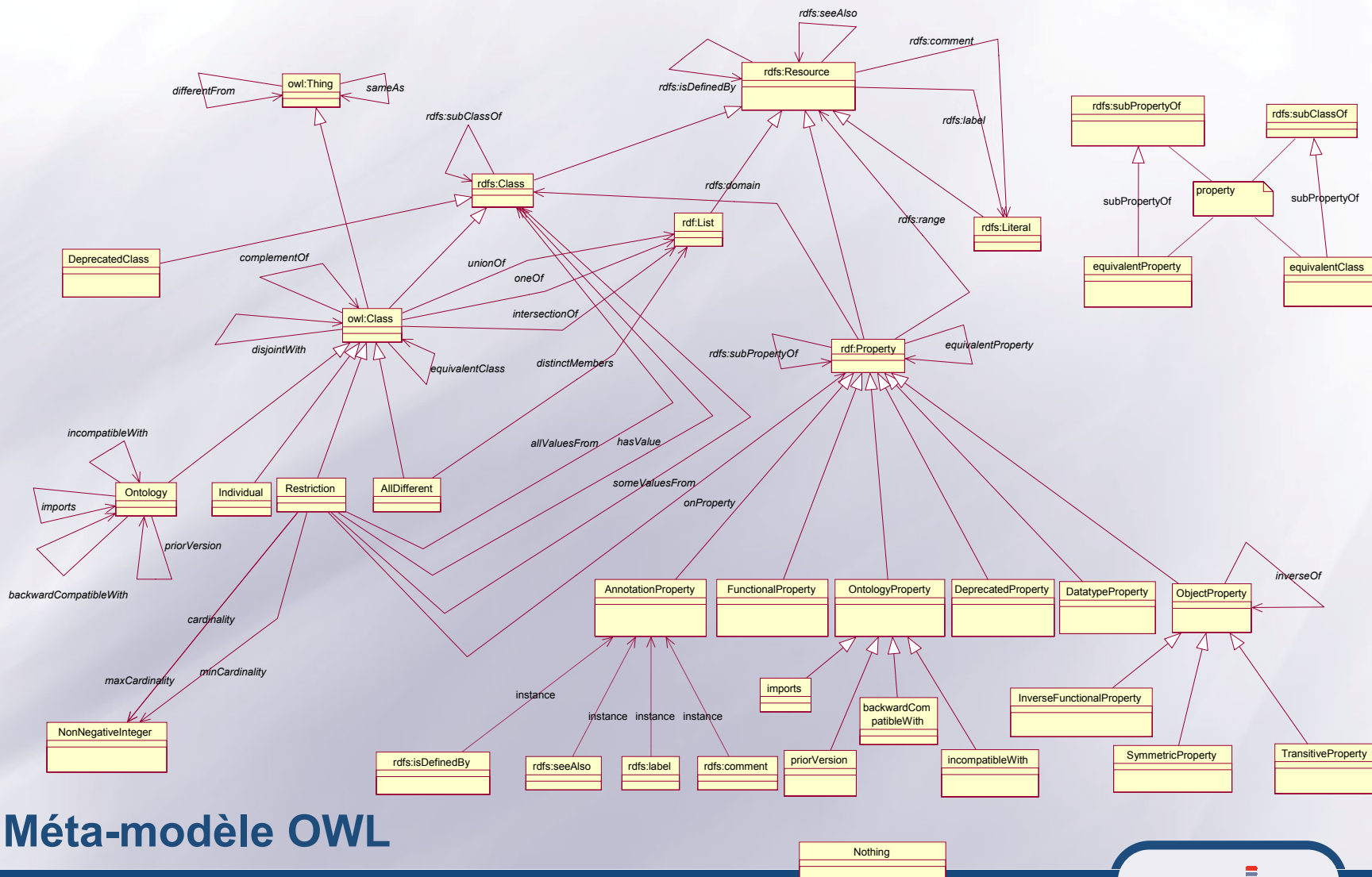
- Conserver les RS d'origine et les « traduire » dans un format pivot
- Identifier les correspondances
- Valuer ces correspondances par un calcul de distance dans l'espace de conception

Démarche adoptée

☰ Homogénéisation des formats de représentation

- Typologie des formats de représentation
 - Taxonomies
 - Graphes
 - Ontologies
- Élaboration de méta-modèles UML
- Sélection d'un format pivot : OWL DL
- Conversion de formats générique vers ce format pivot
 - Algorithme de Binding/Unmarshalling pour créer des classes Java à partir de schémas XML
 - Utilisation de l'API Jena2 pour obtenir les instances des classes OWL

Démarche adoptée



Méta-modèle OWL

Démarche adoptée

☰ Élaboration de correspondances

- Entre ontologies exprimées dans le format pivot
- Création d'une ontologie d'articulation
- Alignement/appariement automatiques
 - Appariement de contenus textuels (nœud à nœud)
 - Adaptation de techniques d'appariement de textes (parties de graphes)
- Structuration des correspondances
 - Règles de mapping valuées
 - Contraintes

Démarche adoptée

☰ Calcul de distances sémantiques

- **Caractérisation de l'espace de conception fondée sur la notion de contextes d'activité**
 - Calcul d'un vecteur sémantique contextuel pour chaque domaine
 - « *Types d'activités* »
 - *Méta-données*
 - *Production documentaire*
- **Appariement « offline » des RS**
 - Indique une distance contextuelle entre deux RS
 - Calcul réalisé *a priori*
- **Appariement « online » spécifique à l'utilisateur**
 - Norme d'un vecteur comprenant les vecteurs sémantiques précédents et des informations spécifiques à l'utilisateur
 - Calculé dynamiquement

Application

- **Présentation de l'équipe**
- **Aspects « multi-ontologies »**
- **Application**
 - **Contexte opérationnel**
 - **Résultats obtenus**
- **Conclusion et perspectives**

Contexte opérationnel

- **Projet FUNSIEC, suite du projet européen e-Cognos :**
 - Étude de faisabilité portant sur l'élaboration ainsi que de la maintenance d'une infrastructure sémantique ouverte pour le secteur européen de construction
- **Domaine d'application : secteur européen du bâtiment et de la construction**
- **Porteur du projet : CSTB**
- **But : un « Serveur d'ontologies » (e-Coser) reçoit des requêtes sur des concepts et doit les apparier**

Contexte opérationnel

Les RS

● Ontologie e-Cognos

- Vocabulaire de termes,
- Super-ontologie des RS du domaine
- Langage d'origine DAML, mais utilisation de l'espace de noms RDFS

● Graphe bcXML

- Vocabulaires multilingues de construction dédiés au e-Business
- Propose un « méta-schéma » permettant de définir une taxonomie : XTD (eXtensible Taxonomy Definition)
- Format des données : XML

● Taxonomie Edibatec

- Catalogue électronique pour la construction
- Format des données : fichiers de données séparés ou XLS

Résultats obtenus

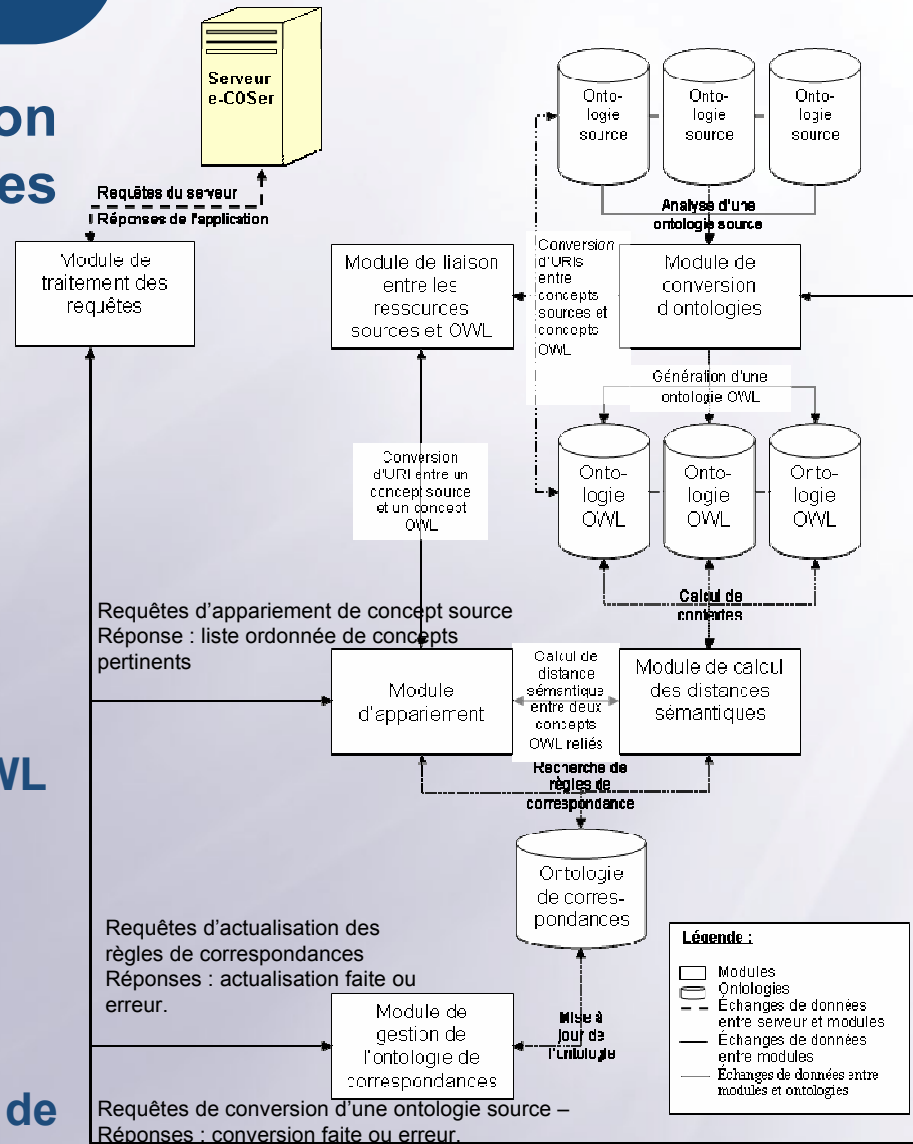
Architecture d'intégration syntaxique et sémantique des ressources hétérogènes

Trois types de requêtes

- Conversion
- Mise à jour
- Appariement

Éléments pris en charge

- Ressources
 - Ontologies converties en OWL
 - Ontologie d'articulation
- Modules
 - Conversion
 - Liaison
 - Gestion de l'ontologie de correspondance



Conclusion et perspectives

- Présentation de l'équipe
- Aspects « multi-ontologies »
- Application
- Conclusion et perspectives