

Intégration de données environnementales : une approche basée sur les entrepôts de documents XML et les ontologies

Ousmane SALL* ** Moussa LO*

* Laboratoire d'Analyse Numérique et d'Informatique – UFR SAT
Université Gaston Berger de Saint-Louis
BP 234 Saint-Louis (Sénégal)
lom@ugb.sn

** Laboratoire d'Informatique du Littoral
Université du Littoral Côte d'Opale
50, rue Ferdinand Buisson - BP 719
62228, Calais Cedex (France)
sall@lil.univ-littoral.fr

Résumé. Cet article présente l'approche que nous avons adoptée pour résoudre le problème d'intégration de données dans le contexte du projet SIC-Sénégal dont l'objectif est de permettre à plusieurs organismes partenaires de partager leurs sources de données environnementales. Nous réalisons une intégration en deux phases. Une première phase d'intégration structurelle, basée sur l'utilisation d'entrepôts de documents XML, permet de créer un entrepôt pour chaque organisme participant au projet. Une deuxième phase consiste alors à effectuer l'intégration de ces entrepôts de documents XML en associant une ontologie à chaque entrepôt. Cela se fait par une construction automatique d'ontologie OWL¹ à partir des données XML de l'entrepôt et par une réutilisation de l'ontologie AGROVOC².

1 Introduction

L'intégration est guidée par le besoin de regrouper des données ou documents présentant une hétérogénéité à deux niveaux : structurelle et sémantique. Du point de vue structurel, les documents pouvant exister sous plusieurs formes et les données organisées suivant des structures différentes. Du point de vue sémantique, les données portent des sens et significations différentes selon leurs identificateurs.

Nos travaux se situent dans le contexte du projet SIC³-Sénégal initié à l'Université Gaston Berger pour offrir des solutions aux problèmes rencontrés dans la gestion et l'exploitation des données liées à la mise en valeur de la vallée du fleuve Sénégal (BDSIC, 2004). Les données sont distribuées sur plusieurs sources hétérogènes et appartiennent souvent à des organismes différents appelés *partenaires* (dans le cadre du projet). Pour résoudre le problème d'hétérogénéité structurelle, nous proposons une approche dataweb (Lô, 2002). Cela consiste à réaliser, pour chaque partenaire, un entrepôt de documents XML intégrant les sources de données dudit partenaire.

¹ <http://www.w3.org/TR/owl-features/>

² <http://www.fao.org/agrovoc/>

³ Système d'Information et de Connaissances