

L'outil SDET pour le complètement des données descriptives liées aux bases de données géographiques

Khaoula Mahmoudi*
Sami Faïz ** ***

* Laboratoire URISA -Unité de Recherche en Imagerie Satellitaire et ses Applications
Ecole Supérieur des communications de Tunis (SUPCOM)

khaoula.mahmoudi@insat.rnu.tn

** Laboratoire de Télédétection et Systèmes d'Informations à Références Spatiales
(LTSIRS)

*** Institut National des Sciences Appliquées et de Technologie (INSAT)
sami.faiz@insat.rnu.tn

Résumé. L'enrichissement des bases de données est un moyen visant à offrir un supplément informationnel aux utilisateurs. Dans le cas des données géographiques, cette activité représente de nos jours un problème crucial. Sa résolution permettrait de meilleures prises de décisions ne reposant pas uniquement sur les informations limitées. Notre outil SDET (Semantic Data Enrichment Tool) vient proposer une solution d'enrichissement faisant du Système d'Information Géographiques (SIG) initial une source riche d'informations.

1 Aperçu du processus d'enrichissement

Le processus d'enrichissement que nous avons proposé (Faïz et Mahmoudi, 2005, Mahmoudi et Faïz, 2006_b) émane d'un besoin informationnel réclamé par les utilisateurs des SIG. Pour extraire les connaissances incarnées dans les documents dans des temps raisonnables, nous procédons d'une manière distribuée en adoptant le paradigme multi-agents (Ferber, 1997).

L'approche que nous proposons est modulaire, elle peut être décomposée en trois grandes phases. Il s'agit de la segmentation et de l'identification des thèmes abordés dans les documents initiaux. Suite à cette phase, un nouveau document est généré pour chaque thème regroupant les segments de textes distribués entre les différents agents et traitant le même thème. La seconde phase consiste à affecter pour chaque thème un délégué responsable de l'extraction de l'essentiel d'information de son document généré. Enfin, un filtrage textuel s'opère, il consiste à éliminer toute portion de texte qui s'avère inutile à la compréhension du thème (Mahmoudi et Faïz, 2006_a).

2 SDET : Un outil pour l'enrichissement des données

Notre approche a été mise en œuvre pour permettre un support informationnel pour les utilisateurs de SIG. L'implémentation de notre approche a été réalisée en utilisant le langage

L'outil SDET pour le complètement des données descriptives

Java. Ce choix est fondé sur le fait que Java supporte le multi-threading qui favorise le traitement parallèle entre les différents agents du système. Les fonctionnalités que nous proposons ont été intégrées à un SIG OpenSource : Open Jump¹ (Java Unified Mapping Platform). Ce SIG supporte les principaux standards industriels comme GML ainsi que le modèle objet spatial de l'OpenGIS Consortium. Open Jump offre une grande modularité et de nombreuses extensions.

SDET est une suite de fonctionnalités facilitant l'extraction de nouvelles données qui viennent enrichir celles déjà existantes dans la Base de Données Géographiques (BDG). Nous procédons comme suit : L'utilisateur du SIG à la quête d'informations, va solliciter en premier lieu la BDG du système. En cas d'insatisfaction (information non stockée dans la BDG, détails insuffisants...), l'utilisateur peut faire appel à notre outil d'enrichissement. Ceci est rendu possible en ajoutant une nouvelle option au SIG libre : Open Jump. L'enrichissement est initié par une collecte de corpus de documents. A partir de ces documents bruts, nous extrayons la liste des thèmes traités et leurs documents générés associés.

L'outil propose trois manières pour condenser les documents générés : Le résumé d'un segment sélectionné par l'utilisateur, le résumé de l'intégralité du document généré ou le résumé de l'intégralité du document généré, mais, avec un filtrage appliqué aux segments afin d'éliminer toute redondance d'information. Les résumés résultants peuvent être affichés et stockés pour une consultation ultérieure.

Références

- Faïz, S. et K. Mahmoudi (2005). Semantic Enrichment of Geographical Databases. *Encyclopedia of database technologies and applications*. Edition Idea Group, Etats-Unis, 587-592.
- Ferber, J. (1997). *Les Systèmes Multi-Agents vers une intelligence collective*. Edition InterEditions.
- Mahmoudi, K. et S. Faïz (2006_a). Une approche distribuée pour l'extraction de connaissances : Application à l'enrichissement de l'aspect factuel des BDG, *Revue des Nouvelles Technologies de l'Information*. Edition Cepaduès, 107-118.
- Mahmoudi, K. et S. Faïz (2006_b). *L'apport de l'information spatiale pour l'enrichissement des bases de données*. INFORSID'2006, Hammamet, Tunisie, 323-338.

Summary

Database enrichment is a mean to provide complementary data to end users. In a geographic context, this is essential to make accurate decisions. SDET is our tool built to perform the data enrichment that we integrated to an open GIS.

¹ <http://www.jump-project.org/>.