

Du texte à la base de données géographiques

Nesrine Hassini, Khaoula Mahmoudi, Sami Faiz

Laboratoire de Télédétection et Systèmes d'Information
à Références Spatiales, ENIT BP 37 le Belvédère 1002 Tunis, TUNISIE
hassnes@yahoo.fr, kamahmoudi@yahoo.fr, sami.faiz@insat.rnu.tn

Résumé. Avec la prolifération des données géographiques, il y a un fort besoin de concevoir des outils automatiques pour l'exploitation des connaissances géographiques incarnées dans les documents textuels. C'est dans ce contexte, que nous proposons une approche permettant de générer une base de données géographiques (BDG) à partir de textes. Notre approche s'articule autour de deux grandes phases : la génération du schéma de la BDG et la détermination des données qui serviront au remplissage de cette base. L'implémentation de notre approche a donné naissance à un outil que nous avons baptisé GDB Generator et que nous avons intégré dans le SIG : OpenJUMP.

1 Introduction

L'explosion récente des technologies mobiles et des données géo-référencées a fait émerger un nouveau type de données, qualifiées de géographiques. De ce fait, une prolifération des systèmes d'informations géographiques (SIG) a vu le jour pour assurer une meilleure exploitation de ces informations. En fait, pour atteindre cet objectif on doit disposer d'un ensemble d'informations complet et cohérent. Des moyens d'enrichissement ont été proposés pour aboutir à l'enrichissement des BDG à des coûts réduits. Dans ce contexte, on trouve à titre d'exemple MetaCarta (www.metacarta.com) et PIV (Lesbegueries et al., 2006) qui opèrent l'enrichissement en liant les documents textuels aux entités géographiques correspondantes. Ces travaux, ne proposent pas (ou peu) de moyens pour gérer les contenus textuels des documents. L'outil SDET (Mahmoudi et Faiz, 2010) est un autre moyen d'enrichissement qui cherche à exploiter le contenu textuel pour extraire l'essentiel sous forme d'un résumé. Le présent travail s'inscrit dans le contexte d'enrichissement des SIG tout en proposant une vue structurée des informations sous format de BDG. Concrètement, notre approche s'articule autour de deux grandes phases : une correspondance du texte en schéma conceptuel et une génération de la BDG à partir du schéma suite à son remplissage.

2 Notre approche

Le processus général de génération d'une structure sous forme de BDG à partir de texte vise un enrichissement de la BDG initialement incarnée dans le SIG. L'idée est de chercher un ensemble d'attributs pouvant compléter les données existantes. Notre approche se décompose