

Une approche distribuée pour l'extraction de connaissances : Application à l'enrichissement de l'aspect factuel des BDG

Khaoula Mahmoudi*
Sami Faïz ** ***

* Ecole Supérieure des communications de Tunis (SUPCOM)

khaoula.mahmoudi@insat.rnu.tn

** Institut National des Sciences Appliquées et de Technologie (INSAT)

*** Laboratoire de Télédétection et Systèmes d'Informations à Références Spatiales (LTSIRS)
sami.faiz@insat.rnu.tn

Résumé. Les systèmes d'informations géographiques (SIG) sont utilisés pour améliorer l'efficacité des entreprises et des services publics, en associant méthodes d'optimisation et prise en compte de la dimension géographique. Cependant, les bases de données géographiques (BDG) stockées dans les SIG sont restreintes à l'application pour laquelle elles ont été conçues. Souvent, les utilisateurs demeurent contraints de l'existant et se trouvent dans le besoin de données complémentaires pour une prise de décision adéquate. D'où, l'idée de l'enrichissement de l'aspect descriptif des BDG existantes. Pour atteindre cet objectif, nous proposons une approche qui consiste à intégrer un module de fouille de données textuelles au SIG lui même. Il s'agit de proposer une méthode distribuée de résumé de documents multiples à partir de corpus en ligne. L'idée est de faire coopérer un ensemble d'agents s'entraïdant afin d'aboutir à un résumé optimal.

1 Introduction

Le but d'un SIG est de fournir une aide à la décision dans des domaines divers. Souvent, il sert à produire des cartes répondant à un besoin spécifique. Il peut être utilisé pour associer une densité de population à chaque région sur une carte, la représentation de la présence de consommateurs potentiels d'un produit ou d'un service dans une région, etc. Les données sont dans tous les cas restreintes à l'application en cours et parfois on a besoin d'avoir des informations au-delà de ce qui est stocké dans la BDG. A titre d'exemple, une BDG créée pour une application de découpage administratif ne permet pas de fournir une réponse à une requête faisant intervenir des informations d'ordre économique, historique, etc. D'où, l'idée d'offrir des sources complémentaires d'informations sans nuire aux données préalablement fournies (Faïz et Mahmoudi, 2005). Pour atteindre cet objectif, nous avons bâti une approche pour la génération automatique de résumés de documents multiples pour fournir les informations complémentaires relatives aux entités géographiques manipulées par le SIG. Cette approche est basée sur trois types d'agents coopérant afin d'aboutir à un résumé optimal. Il s'agit d'un agent *interface*, des agents *entité* (géographiques) et des agents *tâche*. La communication entre ces agents est assurée par l'envoi de messages. L'approche est