

Extraction des connaissances pour l'enrichissement des bases de données géographiques

Sami Faïz* **
Khaoula Mahmoudi***

* 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

***Ecole Supérieur des communications de Tunis (SUPCOM)
khaoula.mahmoudi@insat.rnu.tn

Utilisés de façon croissante depuis une quinzaine d'années, les Systèmes d'Informations Géographiques (SIG) permettent le stockage, l'interrogation, la manipulation, le partage, la diffusion et la restitution d'informations à composante géographique, telles que les villes, les rues ou les fleuves. Un SIG comme son nom l'indique traite principalement des aspects cartographiques. Dans une situation où l'utilisateur désire obtenir des informations d'ordre économiques, historiques, etc..., le SIG ne peut être à lui tout seul suffisant. D'où, la nécessité d'un enrichissement sémantique des bases de données géographiques manipulées, ces systèmes deviennent alors des sources riches de données.

Dans ce contexte, nous profitons des informations abondantes sur le Web pour ajouter des informations complémentaires aux SIG. En fait, le volume d'informations disponibles sur Internet ne cessant d'augmenter chaque jour, on utilisera la technique du textmining pour permettre à partir des textes du web d'extraire et d'explicitier l'essentiel de l'information et ce sous forme de résumés. L'application de ces techniques d'analyse textuelle est initiée en fournissant des éléments localisés par pointé ou fenêtrage via le SIG. Par la suite, un accès au Web prend lieu pour le téléchargement des documents pertinents.

La résolution de ce problème se réalise d'une manière distribuée entre un ensemble d'agents interagissant, afin, de générer le meilleur résumé. Un agent *interface* qualifié de méta-agent est créé par le système. Cet agent crée à son tour des agents *summarizer* (autant d'agents *summarizer* que de documents). Chacun de ces agents prend à sa charge la segmentation du texte de son document en fixant les frontières thématiques. Une fois les segments cernés, ils sont envoyés à l'agent *interface* qui se met à les regrouper ensemble selon le degré de similarité des thèmes. Par la suite, il affecte un délégué pour chaque thème et ce en tenant compte du coût engendré par sa décision. Il s'agit en fait de minimiser la surcharge de chaque délégué tout en distribuant au maximum la tâche de génération de résumés entre le plus grand nombre d'agents.

A l'issue de cette décision chaque délégué doit interagir avec ses accointances (les agents qui traitent le(s) thème(s) sous sa responsabilité) par envoi de messages afin de collecter les segments correspondants. Désormais, le délégué tient à sa disposition un document virtuel. Il est appelé ainsi puisqu'il ne résulte pas d'une tâche de recherche d'informations, mais, de la collecte de segments issus des différents agents *summarizer*. Enfin, chaque délégué extrait les informations les plus saillantes qui résumement mieux ses documents virtuels. Le résumé final résulte de l'assemblage des résumés partiels générés par chaque agent délégué. Ce résumé final étant transféré au SIG.