

Utilisation d'une ontologie du domaine pour la découverte du contenu de bases de données géographiques

Ammar Mechouche, Nathalie Abadie, Emeric Prouteau, Sébastien Mustière

Institut Géographique National, Laboratoire Cogit, 73 Av. de Paris, 94160 St-Mandé, France

Résumé. L'essor récent des technologies associées à la géomatique a permis la production rapide de nombreuses données géographiques. Or, pour tirer profit de ces données, il convient de pouvoir évaluer leur pertinence et leur complexité vis-à-vis de l'application à laquelle on les destine. Dans cet article, nous présentons une application permettant à un utilisateur de découvrir le contenu de bases de données géographiques, à savoir, quels types d'entités géographiques sont représentés au sein de chaque base et comment. Pour accéder à ces informations, l'utilisateur interroge le système via une ontologie globale du domaine qui décrit les types d'entités topographiques du monde réel. Des ontologies locales (ou d'application) sont utilisées pour formaliser les spécifications de chaque base de données décrite. Elles sont annotées à l'aide de concepts issus de l'ontologie globale. Ce système est implémenté sous la forme d'une interface Web et inclut un affichage cartographique d'échantillons de données.

1 Contexte et objectifs

Au cours des dernières décennies, les Systèmes d'Information Géographique (SIG) n'ont cessé de gagner en importance, suscitant le développement rapide de la production de données géographiques, de leur diffusion, ou encore d'applications de géolocalisation (Craglia et al., 2008). Les nombreuses sources de données géographiques désormais disponibles s'avèrent utiles pour de nombreuses applications telles que l'environnement, l'urbanisme ou l'agriculture. Cependant, d'une base à l'autre, un même type d'entités géographique pourra être représenté différemment selon le point de vue adopté par son producteur lors de la conceptualisation de la base et l'établissement de ses spécifications (Fonseca et al., 2003), phénomène que l'on qualifie d'hétérogénéité sémantique des données (Partridge, 2002), et qui est à l'origine de nombreuses difficultés dans la réutilisation conjointe de bases de données hétérogènes.

En raison de la variété des données géographiques disponibles et de la complexité de leurs spécifications, les utilisateurs peuvent éprouver de grandes difficultés à évaluer et comprendre précisément le contenu de ces bases de données. Le développement de portails Web permettant aux utilisateurs de visualiser les données géographiques disponibles a permis d'améliorer la compréhension de ces données. Cependant, en dépit de ces géoportails, un grand nombre d'informations concernant les données restent inaccessibles à des utilisateurs plus expérimentés en géomatique ; il demeure impossible pour des spécialistes de différents domaines désireux d'apprécier et de comparer les contenus des diverses bases de données disponibles vis-à-vis de leurs besoins spécifiques d'accéder simplement aux informations utiles, et en particulier aux spécifications de ces bases.