

1.1 Retombées scientifiques et techniques

D'un point de vue méthodologique, le principal objectif du projet est d'éprouver et d'adapter des outils et méthodes existants pour créer, enrichir, aligner et exploiter des ontologies à partir de données et de connaissances qui ont la particularité d'être, d'une part réelles, et donc entachées d'erreurs ou d'imprécisions, et d'autre part hétérogènes en terme de structure et de niveaux de structuration.

Du point de vue du cadre applicatif, à savoir la description et l'intégration de données géographiques, le principal objectif est de créer des ontologies riches en termes de quantité de concepts manipulés et en termes de sémantique afin de permettre l'accès à une masse de données géographiques hétérogènes, qu'il s'agisse de contenus de bases de données ou de textes grand public. Le deuxième objectif est d'appréhender de manière fine les différences de point de vue sous-jacentes à diverses sources de connaissances du domaine.

Les résultats attendus dans ce projet sont les suivants :

- Des outils nouveaux d'extraction de concepts et de relations dans des textes, basés sur la définition de patrons, sur le repérage d'Entités Nommées Géographiques, sur l'exploitation de la structure des textes et des relations argumentatives,
- Des techniques et une méthodologie de création, d'enrichissement et de restructuration d'ontologie, combinant l'utilisation de patrons, de ressources lexicales externes, de techniques d'alignement d'ontologies,
- Une étude de l'efficacité des techniques d'alignement, l'adaptation de techniques actuelles et le développement de nouvelles techniques adaptées à la fusion efficace de deux ontologies hétérogènes de qualité moyenne (cas des ontologies réelles),
- Des techniques et une méthodologie de comparaison de deux ontologies reflétant des niveaux d'échelle spatiale différents ou issues de pays différents, afin d'étudier si ces ontologies reflètent de véritables points de vue différents ou si elles n'ont que des différences marginales et gagneraient donc à être fusionnées.
- Une ontologie du domaine de l'information géographique, et plus particulièrement de la description topographique du paysage,
- Des techniques et méthodologies exploitant une ontologie pour indexer automatiquement le contenu de documents diversifiés et pour appairer des schémas de bases de données hétérogènes.

Le projet étant motivé par la volonté d'intégrer des informations à caractère géographiques, les techniques et les méthodes proposées, bien que génériques, c'est-à-dire applicables a priori à tout domaine d'application, seront validées sur ce domaine. Tous les résultats seront évalués d'un point de vue géographique par les chercheurs du laboratoire COGIT de l'IGN, experts du domaine.

Une ontologie topographique riche sera mise à disposition de la communauté scientifique. Elle sera en bilingue français/anglais et répondra aux attentes des groupes de travail associés à la directive européenne INSPIRE qui ont exprimé le besoin de disposer d'ontologies dans le domaine géographique, les ontologies existant aujourd'hui étant très restreintes et décrivant des domaines très ciblés.

Le contexte applicatif du projet permettra d'assurer la pérennité des résultats. Cela aura lieu d'une part, à travers un site web maintenu par l'IGN mettant à disposition les ontologies définies et fournissant un certain nombre d'outils associés (moteur d'accès aux schémas des bases de données, résultats de l'appariement de schémas, etc.). Par ailleurs, comme déjà indiqué, le LIUPPA dispose d'une plateforme expérimentale basée sur des Services Web, ces services, décrits en WSDL et instanciés dans un service UDDI, sont, aujourd'hui déjà, une preuve de notre capacité à pérenniser nos travaux¹.

¹ A cause de tests nécessaires liés à la sécurité, ces services, bien qu'étant déclarés dans un annuaire, ne sont, pour le moment, disponibles que sur demande explicite.

Les outils mis au point pour l'appariement de schémas bases de données au COGIT seront développés dans le système d'information géographique OpenSource GéOxygène. L'opportunité de leur dépôt en OpenSource ainsi que leur éventuelle intégration au sein de services web sera étudiée.

Au niveau dissémination scientifique des résultats, ceux-ci seront publiés dans des conférences internationales (International Conference on Knowledge Capture (K-CAP), International Conference on Knowledge Engineering and Knowledge Management (EKAW), International Semantic Web Conference (ISWC), Ontology-Mapping Workshop), nationales (RFIA, IC, TIA, EGC, BDA) ou revues réputées sur l'intégration de bases de données ou les ontologies (Knowledge Engineering Review, Journal on Data and Knowledge Engineering, Very Large Database Journal, TSI), ainsi que sur l'information géographique (par exemple Revue Internationale de Géomatique, IJGIS, AGILE). Les résultats seront également publiés sur le site Web du projet, ce qui permettra aux organismes intéressés de suivre l'avancement du projet et de prendre connaissance des résultats obtenus.