



HAL
open science

Création d'une ontologie à partir des spécifications textuelles pour l'intégration des bases de données géographiques

Nathalie Abadie, Nils Gesbert, Sébastien Mustière

► To cite this version:

Nathalie Abadie, Nils Gesbert, Sébastien Mustière. Création d'une ontologie à partir des spécifications textuelles pour l'intégration des bases de données géographiques. 17ème conférence en Ingénierie des Connaissances (IC'06), Jun 2006, Nantes, France. hal-02411432

HAL Id: hal-02411432

<https://hal.science/hal-02411432>

Submitted on 14 Dec 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Création d'une ontologie à partir des spécifications textuelles pour l'intégration des bases de données géographiques

Nathalie Abadie, Nils Gesbert et Sébastien Mustière

Laboratoire COGIT, IGN, 2/4 Avenue Pasteur, F-94165 Saint-Mandé cedex, <http://www.ign.fr>, {nathalie-f.abadie ; nils.gesbert ; sebastien.mustiere}@ign.fr

Résumé

Les bases de données géographiques résultent d'une conceptualisation particulièrement complexe du monde réel. Aussi, les organismes chargés de leur saisie se basent-ils sur des documents textuels très volumineux, les spécifications, qui décrivent le contenu de la base de données à créer. Nous proposons ici d'utiliser ces connaissances pour créer une ontologie géographique et définir les liens existant entre cette ontologie et les schémas conceptuels de différentes bases de données géographiques. La démarche présentée vise à faciliter l'intégration de plusieurs bases de données géographiques.

Mots clés : Spécifications ; bases de données géographiques ; ontologie ; intégration.

1 Introduction

Dans le cas des bases de données géographiques, l'essentiel des connaissances décrivant les données est consigné sous forme de textes qui sont les spécifications de contenu des bases. Nous présentons ici l'approche proposée par Nils Gesbert [2] [3] pour déterminer les correspondances entre les schémas de différentes bases de données en s'appuyant sur la sémantique contenue dans ces spécifications. Celle-ci se base sur une ontologie du domaine, et sur un modèle formel de représentation des spécifications.

2 Contexte

2.1 Les bases de données géographiques et leurs spécifications

Les bases de données topographiques telles que celles produites à l'Institut Géographique National visent à décrire la topographie du terrain réel. Ces bases de données sont basées sur un modèle orienté objet, chaque classe de la base représentant un type d'entités géographiques correspondant aux concepts géographiques communément admis (les mers, les forêts, les états, etc.). Ces concepts présentent la particularité d'être indissociables de leur localisation [5]. Or, cette conceptualisation du monde réel

constitue un processus cognitif complexe. En outre, ces bases de données ne sont jamais exhaustives dans la mesure où il est impossible d'appréhender le monde réel dans sa totalité.

Lors de la création des bases de données géographiques, on procède donc à une sélection des entités géographiques devant figurer dans la base, basée sur divers critères (nature, taille, importance relative, etc.). Afin de garantir une homogénéité satisfaisante du contenu des bases, ces critères de sélection sont détaillés au sein de volumineux documents papier, rédigés en langage naturel, qui sont les spécifications de contenu des bases de données géographiques.

2.2 Le besoin d'intégration des bases de données géographiques

Il existe fréquemment plusieurs bases de données géographiques couvrant une même zone, produites indépendamment les unes des autres, dans des buts et avec des moyens différents.

Or, l'indépendance de ces bases de données géographiques suscite des problèmes liés à la répétition des opérations de mise à jour, au manque de cohérence entre les différentes bases de données, et à l'impossibilité d'effectuer des analyses utilisant des données issues de bases différentes.

Il convient donc d'intégrer ces bases afin de disposer d'un ensemble de données structuré et cohérent. [1]

3 Modèle proposé pour la représentation des spécifications

3.1 Démarche proposée pour l'intégration des BDG

Pour effectuer le travail nécessaire de mise en correspondance sémantique des schémas conceptuels des différentes bases de données à intégrer, [4] propose d'explicitier les déviations des schémas conceptuels de données des bases par rapport à une ontologie du domaine, puis de se baser sur cette ontologie pour définir le schéma intégré.

C'est cette démarche que nous avons adoptée et mise en œuvre pour les bases de données géographiques [2] [3]. Ainsi, l'ontologie du domaine représentera l'ensemble des concepts partagés, relatifs à l'étude de l'information topographique.

3.2 Création de l'ontologie du domaine

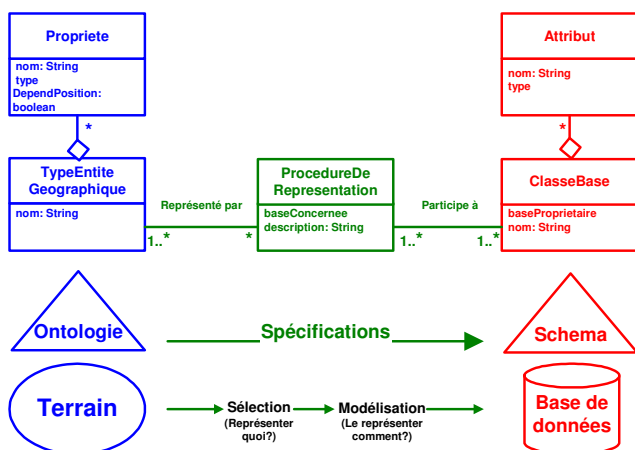
Il s'agit donc, dans un premier temps, de créer une ontologie permettant de représenter le terrain conceptualisé, c'est-à-dire l'ensemble des concepts géographiques utilisés par les experts du domaine lors de la rédaction des spécifications. Une ontologie géographique, portant sur le thème « Hydrographie », a donc été réalisée manuellement à partir de mots-clés extraits des spécifications.

3.3 Liens entre l'ontologie et les bases de données

Une fois l'ontologie définie, il convient d'établir les liens qui existent entre celle-ci et les schémas conceptuels des bases de données à intégrer. Les spécifications décrivant la sémantique du contenu de la base à travers le processus d'acquisition des données, leur lecture permet de déterminer quels types d'entités géographiques sont représentés par les instances d'une classe donnée.

3.4 Modèle proposé pour la représentation des spécifications

Le modèle de représentation des spécifications proposé s'appuie donc sur une ontologie du domaine pour décrire le terrain conceptualisé, et sur le schéma conceptuel de données de la base. Les spécifications sont considérées comme des règles de représentation faisant le lien entre l'ontologie et le schéma conceptuel. (voir FIG.1)



(FIG. 1) Structure générale du modèle proposé

L'ensemble des règles formalisées permettant de décrire comment un type d'entités géographiques est représenté dans la base constitue ce que nous appelons une « procé-

ture de représentation ». Afin d'explicitier ces procédures de représentation, et de faciliter leur utilisation informatique, un langage de description a été défini à l'aide d'une grammaire BNF. Ce dernier permet de décrire entièrement les conditions de *sélection*, d'*agrégation*, de *découpage*, et d'*instanciation* des entités du monde réel au sein des classes de la base. Par exemple, pour la classe des « Surfaces d'eau », il est spécifié que l'on sélectionnera « Toutes les surfaces d'eau de plus de 20m de long [...] ainsi que les cours d'eau de plus de 7,5m de large. ».

4 Conclusion

La démarche proposée pour l'intégration des bases de données géographiques consiste donc à utiliser les connaissances issues des spécifications pour créer une ontologie du domaine et expliciter les liens entre les schémas conceptuels des bases à intégrer et cette ontologie, sous la forme de procédures de représentation [2] [3].

A ce jour, ce modèle de représentation des spécifications a été testé sur le thème « Hydrographie » des deux principales bases de données géographiques produites par l'IGN. Les résultats obtenus étant satisfaisants, ces tests seront étendus à l'ensemble des thèmes de ces bases. Par ailleurs, ces premiers tests se sont basés sur une ontologie du domaine, réalisée manuellement, par détection de mots-clés dans les textes des spécifications. Il est prévu, par la suite, d'améliorer la procédure de création de l'ontologie.

Références

- [1] T. Devogele, C. Parent et S. Spaccapietra. On spatial database integration. *International Journal of Geographical Information Science*, vol 12, n° 4, pages 335-352, Taylor & Francis, 1998.
- [2] N. Gesbert. Etude de la formalisation des spécifications de bases de données géographiques en vue de leur intégration. Thèse de doctorat, Université de Marne-la-Vallée, 2005.
- [3] N. Gesbert, T. Libourel et S. Mustière. Apport des spécifications pour les modèles de bases de données géographiques. *Revue internationale de Géomatique*, vol 14, n° 2, pages 239-257, Lavoisier, 2004.
- [4] C. Partridge. The role of ontology in integrating semantically heterogeneous databases. *Rapport technique 05/02 LADSEB-CNR*, Padoue, 2002.
- [5] B. Smith et D.M. Mark. Ontology and Geographic Kinds. *Proceedings of the 8th international symposium on spatial data handling (SDH'98)*, pages 308-320, Vancouver : International Geographical Union, Geographic Information Science Study Group, 1998.