



HAL
open science

Constitution d'une taxonomie géographique à partir des spécifications de bases de données

Nathalie Abadie, Sébastien Mustière

► To cite this version:

Nathalie Abadie, Sébastien Mustière. Constitution d'une taxonomie géographique à partir des spécifications de bases de données. Colloque International de Géomatique et d'Analyse Spatiale SAGEO 2008, Jun 2008, Montpellier, France. hal-02411372

HAL Id: hal-02411372

<https://hal.science/hal-02411372v1>

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.

Constitution d'une taxonomie géographique à partir des spécifications de bases de données

Nathalie Abadie — Sébastien Mustière

*Laboratoire COGIT, Institut Géographique National
2 Avenue Pasteur 94160 Saint-Mandé Cedex
{nathalie-f.abadie, sebastien.mustiere }@ign.fr*

RÉSUMÉ. Les ontologies sont des ressources essentielles pour gérer l'hétérogénéité sémantique entre diverses sources d'informations. En pratique, il existe malheureusement peu d'ontologies géographiques. Nous présentons ici des travaux réalisés pour créer une taxonomie de concepts topographiques, étape préliminaire à la constitution d'une véritable ontologie, à partir de l'analyse semi-automatique de documents textuels particuliers: des spécifications de bases de données géographiques. Plus de 700 concepts ont ainsi été identifiés et hiérarchisés. Cette taxonomie ouvre la voie à diverses applications, comme l'appariement de données et de schémas, la formalisation et la consultation des spécifications.

ABSTRACT. Ontologies are central tools to manage semantic heterogeneity between various sources of information. Unfortunately, only few geographic ontologies exist in practice. In this paper, we present works undertaken to build one taxonomy of geographical concepts, a necessary step toward an actually rich ontology, through the semi-automatic analysis of textual specifications of geographic databases. More than 700 hundreds concepts have been identified and organised in this taxonomy, paving the way to several applications like data and schema matching, or intuitive portals to access geographic data.

MOTS-CLÉS: ontologie, taxonomie, traitement automatique du langage naturel, spécifications.

KEYWORDS: ontology, taxonomy, natural language processing, specifications.

1. Motivation

Avec d'une part l'essor des technologies de l'information, et d'autre part le développement des techniques de localisation spatiale, les données géographiques sont de plus en plus nombreuses et diverses. Or, chacune de ces bases de données géographiques reflète une conceptualisation particulière du monde réel, propre à son producteur [Aerts et al. 2006]. Aussi, d'une base de données à l'autre, un même phénomène du monde réel pourra-t-il avoir différentes descriptions et interprétations, difficulté qui peut être qualifiée « d'hétérogénéité sémantique » des données [Partridge 2002]. S'il ne s'agit en aucun cas d'une caractéristique exclusive des données géographiques, ce phénomène d'hétérogénéité sémantique est particulièrement vérifié dans ce domaine, en raison de la complexité du monde géographique représenté [Smith et al. 1998], et pose problème pour diverses applications. [Cullot et al. 2003] prônent ainsi une meilleure compréhension des données et une automatisation des processus de navigation, d'extraction et de mise à jour des informations géographiques sur Internet, grâce aux outils du Web sémantique, rejoignant ainsi [Egenhofer, 2002] qui, parmi d'autres, appelle de ses vœux la mise en place d'un Web sémantique géospatial. [Partridge 2002] et [Gesbert 2005] insistent également sur la nécessité d'appréhender les différences de point de vue entre bases de données géographiques, pour permettre leur intégration.

Pour résoudre les divers problèmes liés à l'hétérogénéité sémantique des données, un consensus se fait globalement autour de la nécessité de disposer d'ontologies, c'est-à-dire d'outils permettant de spécifier sans ambiguïté la sémantique de systèmes terminologiques donnés [Gruber 1993], [Guarino 1998]. Celles-ci sont encore peu nombreuses dans le domaine de la géographie, même si l'on recense quelques travaux de mise en place d'ontologies de haut niveau ou d'ontologies du domaine, présentés de façon exhaustive par [Bittner et al. 2005] et [Agarwal 2005]. De façon générale, la création d'une seule et même ontologie du domaine permettant l'intégration de tous les points de vue existants reste extrêmement coûteuse et difficilement réalisable [Rodriguez 2003], voire peu souhaitable. Il nous semble donc important que chaque communauté d'acteurs de l'information géographique constitue sa propre ontologie du domaine, reflétant son point de vue sur le monde, et explicitant la sémantique des concepts partagés par ses membres. A ce titre, les projets *Towntology* [Roussey et al., 2004] et *FoDoMuSt* [Brisson et al., 2007], qui ont conduit respectivement à la création d'une ontologie dédiée au domaine de l'aménagement et de l'urbanisme, et à celle d'une ontologie des objets géographiques susceptibles d'être visibles sur des images aériennes ou satellites, constituent deux références de mise en place d'ontologies géographiques relatives à des domaines bien spécifiques.

Dans ce contexte, notre objectif est de créer une ontologie géographique du point de vue des topographes. Pour ce faire, une première étape est, plus modestement, de constituer une taxonomie de termes liés à la description de l'espace et reflétant le point de vue adopté à l'Institut Géographique National pour définir ses données de

référence. Nous proposons ici de réaliser cette tâche en exploitant les spécifications des bases de données de l'IGN, documents textuels particuliers présentés en détails dans la partie 2. La partie 3 retrace le processus mis en place pour analyser automatiquement le contenu des spécifications et constituer notre taxonomie topographique. Enfin, avant de conclure, nous exposons dans la partie 4 quelques utilisations de cette taxonomie, réalisées ou envisagées.

2. Les spécifications des bases de données géographiques

Dans le cas des bases de données géographiques, les informations décrivant les données se situent à plusieurs niveaux. Aussi pourra-t-on, en première approche, se référer aux schémas conceptuels des bases de données pour en comprendre le contenu. En outre, l'expertise des concepteurs et des administrateurs de ces bases de données constitue une source de connaissances supplémentaire très appréciable, mais malheureusement bien souvent incomplète [Partridge 2002]. En fait, la majeure partie de la sémantique des bases de données géographiques se trouve dans les spécifications mêmes de ces bases. Celles-ci, en effet, décrivent la sémantique du contenu d'une base de données à travers le processus d'acquisition des données.

2.1 La structure des spécifications des bases de données de l'IGN

La saisie des bases de données géographiques fait intervenir un grand nombre de règles de sélection et de représentation des entités géographiques du monde réel au sein des bases. Afin de garantir le respect de chacune de ces règles, ces dernières sont consignées dans de volumineux documents papier, rédigés en langage naturel, qui sont les spécifications de saisie de la base de données en question.

Une première partie de ces spécifications concerne la description générale de la base de données. Elle présente globalement son contenu, sa résolution, ses sources, etc. La seconde partie est organisée par thèmes : occupation du sol, hydrographie, routier, etc. Chaque thème regroupe des fiches descriptives des classes du schéma conceptuel de données. Au sein d'une même base, toutes ces fiches sont organisées selon le même modèle (voir Fig. 1). Tout d'abord y figure le nom de la classe décrite. Puis une partie *définition* précise la sémantique du (ou des) concept(s) représenté(s). Viennent ensuite les éventuelles *relations* concernant cette classe. La partie *regroupement* établit la liste des types d'entités topographiques du monde réel devant être saisis au sein de cette classe, tandis que les *critères de sélection* détaillent les conditions devant être respectées par les entités topographiques de chaque type énuméré dans la partie *regroupement* pour figurer dans cette classe de la base de données. Le paragraphe *modélisation géométrique* définit la géométrie de l'objet à représenter. Enfin, la partie *attributs* dresse l'inventaire des attributs de l'objet, autres que sa géométrie.

Ainsi, les spécifications des bases de données font référence à un très grand nombre de concepts topographiques, à travers les valeurs d'attributs ou de manière plus informelle au milieu de phrases en langage naturel. Rédigées pour un producteur de données de référence, les spécifications se veulent aussi le fruit d'un consensus sur le monde topographique. Destinées, en premier lieu, à garantir un niveau d'homogénéité satisfaisant au sein des bases de données géographiques, elles visent à réduire, autant que faire se peut, les ambiguïtés concernant l'interprétation du contenu de ces bases. Enfin, la structure même de ces spécifications est source d'information concernant notamment les relations entre les concepts topographiques rencontrés au fil du texte.

Canalisation	
<p>Définition : Canalisation ou tapis roulant. Géométrie : Linéaire tridimensionnelle</p>	<p>Attributs</p> <ul style="list-style-type: none"> • Source géométrique des données ⁽¹⁾ • Nature • Position par rapport au sol ⁽¹⁾ • Z_Initial ⁽¹⁾⁽²⁾ • Z_Final ⁽¹⁾⁽²⁾ <p><small>(1) voir les spécifications générales (2) uniquement pour les formats 2D</small></p>
<p>Regroupement : Voir les différentes valeurs de l'attribut <nature>. Sélection : Uniquement les canalisations aériennes et celles qui figurent sur la carte au 1 : 25 000 en service. Modélisation géométrique : A l'axe et sur le dessus de la canalisation.</p>	
Attribut : Nature	
<p>Définition : Attribut permettant de différencier les canalisations d'eau des autres. Type : Énuméré Valeurs : Eau / Autre</p>	
<p>Nature = « Eau »</p> <p>Définition : Aqueduc (ouvrage maçonné et couvert destiné à transporter l'eau potable suivant une faible pente) ou conduite forcée (canalisation permettant le transfert d'eau en charge (gravitaire et sous pression) vers un ouvrage hydraulique). Regroupement : Aqueduc Conduite forcée Galerie d'amenée d'eau</p>	
<p>Nature = « Autre »</p> <p>Définition : Canalisation ou tapis roulant utilisé pour le transport de matière première (gaz, hydrocarbure, minéral, etc.) ou canalisation de nature inconnue. Regroupement : Conduite de matière première Gazoduc Oléoduc Pipe-line Tapis roulant industriel</p>	

Figure 1. Extrait des spécifications de la BDTOP pour la classe "Canalisation"

2.2 Des spécifications complexes

L'organisation actuelle des spécifications, qui suit celle du schéma conceptuel de la base de données, peut cependant prêter à confusion. En effet, il pourrait être tentant de considérer que chaque classe ainsi présentée correspond directement à un concept géographique du monde réel, et que les spécifications fournies ne servent qu'à représenter les critères de sélection et de modélisation des objets géographiques en vue de leur instanciation.

Cependant, lorsque l'on examine plus précisément ces spécifications, on découvre que les regroupements d'entités topographiques au sein d'une classe diffèrent d'une base de données à l'autre, ceci en raison de niveaux de détail ou bien de choix de modélisation différents entre les bases. Par exemple, dans la BDCARTO®, les aqueducs appartiennent à la classe des « tronçons de cours d'eau », tandis qu'ils rejoignent les gazoducs et les oléoducs au sein de la classe des « canalisations » dans la BDTPOPO®.

Ainsi, seule une parfaite connaissance des spécifications des bases de données géographiques permet d'appréhender avec exactitude le contenu de chaque base de données. Ceci permet, entre autres, d'interpréter les différences de représentations entre bases, et de détecter d'éventuelles incohérences entre ces bases, lorsque ces différences ne sont pas prévues par les spécifications [Sheeren 2005] [Sheeren et al. 2004]. Enfin, bien que très structurées, ces spécifications n'en demeurent pas moins destinées à des humains. Il s'agit donc encore largement de textes non formels qui contiennent de nombreuses parties exprimées en langage naturel. Un processus semi-automatique d'analyse de la structure et du contenu de ces documents doit donc être mis en place pour espérer les exploiter au mieux.

3. Des spécifications vers une taxonomie

Les spécifications de contenu des bases de données géographiques constituent la meilleure source de connaissances disponible sur ces bases. Nous proposons ici de mettre à profit ces informations pour permettre la construction d'une taxonomie portant sur le domaine de la topographie.

Selon [Partridge 2002], chaque schéma conceptuel de base de données s'appuie sur une ontologie, commune à toutes les bases du domaine, nommée "ontologie sous-jacente". Ainsi, un schéma conceptuel dont les éléments correspondraient exactement aux concepts de cette ontologie serait qualifié d'ontologique. Nous avons vu précédemment que cela n'est pas le cas pour les bases de données géographiques, puisque les relations entre les classes de ces bases et les concepts du terrain réel, décrites dans les spécifications, ne sont pas triviales. Cependant, pour reconstituer cette ontologie sous-jacente, nous proposons d'exploiter les connaissances contenues dans les spécifications de ces bases. Ainsi, pour retrouver quels sont les concepts topographiques constitutifs de cette ontologie sous-jacente,

nous suggérons de nous référer à la description précise des objets géographiques contenus dans les bases, fournie par les spécifications de ces bases. En outre, pour reconstituer la structure de l'ontologie, nous proposons de mettre à profit l'organisation hiérarchique des textes de ces spécifications [Mustière et al. 2007].

3.1 Traitement automatique du langage pour la construction de taxonomies

La méthode de reconstitution de l'ontologie sous-jacente proposée a été mise en œuvre pour deux bases de données topographiques de l'IGN, la BDTPOPO® et la BDCARTO®. Cette tâche s'avérant particulièrement longue, nous avons choisi, dans un souci d'automatisation, de recourir à des techniques de traitement automatique du langage naturel (TALN). Il s'agit donc, dans un premier temps, de retrouver dans les textes les concepts topographiques, puis de reconstituer automatiquement la hiérarchie de ces concepts à partir de leur localisation au sein des textes. Notons que les deux bases de données sont traitées dans un premier temps indépendamment, avant de fusionner les résultats (cf. 3.3).

L'utilisation de méthodes par règles a été privilégiée. En effet, les spécifications de ces deux bases sont des textes relativement courts en termes de TALN, de l'ordre d'une centaine de pages pour chaque base, au sein desquels certains concepts topographiques n'apparaissent que rarement. Elles s'avèrent donc peu propices à l'utilisation de méthodes statistiques et probabilistes de TALN. En revanche, il s'agit de documents extrêmement organisés. C'est pourquoi le choix a été fait de se baser sur la structure même des documents, et sur des heuristiques linguistiques, afin de détecter les concepts topographiques. Aussi, deux parseurs utilisés en début de traitement (voir Fig.2) ont-ils pour but de rendre explicite la structure des spécifications. Ainsi, en sortie de ces traitements, les spécifications sont converties en un fichier XML dont la structure reprend celle des spécifications en langage naturel: *nom de classes*, *nom d'attributs*, *parties définition* et *regroupement* sont extraits et stockés dans des balises de mêmes noms.

De plus, on émet l'hypothèse selon laquelle les concepts géographiques recherchés seront systématiquement représentés dans le texte par des groupes nominaux. Ainsi, l'étape suivante sera consacrée à l'extraction de tous les groupes nominaux du texte des spécifications à l'aide d'un transducteur (voir Fig.2). Cependant, la sélection de groupes nominaux effectuée par le transducteur demeure trop large, et doit être filtrée et triée. Ce filtrage est réalisé par comparaison avec un corpus non géographique, constitué de textes relatifs à des domaines très éloignés, comme la chimie ou la mécanique. D'autre part, un dictionnaire de langues permettra de trier cette sélection. Ainsi les groupes nominaux qui semblent être les plus pertinents seront présentés en priorité, pour validation, à un expert.

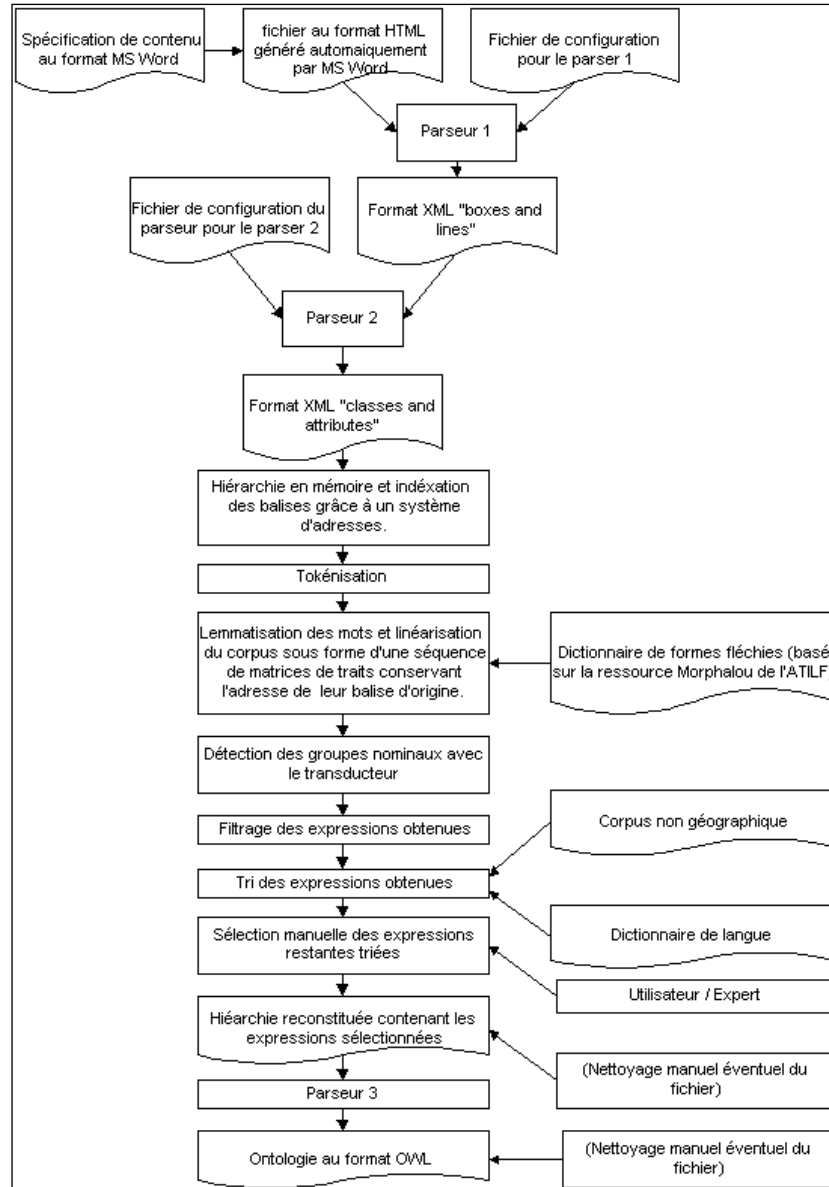


Figure 2. Enchaînement des traitements mis en oeuvre pour la constitution d'une taxonomie des concepts topographiques pour chaque base de données

Enfin, les relations de subsomption entre concepts topographiques sont inférées à partir de la distribution hiérarchique des groupes nominaux retenus dans le texte des spécifications. Ainsi, au sein de la fiche de spécifications d'une classe donnée, un concept topographique désigné par un groupe nominal issu de la partie *nom de classe* sera considéré comme plus général qu'un concept topographique qualifié par un groupe nominal extrait d'une partie *définition* de cette fiche de spécifications. La hiérarchie de concepts ainsi obtenue est finalement convertie au format standard préconisé par le W3C pour la représentation d'ontologies, OWL.

3.2 Résultats obtenus

L'approche présentée ici a été appliquée avec succès aux spécifications de la BDTOPO® et de la BDCARTO® afin d'en extraire deux taxonomies au format OWL. En effet, à partir des spécifications de la BDTOPO®, nous avons obtenu une taxonomie de 580 concepts topographiques, répartis sur quelques 11 niveaux hiérarchiques (voir extrait Fig. 3, en haut). Le même processus appliqué aux spécifications de la BDCARTO® a fourni une taxonomie de 505 concepts organisés sur 5 niveaux de hiérarchie (voir extrait Fig.3, en bas). Cependant, en dépit de ces résultats satisfaisants, un post-traitement manuel de l'ordre d'une journée reste nécessaire afin de pallier les manques de l'analyse automatique.

En effet, les concepts pertinents peuvent se rencontrer à tous niveaux de la structure des spécifications (noms de classes, définitions, valeurs d'attributs ou définitions d'attributs), ce qui rend leur détection parfois aléatoire. De plus, certains mots-clés obtenus traduisent une conceptualisation du monde réel trop proche des préoccupations de modélisation de type SIG. Ainsi, les concepts de « construction surfacique », « construction linéaire », et « construction ponctuelle » devront être fusionnés au profit du concept unique de « construction ». En outre, du fait de la structure même des spécifications, certaines hiérarchies de concepts sont mises à plat, et il est impossible de les reconstituer automatiquement. Ainsi, l'attribut « Nature » de la classe « Point remarquable du relief » pouvant prendre les valeurs « dune », « grotte », « sommet » ou « pic », l'analyse automatique des spécifications en déduit la hiérarchie de concepts « Point remarquable du relief » : « dune », « grotte », « sommet », « pic », sans que la proximité sémantique entre ces deux derniers concepts ne soit particulièrement mise en valeur. Par ailleurs, certains concepts ne peuvent être généralisés sans recourir à un sous-concept représentatif. Le concept de « gorge » se spécialise donc à la fois en « canyon », « cluse », « défilé » et « gorge », faute de terme général plus approprié. De surcroît, si dans certains cas les sous-concepts extraits sont de réelles spécialisations (pour « vallée » : « combe », « ravin », « thalweg », etc.), il s'agit parfois de synonymes, comme c'est le cas pour la hiérarchie « mairie » : « hôtel de ville », ou de véritables contresens dans l'exemple suivant « île » : « îlot », « presque île ».

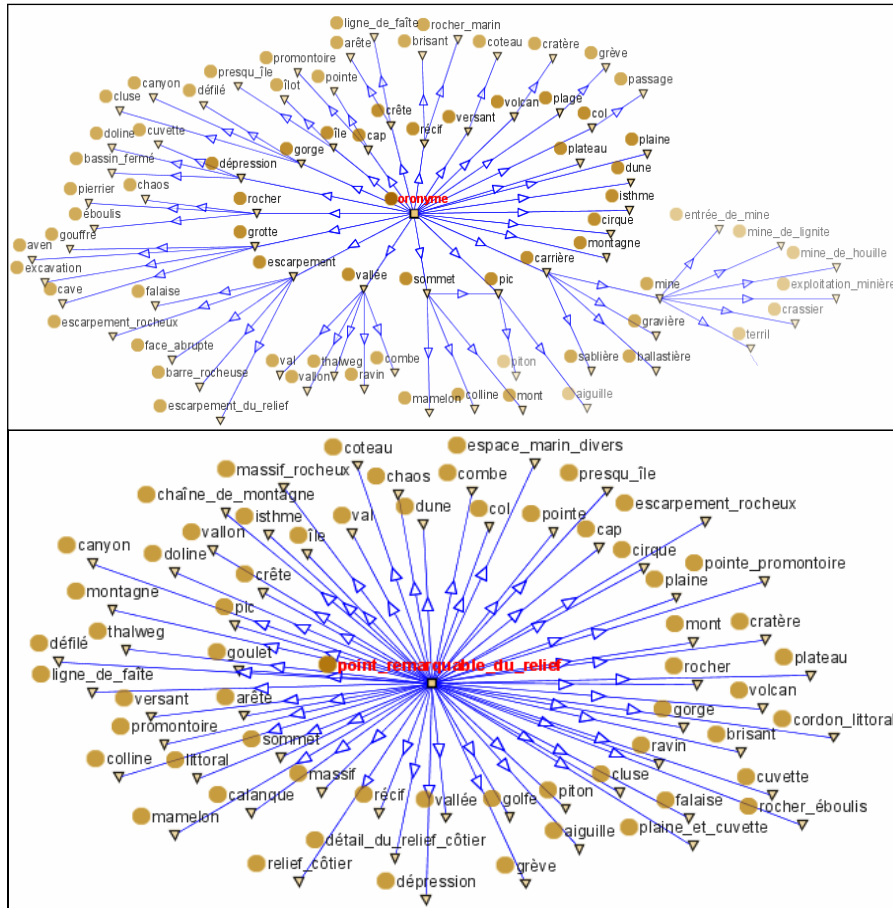


Figure 3. Taxonomies des éléments remarquables du relief obtenues par analyse automatique des spécifications de la BDTOPO® (haut) et BDCARTO® (bas)

Dans le cas de la BDTOPO®, la taxonomie obtenue a pu être comparée à un index créé manuellement et présent à la fin des spécifications. Celle-ci comprend 580 concepts au total, dont 380 sont également présents dans l'index. 80 concepts de l'index, en revanche, ne figurent pas dans notre taxonomie. Ceci s'explique en grande partie par le fait que cet index a été créé pour une version des spécifications antérieure à celle utilisée pour l'analyse automatique. Ainsi, une trentaine de termes comme « fortin », « bergerie », « kiosque » ou « râtelier paravalanche », figurant dans cet index, n'ont pu être détectés dans notre corpus de travail tout simplement parce qu'ils ne s'y trouvent pas. D'autres termes figurant à la fois dans l'index et dans le texte n'ont cependant pas été détectés. Il s'agit pour la plupart de groupes nominaux complexes, plus délicats à traiter comme « tremplin de saut à ski », ou de

concepts trop spécifiques pour avoir été retenus lors de l'étape de sélection interactive, comme celui de « rue piétonne » auquel on aura préféré le concept plus générique de « rue ». En revanche, 200 concepts ne figurant pas dans l'index ont été extraits du texte des spécifications, ce qui témoigne du bon niveau d'exhaustivité de l'ontologie obtenue, vis-à-vis du corpus de base.

3.3 Alignement des deux taxonomies obtenues

La hiérarchie des concepts des taxonomies obtenues en sortie de ces traitements étant inférée automatiquement à partir de celle des textes des spécifications, un texte bien structuré aboutira nécessairement à une hiérarchie de concepts plus élaborée. C'est le cas notamment de la taxonomie créée à partir des spécifications de la BDTOPO®. A l'inverse, l'organisation des spécifications de la BDCARTO® étant moins détaillée, la taxonomie qui en découle est presque totalement dénuée de structure hiérarchique. Par ailleurs, les deux taxonomies résultantes ne comportent que 170 concepts rigoureusement identiques. Il convient donc de fusionner ces deux taxonomies en conservant la structure de celle issue des spécifications de la BDTOPO®, et en s'appuyant sur les concepts communs aux deux taxonomies. Les concepts supplémentaires présents dans chacune des deux taxonomies d'origine devront également être intégrés au résultat de cet alignement.

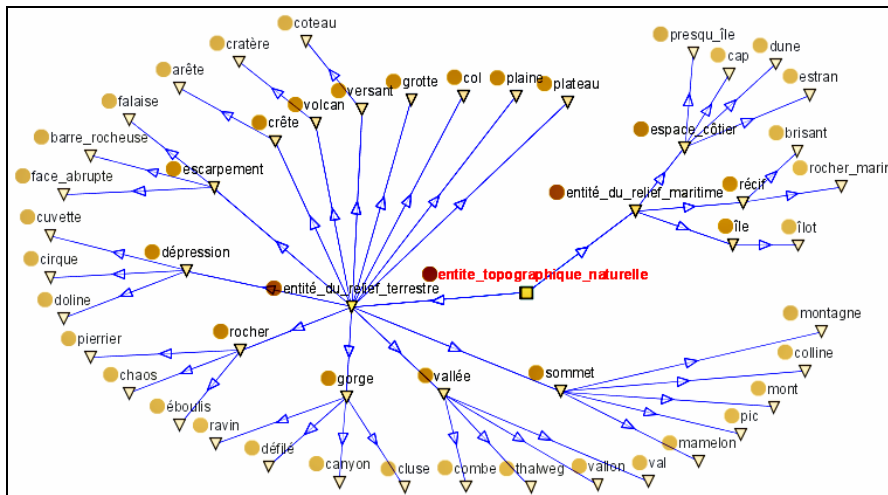


Figure 4. Extrait de la taxonomie obtenue après post-traitement manuel et alignement automatique des taxonomies présentées en Fig. 3 et 4

Cette opération de fusion des deux taxonomies obtenues a été réalisée à l'aide de Prompt, une surcouche de Protégé dédiée à l'alignement d'ontologies. Les résultats de cette opération tendent à prouver qu'un alignement automatique de ces taxonomies est possible, même s'il doit être affiné (voir Fig.4).

4. Exploitation de la taxonomie créée

Pour illustrer le potentiel de la taxonomie créée, nous en présentons ici quelques applications, en cours ou prévues au laboratoire COGIT.

Une première application est l'appariement de données [Olteanu 2007]. Les principales approches existant en appariement de données géographiques s'appuient essentiellement sur la comparaison des géométries et organisations topologiques des objets. La taxonomie que nous avons définie peut permettre d'affiner ce processus en comparant les natures des objets, c'est-à-dire en analysant les positions relatives dans l'ontologie des concepts décrivant cette nature [Abadie et al. 2007].

Une deuxième utilisation de notre taxonomie est la formalisation des spécifications des bases de données. Ces documents ont servi à créer la taxonomie, mais ils contiennent bien d'autres informations. On peut les voir comme l'expression des règles de passage des phénomènes du monde réel à des objets d'une base (comme par exemple « tous les bâtiments de plus de 100m² sont représentés dans la classe bâtiment par leur contour ») [Gesbert 2005, Gesbert et al. 2004]. Un modèle formel a été proposé pour représenter ces règles, dans lequel les spécifications sont modélisées comme un lien entre le schéma d'une base et une ontologie du domaine [Gesbert 2005]. Une fois notre taxonomie créée, ce modèle des spécifications a pu être instancié en très grande partie automatiquement, une fois encore grâce à des outils de traitement automatique du langage basés sur l'utilisation de patrons lexico-syntaxiques [Picard 2007].

Une fois les spécifications formalisées, on peut les exploiter pour définir des outils d'appariement de schémas fins. Illustrons cela par un exemple simple. Si les spécifications formalisées représentent le fait que dans une base « tous les aqueducs sont représentés dans la classe *canalisation* », et que dans une autre base « les aqueducs de plus de 100 m de long sont représentés dans la classe *tronçon hydrographique* avec l'attribut *nature* égal à 'aqueduc' », on peut déduire que ces deux classes sont en correspondance et comment elles le sont, c'est-à-dire quelles contraintes respectent les objets en correspondance : pour notre exemple, tout objet de la classe *tronçon hydrographique* avec l'attribut *nature* égal à 'aqueduc' doit avoir un homologue dans la classe *canalisation*. Dans les travaux passés du laboratoire COGIT, un modèle d'expression des correspondances entre schémas de données géographiques a été défini [Devoegele 1997]. L'analyse des correspondances a également été exploitée pour identifier les incohérences entre deux jeux de données [Sheeren 2005]. Actuellement, une thèse débute pour définir une méthodologie complète d'appariement de schémas en exploitant les spécifications formelles.

Enfin, une dernière application envisagée à court terme est une interface intuitive d'exploration du contenu des bases de données de l'IGN. L'utilisateur devrait y être capable d'exprimer son besoin en termes relativement intuitifs, en s'appuyant sur les concepts de l'ontologie. Le système devrait ensuite être capable de spécifier à l'utilisateur dans quelles classes des bases de données disponibles est représentée l'information qu'il recherche, et comment elle y est représentée.

5. Conclusion

Nous avons présenté différents travaux qui nous ont permis de créer, puis d'exploiter, une taxonomie de termes du domaine de la topographie. Comme nous l'avons exprimé auparavant, cette taxonomie nous semble un premier pas nécessaire vers la création d'une ontologie topographique plus riche. Pour être plus intensément exploitée, notre taxonomie mériterait en effet d'être enrichie, à la fois de nouveaux concepts, mais aussi de définitions et propriétés décrivant ces concepts, et enfin d'une description fine des relations entre ces concepts. Afin de réaliser et exploiter une telle ontologie, nous participons à un projet de recherche nommé GeOnto (ANR-07-MDCO-005). Ce projet vise à créer des ontologies topographiques riches, ainsi qu'à mettre au point des méthodologies de comparaison, enrichissement, fusion et exploitation de ces ontologies.

6. Bibliographie

- Abadie N., Olteanu A.-M. et Mustière S., Comparaison de la nature d'objets géographiques, dans les actes de la conférence Ingénierie des Connaissances, journée « Ontologies et Gestion de l'Hétérogénéité Sémantique », Grenoble, 3 juillet 2007, pp.59-68.
- Abadie N. et Mustière S. 2008. *Constitution d'une taxonomie géographique à partir des spécifications de bases de données*. Proc. of SAGEO'2008, Montpellier, June 2008.
- Aerts K., Maesen K. et van Rompaey A., A practical example of semantic interoperability of large-scale topographic databases using semantic web technologies. *9th AGILE Conference on Geographic Information Science*. College of Geoinformatics, University of West Hungary, pages 35-42, 2006.
- Agarwal P., Ontological considerations in GIScience. *International Journal of Geographic Information Science*, vol 19, n° 5, pages 501-536, Taylor & Francis, 2005.
- Bittner, T., Donnelly, M. et Winter S., Ontology and Semantic Interoperability. *Large-scale 3D Data Integration: Challenges and Opportunities*, Prospery, D. & Zlatanova, S. (dir.), CRC press (Taylor & Francis), pages 139 - 160, 2005.
- Brisson R., Boussaïd O., Gançarski P., Puissant A., Durant N., Navigation et appariement d'objets géographiques dans une ontologie. In *EGC '07.*, p. 391-396. 2007.
- Cullot N., Parent C., Spaccapietra S., et Vangenot C., Des SIG aux ontologies géographiques. *Revue internationale de Géomatique*, vol 13, n° 3, pages 285-306, Lavoisier, 2003.

- Devoegele T., Processus d'intégration et d'appariement des bases de données géographiques. Application à une base de données routière multi-échelles. Thèse de l'Université de Versailles, IGN/Laboratoire COGIT. 1997.
- Egenhofer M.J., Toward the Semantic Geospatial Web. *Proceedings of the 10th international symposium on Advances in Geographic Information Systems*, McLean, Virginia, A. Voisard et S.-C. Chen (dir.), pages 1-4, 2002.
- Gesbert N., Etude de la formalisation des spécifications de bases de données géographiques en vue de leur intégration. Thèse de doctorat, Univ. Marne-la-Vallée, 2005.
- Gesbert N., Libourel T. et Mustière S., Apport des spécifications pour les modèles de bases de données géographiques. *Revue internationale de Géomatique*, vol 14, n° 2, pp. 239-257, Lavoisier, 2004.
- Gruber T.R., Toward principles for the design of ontologies used for knowledge sharing. *Formal ontology in conceptual analysis and knowledge representation*. N. Guarino et R. Poli (dir.). Dordrecht: Kluwer academic, 1993.
- Guarino N., Formal ontology and information systems. *Formal ontology in information systems: proceedings of FOIS'98, Trento, Italy, 6-8 Juin 1998*. N. Guarino (dir.) Amsterdam: IOS Press, pages 3-15, 1998.
- Laurens F. 2006. Création d'une ontologie à partir de textes en langage naturel. Internship report, Master 1 Linguistique-Informatique, University Paris 7.
- Mustière S., Abadie N. et Laurens F. Appariement de schémas de BD géographiques à l'aide d'ontologies déduites des spécifications. In *EGC '07.*, p. 22-27. 2007.
- Olteanu A.-M., A multi-criteria fusion approach for geographical data matching. *Proc. of the 5th International Symposium on Spatial Data Quality (ISSDQ)*, 13-15 Juin, Enschede, Pays-Bas, 2007.
- Olteanu Raimond A.M., Mustière S., A fuzzy approach for matching networks. En cours de soumission à SDH'2008, 2008.
- Partridge C., The role of ontology in integrating semantically heterogeneous databases. *Rapport technique 05/02 LADSEB-CNR*, Padoue, 2002.
- Picard V., Instanciation automatique des liens entre ontologies et schémas de bases de données géographiques à partir des spécifications en langage naturel. *Stage de Master 2 Document Electroniques et Flux d'Information, Université Paris 10 – Nanterre*, 2007.
- Rodriguez M.A. et Egenhofer M.J., Determining Semantic Similarity Among Entity Classes from Different Ontologies. *IEEE Transactions on Knowledge and Data Engineering*, vol. 15, n° 2, pages 442-456, 2003.
- Roussey C., Laurini R., Beaulieu C., Tardy Y. et Zimmermann M., Le projet Towntology : Un retour d'expérience pour la construction d'une ontologie urbaine. *Revue internationale de Géomatique*, vol 14, n° 2, pp. 217-237, Lavoisier, 2004.
- Sheeren D., Méthodologie d'évaluation de la cohérence inter-représentations pour l'intégration de bases de données spatiales – Une approche combinant l'utilisation de métadonnées et l'apprentissage automatique. Thèse de doctorat, Université Paris 6. 2005.

Sheeren D., Mustière S. et Zucker J.-D., Consistency Assessment Between Multiple Representations of Geographical Databases: a Specification-Based Approach. *Proceedings of the 11th International Symposium on Spatial Data Handling (SDH'04)*, pages 617-627, Leicester: Springer, 2004.

Smith B. et Mark D.M., 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.