

OntoGéo 2010

Atelier « *Ontologies Géographiques* », dans le cadre de
Sagéo'10 « *Spatial Analysis and GEomatics* »

Toulouse, 18 novembre 2010

Atelier organisé dans le
cadre du projet ANR
GeOnto

Actes électroniques

Objectifs et thématiques
Comité Scientifique
Programme
Résumés des articles
Résumés des démonstrations
Liste des articles et des résumés de démonstrations

Objectifs et thématiques

La description et l'intégration cohérente des données géographiques se révèle très difficile car ces données sont très nombreuses et diverses. Une approche de plus en plus privilégiée pour réaliser cette intégration, autant dans le monde des bases de données que dans celui de la recherche d'information et des systèmes d'information géographiques, est d'appuyer l'intégration des données sur une ontologie du domaine concerné.

Dans cet atelier, nous proposons ainsi de faire un panorama des recherches en cours sur les ontologies géographiques et de présenter les résultats obtenus dans le cadre de projets de recherche menés dans cette thématique. Cet atelier se veut un lieu de rencontre et d'échanges de façon à favoriser le partage d'expérience et d'expertise concernant la construction et l'exploitation d'ontologies géographiques et initier d'éventuelles coopérations.

Quatre présentations orales, suivies de questions et discussions ont été programmées pour cet atelier d'une demi-journée : un exposé d'un conférencier invité et trois exposés d'articles acceptés. Une session démonstrations de résultats de recherche fait suite à ces présentations.

Le processus de sélection a été réalisé sur la base de soumissions de papiers de 12 pages maximum pour les articles et de résumés pour les démonstrations. Chaque article a fait l'objet de deux relectures.

L'atelier est organisé dans le cadre du projet ANR GeOnto, qui a financé les frais de transport d'un des auteurs de chaque papier et démonstration et qui a permis l'invitation d'Oscar Corcho-Garcia de l'Université Polytechnique de Madrid.

Nous adressons tous nos remerciements aux auteurs des articles et résumés soumis ainsi qu'à tous les membres du comité de programme.

Liste (non exhaustive) des thèmes :

- Construction d'ontologies géographiques
- Enrichissement, évolution d'ontologies géographiques
- Ontologies géographiques et interopérabilité
- Exploitation d'ontologies géographiques pour la recherche d'information, l'indexation de documents, dans le cadre de l'apprentissage humain, ...

Comité d'organisation

- Chantal Reynaud (LRI Univ. Paris-Sud 11 & CNRS, INRIA Saclay Île-de-France) : Présidente
- Nathalie Abadie (Lab. COGIT, IGN)
- Fayçal Hamdi (LRI Université Paris-Sud 11 & CNRS, INRIA Saclay Île-de-France)
- Ammar Mechouche (Lab. COGIT, IGN)
- Yassine Mrabet (LRI Univ. Paris-Sud 11 & CNRS, INRIA Saclay Île-de-France)

Comité de programme

- Nathalie Abadie (Lab. COGIT, IGN)
- Nathalie Aussenac-Gilles (IRIT, Université Paul Sabatier)
- Alain Bouju (Lab. Informatique, Image, Interaction, Université de La Rochelle)
- Mauro Gaio (LIUPPA, Université de Pau et des Pays de l'Adour)
- Jérôme Gensel (LIG, Université Pierre Mendès France, Grenoble)
- Fayçal Hamdi (LRI Université Paris-Sud 11 & CNRS, INRIA Saclay Île-de-France)
- Eric Kergosien (LIUPPA, Université de Pau et des Pays de l'Adour)
- Marion Laignelet (IRIT, Université Paul Sabatier)
- Thérèse Libourel (Laboratoire d'Informatique, Université de Montpellier II)
- Ammar Mechouche (Lab. COGIT, IGN)
- Yassine Mrabet (LRI Univ. Paris-Sud 11 & CNRS, INRIA Saclay Île-de-France)
- Sébastien Mustière (Lab. COGIT, IGN)
- Van Tien Nguyen (LIUPPA, Université de Pau et des Pays de l'Adour)
- Chantal Reynaud (LRI Univ. Paris-Sud 11 & CNRS, INRIA Saclay Île-de-France)
- Raphaël Troncy (EURECOM, Sophia Antipolis)
- Christelle Vangenot (Université de Genève, Suisse)

Programme

18 novembre, 2010

14h30-15h30	Conférence invitée, <i>Oscar Corcho</i>	
	Ontology-based data integration in the geospatial domain: a case study in Spain. <i>Oscar Corcho, Universidad Politécnica de Madrid</i>	p. 8
15h30-17h00	Présentations d'articles	
15h30-16h00	Recherche automatique de concepts et de relations sémantiques dans des documents géographiques. <i>Marion Laignelet, Mouna Kamel, Nathalie Aussenac- Gilles</i>	p.9
16h00-16h30	Une approche basée sur les ontologies et les agents pour l'interopérabilité des SIG <i>Alnafie Emdjed, Hamadou Djamila</i>	p. 24
16h30-17h00	AOC, une ontologie formelle pour la modélisation de systèmes complexes en géographie <i>Patrice Langlois, Eric Daudé, Baptiste Blanpain, Emmanuel Sapin</i>	p.36
17h00-17h30	Session démonstrations	
17h00-17h15	TaxoMap Framework appliqué à l'alignement d'ontologies géographiques au sein du projet Geonto <i>Fayçal Hamdi, Chantal Reynaud, Brigitte Safar</i>	p. 51
17h15-17h30	Vers un système en ligne d'aide à la découverte du contenu des bases de données géographiques <i>Ammar Mechouche, Nathalie Abadie, Emeric Prouteau, Sébastien Mustière.</i>	p. 54

Résumés des articles

Ontology-based data integration in the geospatial domain: a case study in Spain.

Oscar Corso, Universidad Politécnica de Madrid, Madrid, Espagne

In this talk we will present the work that is being carried out under the geo.linkeddata.es initiative, where geographical data provided by the Spanish Instituto Geográfico Nacional is being migrated to the Web of Linked Data. Methodological and technological details about the migration process will be provided, together with a discussion on guidelines for naming geographical features, generation and management of feature geometries, integration of heterogeneous geographical and statistical databases, identification of duplicated individuals, and best practices on the usage of these data in applications. Finally a list of open research topics will be identified and discussed, including aspects related to the maintenance of database to ontology mappings, to the management of multilinguality and to provenance recording.

Recherche de concepts et de relations sémantiques pour la géographie. Traiter automatiquement des documents de spécification de bases de données.

Marion Laignelet^{}, Mouna Kamel^{**}, Nathalie Aussenac-Gilles^{**}*

^{}Maison de la recherche – Université de Toulouse le Mirail, Toulouse, France*

*^{**}Institut de Recherche en informatique de Toulouse (IRIT) – CNRS, Toulouse, France*

Basée sur une méthodologie exploitant la notion de marqueur, nous proposons de repérer des marques linguistiques de concepts et de relations sémantiques en utilisant non seulement des informations de type syntaxique mais également la structure hiérarchique des documents. Nous travaillons sur des documents de spécification de bases de données géographiques dont la structure et l'organisation, du fait notamment de leur fonction, est sémantiquement explicite. Matérialisée par un balisage XM, cette structure permet de construire un noyau d'ontologie qui est ensuite enrichi par des connaissances extraites à l'aide de patrons combinant analyse lexico-syntaxique et analyse structurelle.

Une approche basée sur les ontologies et les agents pour l'interopérabilité des SIG

Emdjed Alnafie^{}, Djamilia Hamdadou^{*}*

^{}Université d'Oran Es-senia, laboratoire LIO, Algérie*

Le domaine de l'information a connu un progrès extraordinaire permettant à de nombreuses spécialités et disciplines, de partager des sources d'information hétérogènes et de grandes masses. Cette communication, en particulier entre les systèmes d'information géographique (SIG), est malheureusement confrontée au problème de l'hétérogénéité des sources d'information. L'interopérabilité des SIGs est la solution qui permet de créer un unique et homogène système « virtuel » avec un accès simple, transparent et efficace à l'information nécessaire. Cet accès est assuré quelle que soit la localisation des SIGs ou de leurs méthodes de stockage. L'objectif principal de cette étude, et dont les contributions seront en mesure de faciliter la prise de décision des gestionnaires concernés, est de proposer une approche fondée

sur l'utilisation combinée des géo-ontologies et des agents. Cette approche adopte une politique de traitement permettant de trouver rapidement des réponses aux requêtes tout en garantissant un accès direct à ces réponses dans le cas d'une prochaine éventuelle requête de même nature.

AOC, une ontologie formelle pour la modélisation de systèmes complexes en géographie

*Patrice Langlois, Eric Daudé, Baptiste Blanpain, Emmanuel Sapin
Université de Rouen, Mont-Saint-Aignan, France*

Cet article présente un méta-modèle nommé AOC (Agents-Organisations-Comportements), qui s'exprime par une ontologie formelle et sert de cadre conceptuel à une plateforme de simulation de systèmes complexes spatialisés, compatible avec les problématiques de dynamiques géographiques les plus larges. Le modélisateur géographe doit pouvoir concevoir son modèle et le formaliser dans un langage cohérent relativement à son cadre de pensée. AOC offre ainsi la possibilité de structurer les relations entre agents via le concept d'organisation, qui permet de créer des modèles multi-niveaux. De plus, une formalisation générale de la notion de comportement permet de modéliser la dynamique d'une classe très large de systèmes.

Résumés des démonstrations

TaxoMap Framework appliqué à l'alignement d'ontologies géographiques dans le projet GéOnto.

Fayçal Hamdi, Chantal Reynaud, Brigitte Safar

Les outils actuels d'alignement d'ontologies ne sont pas performants dans tous les domaines ni quelles que soient les ontologies. La qualité des mappings obtenus peut être améliorée si le processus d'alignement tient compte des spécificités des ontologies alignées. L'environnement TaxoMap Framework a été conçu pour aider un expert du domaine à spécifier des traitements sur les mappings produits par l'outil d'alignement TaxoMap, afin de les raffiner ou bien de fusionner, restructurer ou enrichir des ontologies. Nous proposons une démonstration de deux modules implémentés dans cet environnement : le module d'alignement et le module de raffinement.

Vers un système en ligne d'aide à la découverte du contenu de bases de données géographiques.

Ammar Mechouche, Nathalie Abadie, Emeric Prouteau, Sébastien Mustière

Nous présentons ici un prototype d'une application web proposant à un utilisateur quelconque de pouvoir découvrir le contenu de bases de données géographiques, en lui permettant de savoir comment et où sont représentées les données qui répondent à son besoin. Cette application s'appuie sur une ontologie du domaine décrivant les entités topographiques du monde réel et permettant de guider l'utilisateur dans l'expression de son besoin. Elle s'appuie également sur des ontologies formalisant les spécifications des bases de données géographiques, et inclut une solution de cartographie en ligne pour un aperçu des données.

Ontology-based data integration in the geospatial domain: a case study in Spain.

Oscar Corcho

*Universidad Politécnica de Madrid
28660 Boadilla del Monte, Madrid, Spain*

ABSTRACT. In this talk we will present the work that is being carried out under the geo.linkeddata.es initiative, where geographical data provided by the Spanish Instituto Geográfico Nacional is being migrated to the Web of Linked Data. Methodological and technological details about the migration process will be provided, together with a discussion on guidelines for naming geographical features, generation and management of feature geometries, integration of heterogeneous geographical and statistical databases, identification of duplicated individuals, and best practices on the usage of these data in applications. Finally a list of open research topics will be identified and discussed, including aspects related to the maintenance of database to ontology mappings, to the management of multilinguality and to provenance recording.

Recherche de concepts et de relations sémantiques pour la géographie

Traiter automatiquement des documents de spécification de bases de données

Marion Laignelet*, Mouna Kamel**, Nathalie Aussenac-Gilles**

* *CLLE-ERSS - UMR 5610*

Maison de la Recherche - Université de Toulouse-Le Mirail

5, allée Antonio-Machado, F-31058 Toulouse Cedex 9

marion.laignelet@univ-tlse2.fr

** *Institut de Recherche en Informatique de Toulouse (IRIT) - CNRS*

UPS, 118, route de Narbonne, F-31062 Toulouse Cedex

mouna.kamel@irit.fr

aussenac@irit.fr

RÉSUMÉ. Basée sur une méthodologie exploitant la notion de marqueur, nous proposons de repérer des marques linguistiques de concepts et de relations sémantiques en utilisant non seulement des informations de type syntaxique mais également la structure hiérarchique des documents. Nous travaillons sur des documents de spécification de base de données géographiques dont la structure et l'organisation, du fait notamment de leur fonction, est sémantiquement explicite. Matérialisée par un balisage XML, cette structure permet de construire un noyau d'ontologie qui est ensuite enrichi par des connaissances extraites à l'aide de patrons combinant analyse lexico-syntaxique et analyse structurelle.

ABSTRACT. Based on the notion of pattern, our methodology aims at tracking linguistic evidences of concepts and semantic relations by using not only syntactic information but also the hierarchical structure of documents. The documents that we analyse are geographical and cartographic databases. Due to their role, their structure is semantically meaningful. This structure is implemented with XML tags, which makes it possible to automatically build an ontology kernel. This kernel is enriched with pieces of knowledge that are extracted thanks to patterns that combine lexico-syntactic and structural features.

MOTS-CLÉS : ontologie, géographie, TAL, patrons lexico-syntaxiques, structure de document, repérage automatique de concepts et relations

KEYWORDS: ontology, geography, NLP, lexicale and syntactical patterns, document structure, automatic tracking of concepts and relations

1. Introduction

Nous proposons de trouver en corpus des traces linguistiques de concepts et de relations sémantiques dans le but de construire une ontologie géographique à partir de documents de spécifications de bases de données. La méthode que nous proposons exploite non seulement des patrons lexico-syntaxiques [BOU 03, BOU 04, AUS 09, AUS 08] mais également des patrons basés sur la structure hiérarchique des documents et matérialisée dans nos textes par des balises XML. La mise en forme des documents de spécification de base de données géographiques de l'IGN, du fait notamment de leur fonction descriptive et explicative, est sémantiquement explicite et permet de construire un noyau d'ontologie. Ce noyau est ensuite enrichi par des connaissances analysées et extraites à l'aide de patrons plus locaux, combinant analyse lexicale, syntaxique et structurelle.

Ce travail s'inscrit dans le cadre d'un projet de recherche, le projet Géonto, dont l'objectif est de capitaliser des connaissances hétérogènes à partir de documents géographiques diversifiés. D'un point de vue méthodologique, l'objectif du projet est d'éprouver et d'adapter des outils et méthodes existants pour créer des ontologies à partir de données et de connaissances imprécises et hétérogènes. Du point de vue applicatif, les ontologies créées doivent permettre un accès facilité à de grandes masses de données géographiques. Parallèlement, nous espérons par cette méthode appréhender les différences de point de vue sous-jacentes au domaine.

Dans cet article, nous présentons la méthode que nous avons définie pour exploiter au mieux les documents de spécification de bases de données géographiques de manière à constituer une ontologie reflétant la structuration des connaissances définie par ces spécifications. L'originalité de cette méthode est de rechercher des connaissances en exploitant non seulement le langage naturel présent dans ces textes, mais aussi les indications explicites de structure qui encadrent les zones en langage naturel. Dans la section 2, nous présentons un panorama des différentes méthodes de construction d'ontologies à partir de texte, que ce soit pour la recherche des concepts ou celle des relations entre les concepts. La section 3 présente le projet Géonto et la problématique d'acquisition de connaissances pour la géographie et la cartographie. La section 4 décrit l'approche mise en oeuvre ainsi que les outils développés pour extraire les concepts potentiellement pertinents pour le domaine de la géographie et les relations sémantiques susceptibles de relier ces concepts. Enfin, dans la dernière section, nous évaluons quantitativement et qualitativement la méthode mise en oeuvre.

2. Construire des ontologies à partir de textes en langue naturelle

La construction d'ontologies, et plus généralement de ressources termino-ontologiques (RTO), à partir de textes nécessite la mise en place de méthodologies complexes exploitant massivement des outils du traitement automatique des langues [BOU 04, JAC 06, SCH 05, MAY 09, ROL 06]. L'utilisation de techniques de TAL permet notamment d'extraire les éléments qui vont constituer la RTO, ou qui vont servir d'indices à l'ontographe pour définir des éléments de RTO. Selon le type de RTO visé (hiérarchie de termes, thé-

saurs, hiérarchie de concepts, ontologie plus ou moins riche en relations) les besoins sont variables. Mais dans tous les cas, il est nécessaire d'extraire des termes, lesquels sont potentiellement aptes à être considérés par la suite comme des étiquettes de concepts pour le domaine en question.

2.1. Méthodologies basées sur la notion de marqueur ou patrons

L'utilisation de patrons pour extraire des connaissances est largement utilisée. [HEA 92] fait figure de référence en ce qui concerne les techniques de linguistique computationnelle. Meyer (2001) introduit la notion de patrons de connaissances (*knowledge rich contexts*) pour faire référence à patrons linguistiques exprimant diverses relations sémantiques liées à la définition, telles l'hyponymie, la synonymie ou la méronymie. [HAD 02] définit la notion de marqueur comme une « *forme linguistique faisant partie de catégories prédéfinies (grammaticales, lexicales, syntaxiques ou sémantiques) dont l'interprétation définit régulièrement le même rapport de sens entre les termes.* » Et, dans le contexte de la recherche de relations sémantiques en corpus, [GRA 04] définissent le patron lexicosyntaxique de la manière suivante : « *à la différence des marqueurs, les patrons identifient la relation recherchée plus précisément en définissant également des contraintes syntaxiques ou typographiques sur le contexte des termes.* » Dans tous les cas, un patron définit des contextes suffisamment fins et précis, relevant de niveaux linguistiques variables et révélateurs de connaissances ou de comportements linguistiques particuliers.

Dans le cas de la recherche de relations sémantiques, un patron permet d'indiquer l'existence d'une relation sémantique particulière entre deux entités (au moins). Appliquée à la construction d'ontologies, une telle approche fait l'hypothèse que les relations lexicales peuvent fournir des indices pour définir des relations conceptuelles et, avec elles, de nouveaux concepts et termes associés. [BOU 04] mettent en place une méthodologie fondée sur l'utilisation de patrons. Une première phase d'extraction de termes¹ construit un réseau de mots et de syntagmes (le « réseau terminologique ») dans lequel chaque syntagme est relié à sa tête et à ses expansions. À ce niveau, les éléments du réseau sont des candidats-termes potentiellement aptes à devenir des concepts. La seconde étape est une analyse distributionnelle² : à partir du réseau terminologique, un calcul des proximités distributionnelles entre les unités est effectué. La dernière phase extrait les relations³. Des outils de modélisation⁴ permettent *in fine* de définir les concepts à retenir. Dans cette approche, il est important de noter que les termes sont les occurrences des mots du corpus et les relations sont extraites à l'aide de patrons lexico-syntaxiques. D'autres travaux mettent en oeuvre cette méthodologie ([JAC 06]).

1. À l'aide du logiciel Syntex, analyseur syntaxique.

2. À l'aide du logiciel Upery.

3. À l'aide de Yakwa (concordancier) et Caméléon (logiciel de recherche de relations lexicales à partir de marqueurs linguistiques).

4. TermOnto, Terminae

La question des patrons lexico-syntaxiques pour la création et la population d'ontologies est également traitée dans [MAY 09]. Ces auteurs proposent la notion de patrons lexico-syntaxiques (*lexico-syntactic ontology design patterns*, ODPs) qui génèrent de l'information ontologique à partir de textes non-structurés dans le but de créer une nouvelle ontologie ou enrichir des ontologies existantes.

2.2. Les verbes pour exprimer les relations

Au coeur d'un patron lexico-syntaxique traduisant une relation lexicale, se trouve souvent un verbe comme indice fort de cette relation. Par exemple, *comporter*, *contenir*, *se composer de* peuvent marquer la relation d'un tout vers ses parties. [SCH 05] insiste sur le rôle des verbes comme éléments centraux pour déterminer une relation entre deux concepts d'une ontologie. Pour les auteurs, les verbes spécifient l'interaction entre les participants d'une action ou d'un événement. Leur système identifie des triplets (*i.e.* des paires de concepts reliés par une relation) qui pourront être intégrés à une ontologie existante. Cet outil extrait les termes pertinents et les verbes des textes à l'aide d'une approche combinant des techniques linguistiques et statistiques. L'approche mise en place est de type linguistique. Les structures de dépendance sont reconnues : fonction grammaticale, structure des phrases, étiquetage morpho-syntaxique et lemmatisation, normalisation des syntagmes nominaux complexes sur la base de leur tête nominale. Une reconnaissance des entités nommées et des concepts propres au domaine (le foot) est également menée (utilisation de l'ontologie du domaine) ainsi que la reconnaissance des synonymes. Enfin, un processus statistique fournit des mesures de pertinence et des mesures de cooccurrence. L'approche se veut robuste et adaptable à d'autres domaines.

2.3. Utilisation de la structure des documents

L'idée d'exploiter la structure des documents pour améliorer le processus d'acquisition des connaissances est proposée par [ROL 06]. Les auteurs partent du constat suivant lequel les documents ont la plupart du temps une structure logique cohérente et porteuse de sens. C'est le cas, par exemple, des manuels techniques, des dictionnaires, des codes juridiques, etc. Généralement, cette structure se matérialise en XML. Leur objectif est de construire une ontologie des plantes tropicales en s'appuyant sur la structure du document⁵ à deux niveaux, d'un côté pour obtenir une hiérarchie de classes reflétant la taxinomie botanique traditionnelle (famille, genre, espèce), d'un autre pour cibler plus précisément les traitements linguistiques à effectuer pour compléter la hiérarchie de classes par des indications méronymiques. Les auteurs insistent sur le fait que la botanique est un domaine qui se prête naturellement à la représentation de taxinomies : le découpage en genres et en espèces est représenté dans la structure du document et permet ainsi d'initialiser la hiérarchie de classes de l'ontologie. Le résultat produit consiste en

5. La *Flore du Cameroun* qui représente 40 volumes publiés entre 1963 et 2001. La structure des différents volumes est assez régulière, soit une fiche par espèce.

une première ontologie intermédiaire⁶ qui facilite les traitements linguistiques ultérieurs plus complexes pour créer une ontologie du domaine. D'autres travaux comme [DES 00] utilisent également la structure des documents pour construire des ontologies du domaine.

Dans tous les cas, construire des RTO à partir de textes n'est pas une tâche triviale et l'importance du domaine, de la qualité du corpus ou encore de la finalité de la ressource sont essentiels.

3. Acquisition de connaissances dans le domaine de la géographie-cartographie

Les données géographiques sont à la fois nombreuses et diverses. Par exemple, l'INSEE propose une ontologie en français⁷ qui décrit les événements liés à la construction ou la modification des territoires ; le projet TOWNTOLGY⁸ développe une ontologie dédiée à l'enseignement de l'urbanisme. A notre connaissance, il n'existe pas d'ontologie dédiée aux objets cartographiques, en français du moins⁹. Dans ce contexte, le projet Géonto, dans lequel cet article s'inscrit, propose des méthodes et des outils permettant la description et l'intégration cohérente des données géographiques [MUS 09, KER 09]. Pour répondre à cet objectif, nous proposons la construction d'une ontologie du domaine qui doit permettre l'intégration de sources d'information multiples et hétérogènes. La méthodologie mise en place suit les étapes suivantes :

- création d'une première ontologie (le noyau),
- enrichissement de cette ontologie à partir des spécifications de bases de données topographiques de l'IGN.

Ce qui nous intéresse plus particulièrement concerne la conception et l'enrichissement du noyau d'ontologie. Le noyau de l'ontologie est obtenu à partir de l'exploitation de la structure des documents de spécification des bases de données. L'enrichissement se fait à partir de l'analyse de la structure et de l'analyse linguistique de zones textuelles spécifiques, les zones de définition.

Dans les figures suivantes, un extrait du corpus est présenté, d'abord sous leur format original puis en XML.

D'une manière générale et comme cela a été dit dans la section 2, l'extraction de connaissances à partir de textes nécessite la constitution d'un corpus. Il s'agit d'une étape importante car, dans une approche automatisée, le corpus est la source essentielle d'information : la taille du corpus doit être suffisamment importante pour fournir une couverture large du domaine et il doit être suffisamment homogène pour permettre le repérage de régularités. Par ailleurs, [BOU 04] affirment qu'il est impossible de définir *a priori* des instructions méthodologiques précises car le processus de construction reste lié à l'application et aux exigences des spécialistes du domaine.

6. Dans notre travail, nous parlons de noyau d'ontologie.

7. <http://www.insee.fr/fr/methodes/default.asp?page=xml/xml.htm>

8. <http://liris.cnrs.fr/townto/>

9. La situation pour les ontologies en anglais est différente.

A – Voies de Communication Routière

Tronçon de Chemin

Définition : Voie de communication terrestre non ferrée destinée aux piétons, aux cycles ou aux animaux...	
Regroupement : Voir les différentes valeurs de l'attribut <nature>.	
Sélection : Voir les différentes valeurs de l'attribut <nature>.	
Modélisation géométrique : A l'axe, au sol.	
<hr/>	
Attribut : Nature	
Définition :	Permet de distinguer plusieurs types de voies de communication terrestres.
Type :	Énuméré
Valeurs :	Chemin empierré / Chemin / Sentier / Escalier / Piste cyclable
<hr/>	
Nature = « Chemin empierré »	
Définition : Route sommairement revêtue ou chemin empierré (pas de revêtement de surface ou revêtement très dégradé), mais permettant la circulation de véhicules automobiles de tourisme par tous temps.	
Regroupement : Allée (carrossable) Piste Route empierrée	
Sélection : Toutes les routes empierrées sont incluses.	
...	
<hr/>	
Attribut : Franchissement	
Définition :	Attribut indiquant la présence d'un obstacle physique dans le tracé d'une route et la manière dont il est franchissable.
Type :	Énuméré
Valeurs :	Bac piéton / Gué ou radier / Pont / Tunnel / Sans objet
<hr/>	
Franchissement = « Gué ou radier »	
Définition : Passage naturel ou aménagé permettant de traverser un cours d'eau sans avoir recours à un pont ou un bateau.	
Regroupement : Gué Radier	
...	
<hr/>	
Attribut : Nom	
Définition :	Nom du chemin.
Type :	Caractères
Valeur nulle :	Le champ contient la chaîne de caractères "Valeur non renseignée" pour tous les chemins n'appartenant pas à un grand itinéraire routier nommé (référence = BDCart0).

Figure 1. Un extrait du corpus : sa mise en forme

```

<?xml version="1.0" encoding="ISO-8859-1" ?>
<document> <domaine>
  <nom_domaine>A - Voies de communication routière</nom_domaine>
  <classe>
    <nom_classe>Tronçon de chemin</nom_classe>
    <definition>Voie de communication terrestre ... </definition>
    ...
    <regroupement> Voir les différentes valeurs de l'attribut &lt;nature&gt;
  </regroupement>
  ...
  <attributs>
    <attribut>
      <nom_attribut> Nature </nom_attribut>
      <definition> Permet de distinguer plusieurs... </definition>
      <type> Énuméré</type>
      <valeurs> Chemin empierré/Chemin/Sentier/Escalier ... </valeurs>
      <attribut_valeur>
        <valeur> Chemin empierré </valeur>
      <definition> Route sommairement revêtue ... </definition>
      <regroupement> Allée (carrossable)|Piste|Route empierrée
    </regroupement>
  </attribut_valeur>
    </attribut>
  </classe></domaine> </document>

```

Figure 2. Un extrait de la version XML du corpus

La construction d'une ontologie (dite de référence) dans le projet Géonto a pour but de faire collaborer des données géographiques hétérogènes. Dans ce contexte, le sens des termes et des concepts associés est susceptible de varier selon les sous-domaines de la géographie. Par exemple, le terme de grotte peut être relié à des concepts différents selon que l'on prend le point de vue d'une carte routière ou celui d'une carte de randonnée : dans un cas, "grotte" est relié au concept "curiosité touristique" par une relation d'hyponymie alors que dans un contexte de randonnée, une relation d'hyponymie avec le concept "cavité souterraine" serait sans doute plus appropriée, peut-être même en association avec l'idée de danger. Dans notre corpus, l'exemple du pont est plus concret : le "pont" peut être une sorte de "franchissement" lorsqu'un objet passe par dessus un autre ou une sorte d'"obstacle" lorsqu'il passe par dessous. Dans ce contexte, il est nécessaire de recourir à un corpus suffisamment large et diversifié pour être capable de mettre au jour l'ensemble des points de vue possibles pour un même objet.

Les documents sur lesquels nous travaillons sont des documents de spécification de bases de données géographiques fournies par le COGIT (IGN). Nous disposons de deux textes différents, BDCarto et BDTopo. L'ensemble de ces deux corpus constitue un corpus de 23 884 mots (17 069 pour BDCarto et 6 815 pour BDTopo). Ces textes sont en XML. Les connaissances exprimées dans ces documents est cependant relativement variée puisqu'ils décrivent les objets nécessaires à la création de cartes, qu'elles soient touristiques, routières, pédestres, etc. Des points de vue différents sur les objets manipulés dans le domaine sont ainsi mis en valeur, notamment dans les zones de définition¹⁰. En plus d'un travail sur la structure logique de ces textes (notamment à travers l'organisation des titres et des sous-titres) qui fournit des indications de relation d'hyponymie, nous proposons une étude approfondie des zones de définition. À ce propos, [AGU 09] constatent que dans un contexte définitoire, les relations sémantiques à considérer sont la synonymie, la méronymie, la causalité et le but.

4. Présentation de la méthode

Dans cette section, nous présentons brièvement le module de construction du noyau d'ontologie. Cette partie du travail a été décrit avec précision dans [KAM 09a] et [KAM 09b]. Nous focalisons notre attention sur la chaîne de traitement destinée à l'enrichissement du noyau d'ontologie. Cette chaîne met au jour des concepts et des relations sémantiques entre ces concepts à l'aide de patrons exploitant à la fois des informations lexicales, syntaxiques et structurelles.

4.1. Construction du noyau d'ontologie

Partant du constat que les éléments de structure servent à organiser, subdiviser, hiérarchiser le contenu d'un document, alors cette structure est porteuse de sémantique, soit des relations hiérarchiques entre unités textuelles. Partant de là, il est possible de construire

10. Matérialisées par des balises XML spécifiques.

un noyau d'ontologie valide pour un corpus donné. L'idée de patron structurel proposé et mis en oeuvre par [KAM 09b] permet d'associer une propriété sémantique à un élément structurel. Le patron lexico-syntaxique, quant à lui, prend généralement en compte des éléments lexicaux et syntaxiques présents dans le texte non structuré. Plus précisément, les titres (emboîtement ou parallélisme de titres et sous-titres) ou encore les structures énumératives peuvent permettre de mettre en lumière des relations de subordination ou de juxtaposition entre les termes contenus dans ces mêmes éléments structurels. Les auteurs insistent sur la nécessité de disposer d'un document au format qui rend explicite sa structure et son caractère hiérarchique, par exemple, le XML. Enfin, les termes doivent désigner sinon des concepts, au moins des termes valides pour le domaine visé. La démarche proposée se fait en quatre étapes : (i) associer une sémantique aux balises et aux relations qui les relient ; (ii) ré-annoter (automatiquement) le document pour mettre en évidence les propriétés des balises et pour attribuer des catégories grammaticales ou sémantiques aux unités textuelles ; (iii) définir les patrons structurels qui caractérisent la sémantique portée par un élément de structure et produisent un fragment d'ontologie ; et (iv) projeter les patrons. Cette méthode¹¹ décrite dans [KAM 09a] met en évidence 1183 concepts extraits à partir de la structure.

4.2. Chaîne de traitement pour l'enrichissement du noyau d'ontologie

Cette chaîne de traitement a été produite avec le logiciel LinguaStream¹². Son objectif est d'extraire automatiquement dans des textes en langage naturel des concepts et des relations entre ces concepts pour produire des fragments d'ontologie (de la forme : terme/concept-relation-terme/concept). La chaîne de traitement se décompose en cinq modules principaux : (i) la préparation du corpus en XML, (ii) les pré-traitements linguistiques et le repérage des syntagmes nominaux, (iii) la recherche de relation, (iv) la prise en compte de la structure (les titres exclusivement), (v) l'export en XML.

4.3. La préparation du corpus en XML

Ce module permet d'abord de sélectionner les textes du corpus à analyser ainsi que, à l'intérieur des textes, les zones textuelles délimitées par des balises particulières. Pour ce qui nous concerne, nous exploitons les zones de titres et les zones de définition (cf. 1 et 2.

4.4. Les pré-traitements linguistiques et le repérage des syntagmes nominaux

Cette étape est constituée de plusieurs modules. Tout d'abord, un segmenteur, fourni nativement par LinguaStream, permet de découper le texte en mots. Puis l'analyseur morpho-syntaxique TreeTagger¹³ fournit un étiquetage morpho-syntaxique des mots du

11. Développée à l'aide de Gate (<http://gate.ac.uk>)

12. <http://www.linguastream.org/>

13. <http://www.ims.uni-stuttgart.de/projekte/complex/TreeTagger/>

corpus. Un troisième module, basé sur des macro-expressions régulières, permet le découpage du corpus en phrase. 23 884 mots et 1 978 phrases sont annotés dans le corpus. Le repérage des syntagmes nominaux est réalisé à l'aide d'une grammaire EDCG. Il permet de repérer et d'annoter :

- des SN syntaxiquement simples (*un gué*)
- des SN avec des noms composés, [N de N] (*l'hôtel de police*)
- des SN plus complexes, constitués soit d'adjectifs (*un chemin étroit, un bâtiment industriel et commercial*), soit de compléments du nom "spéciaux" où l'un des éléments introduit une relation sémantique particulière comme *partie, tas, ensemble*, etc. (*un tronçon de route*)

En plus du repérage des syntagmes nominaux, une structure de traits sémantiques est associée automatiquement à chacune des expressions. Les informations suivantes sont associées : le type de marqueur éventuellement présent (méronyme, holonyme, hypéronyme, mot composé, mot simple), le lemme court (dans la plupart des cas le lemme du nom tête du SN), le lemme long (totalité de l'expression nominale) et des informations de type syntaxique. Ce module annote 6 244 syntagmes nominaux. Par exemple, l'occurrence *Ligne de métro* sera annotée comme cela :

```

<lss :sem type="sn" id="296">
<lss :text>Ligne de métro</lss :text>
<lss :value>
<marqueur>compose</marqueur>
<lemmasCourt>ligne</lemmasCourt>
<lemmasLong>ligne de métro</lemmasLong>
<syntaxe>null</syntaxe>
</lss :value>
</lss :sem>
```

Figure 3. Exemple de repérage et d'annotation d'un syntagme nominal

Sur la base de l'ensemble de ces traitements, l'étape suivante consiste à mettre au jour des relations sémantiques entre les syntagmes repérés, sans préfigurer, à ce niveau de l'étude, de leur qualité de concept ou non.

4.5. La recherche de relations sémantiques

Cette étape se concentre sur le repérage des relations sémantiques. Sont repérées les relations traditionnelles telles que l'hyponymie et la méronymie. Les deux grammaires correspondantes ont été créées et adaptées à partir du travail accompli avec Caméléon [AUS 08]. Il s'agit à la fois d'une adaptation et d'un enrichissement de ces travaux relativement nos propres problématiques et nos corpus. Concernant les relations d'artefact et de fonction, elles ont été créées pour le projet Géonto, sur la base d'une étude manuelle du corpus. Dans chacune des grammaires, la relation s'établit entre un X et un Y. Dans certains cas, lorsque que le X ou le Y est absent de la relation intra-phrastique, c'est le titre qui s'avère jouer, la plupart du temps, le rôle du X ou du Y.

La relation de méronymie est transitive et anti-symétrique. Elle est repérée dans notre grammaire à l'aide de règles du type « X est une partie de Y » ou « Y est composé de X ». Le X n'est donc pas forcément le premier élément de l'expression repérée. La grammaire comporte 68 types de règles, chacune étant divisée en plusieurs configurations (présence du verbe être ou de virgules, absence explicite du X, ...). Dans le corpus, 18 cas de méronymie sont repérés¹⁴. Dans l'exemple suivant, la règle r12c s'applique à la phrase *Voie qui fait partie du domaine public* et extrait une relation de méronymie entre "voie" (X) et "partie du domaine public" (Y).

```
<lss :sem type="meronymie" id="1">
<lss :text>Voie qui fait partie du domaine public</lss :text>
<lss :value>
<regleMero>r12c</regleMero>
<lemmasCourtX>voie</lemmasCourtX>
<lemmasLongX>voie</lemmasLongX>
<snX>Voie</snX>
<lemmasCourtY>domaine</lemmasCourtY>
<lemmasLongY>partie domaine public</lemmasLongY>
<snY>partie du domaine public</snY>
</lss :value>
</lss :sem>
```

Figure 4. Exemple de relation de méronymie

La relation d'hyperonymie est une relation transitive, anti-symétrique et hiérarchique du type « X est un Y », « X est une sorte de Y ». Dans l'exemple suivant, la règle r20a trouve une relation d'hyperonymie entre *autoroute* et *route*.

```
<lss :sem type="hyperonymieInSentence" id="1">
<lss :text>Les autoroutes sont des routes</lss :text>
<lss :value>
<regleHyperoS>r20a</regleHyperoS>
<lemmasCourtX>autoroute</lemmasCourtX>
<lemmasLongX>autoroute</lemmasLongX>
<snX>Les autoroutes</snX>
<lemmasCourtY>route</lemmasCourtY>
<lemmasLongY>route</lemmasLongY>
<snY>routes</snY>
</lss :value>
</lss :sem>
```

Figure 5. Exemple de relation d'hyperonymie

La grammaire dédiée à la relation d'hyperonymie comporte 20 règles. Dans notre corpus, 48 relations d'hyperonymie sont instanciées (seulement 2 dans BDTopo). Dans la littérature, cette relation est souvent citée comme étant la plus riche [AGU 09] alors que dans notre corpus, il semble que ce soit le contraire. Cette différence est sans doute liée au type de corpus qui, dans notre cas, est très spécialisé.

14. Il y a une différence entre le nombre de règles et le nombre d'instances : un certain nombre de règles/ patrons ne s'instancient pas sur ce corpus, les patrons ayant été développés et évalués sur des textes diversifiés et pas uniquement sur les textes de spécification de bases de données. L'intérêt de cette méthode est de se donner les moyens de traiter efficacement de nouveaux textes.

La relation d'artefact correspond à une relation du type « X est représenté par » ou « X est employé pour ». C'est une relation symétrique. Ce type de relation se caractérise par le fait que deux concepts sont mis en relation de quasi-synonymie dans un domaine spécifique, dans notre cas, les bases de données de l'IGN. Dans le monde *réel* (*i.e.* général), ces concepts ne seraient en aucun cas considérés comme des synonymes. Cette relation est à mettre en parallèle avec la distinction entre monde réel et monde carto proposé par le Cogit¹⁵ : un artefact est utilisé pour désigner de manière plus ou moins détournée des objets ou des ensembles d'objets du monde réel [KAS 09]. Dans l'exemple ci-dessous, le concept X n'est pas explicite dans la phrase mais doit être repris du titre, d'où la valeur « fromTitre » de « lemmasCourtX », « lemmasLongX » et « snX ».

```
<lss :sem type="artefact" id="5">
<lss :text>représentant un danger potentiel</lss :text>
<lss :value>
<regleArtef>r4aSansX</regleArtef>
<lemmasCourtX>fromTitre</lemmasCourtX>
<lemmasLongX>fromTitre</lemmasLongX>
<snX>fromTitre</snX>
<lemmasCourtY>danger</lemmasCourtY>
<lemmasLongY>danger potentiel</lemmasLongY>
<snY>un danger potentiel</snY>
</lss :value>
</lss :sem>
```

Figure 6. Exemple de relation d'artefact

Cette grammaire est constituée de 6 règles principales, chacune étant divisée en plusieurs configurations possibles (présence du verbe être ou de virgules, absence explicite du X dans la zone de définition, ...). Ce module repère 14 relations d'artefact dans notre corpus.

La relation de fonction indique à quoi un objet sert. C'est une relation anti-symétrique qui introduit un Y composé soit d'un SN seul (*le transport des marchandises*) ou un syntagme verbal qui pourrait être nominalisé (*transporter les marchandises*, *le transport des marchandises*). Dans l'exemple qui suit, une relation fonctionnelle de type *faire passer des véhicules* est associée à *trajet de bateau*.

La grammaire consacrée aux artefacts est constituée de 6 règles principales, chacune étant divisée en plusieurs configurations (présence du verbe être ou de virgules, absence explicite du X, ...). Dans notre corpus, elle repère 82 relations de fonction.

4.6. La prise en compte de la structure

Ce dernier module repère deux types différents de relations nécessitant la prise en compte de la structure du document. Il est basé sur les observations faites dans [KAM 09b] concernant la notion de patron structurel. Trois types de relations sont mises en avant :

15. Laboratoire de l'IGN, partenaire du projet Géonto.

```

<lss :sem type="fonction" id="4">
<lss :text>Trajet du bateau servant à passer des véhi-
cules</lss :text>
<lss :value>
<regleFonct>r2b</regleFonct>
<lemmasCourtX>trajet</lemmasCourtX>
<lemmasLongX>trajet de bateau</lemmasLongX>
<snX>Trajet du bateau</snX>
<lemmasCourtY>véhicule</lemmasCourtY>
<lemmasLongY>passer véhicule</lemmasLongY>
<snY>passer véhicules</snY>
</lss :value>
</lss :sem>

```

Figure 7. Exemple de relation de fonction

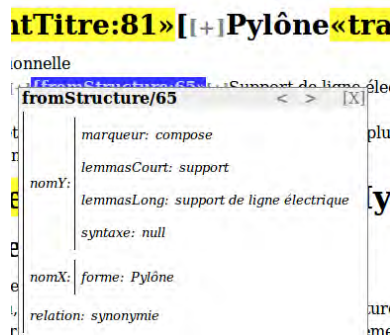


Figure 8. Une relation de synonymie entre deux concepts extrait à l'aide de la structure et d'informations lexico-grammaticales

- le type « synonymie stricte » : un terme-concept est défini par un SN simple (un nom/terme seul). Par exemple : *cascade* / *définition* : *chute d'eau*.
- le type « hyperonymie » : un terme-concept est défini par un un nom/terme auquel sont adjoints des adjectifs/propriétés. Par exemple : *gorge* / *définition* : *vallée étroite et encaissée*
- le type « méronymie » : un terme-concept est défini par un sn introduit par un marqueur de méronymie. Par exemple : *tronçon de route* / *définition* : *portion de voie de communication*

Dans le module, le premier traitement concerne les SN constitués d'un marqueur de méronymie, d'holonymie ou d'hyperonymie. Dans ce cas, le titre est nécessaire pour la création de l'un des deux concepts de la relation (X ou Y selon les cas). Le second cas concerne les SN simples qui suivent immédiatement un titre. Deux possibilités sont envisagées : soit le SN est seul et nous concluons sur une relation de synonymie ; soit le SN ne constitue que le début de la définition et nous concluons à une relation d'hyperonymie. Ce module repère 406 relations. La figure 8 montre un exemple pour la relation de synonymie.

5. Résultats et discussion

L'évaluation des ontologies en termes de performance est un problème reconnu mais délicat [SCH 05]. Dans l'idéal, disposer d'un gold-standart permettrait de fournir une évaluation quantitative intéressante. Ce type de processus nécessite un investissement humain coûteux essentiellement en termes d'annotation manuelle. Nous n'avons malheureusement pas pu mettre un tel système en place.

Nous avons mené une évaluation qualitative et manuelle des concepts et relations sémantiques associées aux concepts. Pour chaque relation annotée nous avons jugé de la pertinence de la relation elle-même ainsi que celle des concepts X et Y. Cinq valeurs sont possibles :

- la valeur est **valide** si la relation est jugée valide ainsi que les deux concepts X et Y reliés par cette relation,
- la valeur est **inverse** si la relation est valide mais est inversée (c'est le cas des relations anti-symétriques comme l'hyperonymie et la méronymie),
- la valeur est **approximative** lorsque la relation est valide mais les concepts sont approximatifs soit parce qu'ils ont été mal extraits par les programmes soit parce qu'il y a une relation de pronominalisation avec le titre et que c'est ce dernier qu'il faudrait prendre en compte,
- la valeur est **incertaine** lorsqu'une expertise particulière est requise (nécessité de recourir à des experts en cartographie),
- enfin, la valeur est **fausse** quand la relation est éronnée (dans ce cas, les concepts sont également souvent non valides).

Les résultats de cette évaluation sont indiqués dans le tableau suivant.

Relation	Artefact	Fonction	Hyperonymie	Synonymie	Méronymie	Holonymie	Total
Valide	16,7 %	14,3 %	38,8 %	25 %	19 %	58,3 %	33,1
Inverse	0 %	0 %	1 %	0 %	43 %	0 %	2,7
Approximative	16,7 %	23,8 %	12 %	25 %	19 %	8,3 %	15,4
Incertaine	66,6 %	55,5 %	19,6 %	0 %	0 %	16,7 %	22,2
Fausse	0 %	6,4 %	28,5 %	50 %	19 %	16,7 %	26,5

Ces résultats mettent en avant de grandes disparités qualitatives selon les types de relations sémantiques. Tout d'abord, les relations d'artefact et de fonction sont les relations présentant le plus haut pourcentage d'incertitude sur la pertinence de la connaissance extraite même si d'un point de vue linguistique, elles sont tout à fait acceptables. Cette incertitude est liée à leur spécificité pour le domaine et au recours nécessaire à des experts géographes pour leur validation. La relation d'hyperonymie présente des résultats corrects mais insuffisants. Presque 30 % des cas sont incorrects. Ces résultats vont dans le sens de ceux de [MAY 09] qui constatent, sur leurs données, une situation de surgénération de leurs patrons d'hyperonymie. Notre proposition pour le traitement de la synonymie n'est clairement pas convainquant : exploiter la structure et la syntaxe n'est pas suffisant pour distinguer l'hyperonymie de la synonymie. Enfin, les résultats concernant la méronymie et l'holonymie sont encourageants même si les patrons de méronymie doivent être revus pour ne pas générer des relations sémantiques inversant les concepts.

Une évaluation quantitative est en cours sur ces données. Nous mettons en place une procédure de comparaison des concepts existants dans le noyau ontologique avec les concepts (introduit à la section 4.4) et les relations fournis par la chaîne de traitement décrite dans cet article. Cette évaluation devrait fournir des résultats quantitatifs à même de rendre compte de l'intérêt de notre méthode d'enrichissement d'ontologie.

6. Conclusion

Cet article présente une méthodologie d'extraction de connaissances à partir de textes mise en place dans le cadre d'un projet de recherche visant l'appariement de connaissances géographiques hétérogènes. Il s'agit plus particulièrement de repérer de manière automatique des termes susceptibles d'être des concepts du domaine et des relations sémantiques particulières entre ces termes.

Dans la tradition des travaux exploitant les notions de marqueur et de patron lexicosyntaxique, notre approche utilise également la structure même des documents pour mettre en évidence des relations sémantiques entre les termes. Les résultats confirment l'idée que la structure logique permet de mettre au jour des relations d'hyponymie entre les termes du titre et les termes contenus dans les zones de définition. Mais la situation est moins convaincante pour ce qui est de la relation de synonymie.

L'évaluation des résultats a uniquement porté sur un jugement humain de la validité des relations entre termes extraits des documents. Une seconde évaluation, en cours, va permettre de mesurer dans quelle mesure les concepts et relations extraites enrichissent le noyau d'ontologie initial. Enfin, une évaluation par les experts est nécessaire pour notamment valider le statut de concept de chacun des termes repérés.

7. Bibliographie

- [AGU 09] AGUADODECEA G., ALVAREZDEMON I., MONTIEL PONSODA E., « From linguistic pattern to ontology structures », *TIA'2009*, 2009.
- [AUS 08] AUSSENAC-GILLES N., JACQUES M.-P., « Designing and evaluating patterns for relation acquisition from texts with Caméléon », *Terminology*, vol. 14, n° 1, 2008, p. 45-73.
- [AUS 09] AUSSENAC-GILLES N., HERNANDEZ N., « Du linguistique au conceptuel : identification de relations conceptuelles à partir de textes », *TIA*, 2009.
- [BOU 03] BOURIGAUT D., AUSSENAC-GILLES N., « Construction d'ontologies à partir de textes », *Conférence TALN 2003*, 2003.
- [BOU 04] BOURIGAUT D., AUSSENAC-GILLES N., CHARLET J., « Construction de ressources terminologiques ou ontologiques à partir de textes : un cadre unificateur pour trois études de cas », *Revue d'intelligence artificielle*, vol. 18, n° 1, 2004, p. 87-110.
- [DES 00] DESMOULINS C., GRANDBASTIEN M., « Des ontologies pour indexer des documents techniques pour la formation professionnelle », *IC'2000 : ingénierie des connaissances (Toulouse, 10-12 mai 2000)*, 2000.

14 Nom de la revue ou conférence (à définir par \submitted ou \toappear)

- [GRA 04] GRABAR N., HAMON T., « Les relations dans les terminologies structurées : de la théorie à la pratique », *Revue d'intelligence artificielle*, vol. 18, 2004.
- [HAD 02] HADDAD M., « Extraction et impact des connaissances sur les performances des systèmes de recherche d'information », PhD thesis, Université de Grenoble, 2002.
- [HEA 92] HEARST M., « Automatic Acquisition of Hyponyms From Large Text Corpora », *Proceedings of the 14th International Conference on Computational Linguistics*, 1992, p. 539-545.
- [JAC 06] JACQUES M.-P., AUSSENAC-GILLES N., « Variabilité des performances des outils de TAL et genre textuel », *TAL*, vol. 47, n° 1, 2006, p. 11-32.
- [KAM 09a] KAMEL M., AUSSENAC-GILLES N., « Construction automatique d'ontologies à partir de spécifications de bases de données », *Conférence IC*, 2009.
- [KAM 09b] KAMEL M., AUSSENAC-GILLES N., « Utiliser la Structure du Document dans le Processus de Construction d'Ontologies (regular paper) », L'HOMME M.-C., SZULMAN S., Eds., *Conférence Internationale sur la Terminologie et l'Intelligence Artificielle (TIA)*, <http://www.irit.fr/>, novembre 2009, IRIT, page (on line).
- [KAS 09] KASSEL G., « Vers une ontologie formelle des artefacts », *20es Journées Francophones en Ingénierie des Connaissances, Hammamet, Tunisie*, 2009.
- [KER 09] KERGOSIEN E., KAMEL M., SALLABERRY C., BESSAGNET M.-N., AUSSENAC-GILLES N., GAIO M., « Construction et enrichissement automatique d'ontologies à partir de ressources externes », *Conférence JFO 2009*, 2009.
- [MAY 09] MAYNARD D., FUNK A., PETERS W., « Using Lexico-Syntactic Ontology Design Patterns for Ontology creation and population », EVA BLOMQUIST KURT SANDKUHL F. S. V. S., Ed., *Proceedings of the Workshop on Ontology Patterns (WOP 2009)*, Washington DC, USA, vol. 516, 2009.
- [MUS 09] MUSTIÈRE S., ABADIE N., AUSSENAC-GILLES N., BESSAGNET M.-N., KAMEL M., KERGOSIEN E., REYNAUD C., SAFAR B., « GéOnto : Enrichissement d'une taxonomie de concepts topographiques », *SAGEO'2009*, Marne-la-Vallée, 2009.
- [ROL 06] ROLE F., ROUSSE G., « Construction incrémentale d'une ontologie par analyse du texte et de la structure du document », *Document numérique*, vol. 9, n° 1, 2006, p. 77-91.
- [SCH 05] SCHUTZ A., BUITELAAR P., « RelExt : A Tool for relation Extraction from Text in Ontology Extension », *Proceedings of the 4th International Semantic Web Conference (ISWC)*, 2005.

Une approche basée sur les ontologies et les agents pour l'interopérabilité des SIG

ALNAFIE Emdjed* — **HAMDADOU Djamila****

* Université d'Oran Es-senia, Laboratoire LIO, ALGÉRIE amazad@hotmail.fr

** Université d'Oran Es-senia, Laboratoire LIO, ALGÉRIE dzhamdadoud@yahoo.fr

RÉSUMÉ. Le domaine de l'information a connu un progrès extraordinaire permettant à de nombreuses spécialités et disciplines, de partager des sources d'information hétérogènes et de grandes masses. Cette communication, en particulier entre les systèmes d'information géographique (SIG), est malheureusement confrontée au problème de l'hétérogénéité des sources d'information. L'interopérabilité des SIGs est la solution qui permet de créer un unique et homogène système "virtuel" avec un accès simple, transparent et efficace à l'information nécessaire. Cet accès est assuré quelle que soit la localisation des SIGs ou leurs méthodes de stockage. L'objectif principal de cette étude, et dont les contributions seront en mesure de faciliter la prise de décision des gestionnaires concernés, est de proposer une approche fondée sur l'utilisation combinée des géo-ontologies et des agents. Cette approche adopte une politique de traitement permettant de trouver rapidement des réponses aux requêtes tout en garantissant un accès direct à ces réponses dans le cas d'une prochaine éventuelle requête de même nature.

ABSTRACT. The domain of information has known an amazing progress which allowed many specialities and disciplines, to share at the same time large and heterogeneous sources of information. This communication, in particular between the geographical information systems (GIS), is unfortunately, confronted to the problem of heterogeneity of the information resources. The GIS's Interoperability is the solution which makes possible to create a unique and homogeneous "virtual" system with a simple, transparent and effective access to required information, whatever their localization or their storage methods. The principal aim of this study, and whose contributions will be able to facilitate the decision-making of the managers concerned, is to propose an approach based on geo-ontologies and agents. This approach adopts a requests treatment policy making possible to find request's answers rapidly and guarantee a direct access to these desired answers in case of next possible requests of same nature.

MOTS-CLÉS : Système d'Information Géographique (SIG), Interopérabilité, Hétérogénéité, Géo-ontologies, Agent.

KEYWORDS: Geographic Information System(GIS), Interoperability, heterogeneity, Geo-ontologies, Agent

1. Introduction

Le domaine de l'information et de la communication a connu une avancée remarquable qui ne cesse d'engendrer de nouvelles techniques. Cette évolution exponentielle a permis à un nombre important de personnes, de différentes spécialités et domaines, de partager des sources d'information à la fois hétérogènes et d'une grande masse.

Il est donc clair que la recherche d'information qu'effectue un utilisateur va s'élargir pour toucher aux ressources en possession d'autres utilisateurs. Cette communication va malheureusement se heurter au problème d'hétérogénéité des ressources d'information, chose qui nécessite une standardisation de l'univers de discours, et c'est là que l'interopérabilité des systèmes joue un rôle primordial dans le partage des informations et des services.

Notre objectif à travers cette étude, est de proposer une approche pour l'interopérabilité des SIGs, fondée sur l'utilisation combinée des ontologies et des agents. Cette combinaison vise à unir la flexibilité des ontologies à l'efficacité des agents pour remédier au mieux du problème de l'hétérogénéité. L'approche permet donc d'échanger de manière cohérente les informations entre les SIGs et de trouver des réponses aux requêtes afin de répondre aux besoins des utilisateurs.

L'article présente en section 2 les concepts de base de l'interopérabilité des SIGs. La section 3 donne un rapide aperçu préliminaire sur les ontologies et présente les spécificités des géo-ontologies en particulier. L'approche proposée est décrite dans sa globalité en section 4 et nous concluons notre propos en section 5.

2. L'interopérabilité des SIGs : État de l'art

L'étude des phénomènes et des objets géographiques englobe des quantités de plus en plus grandes d'informations dispersées géographiquement et en possession d'utilisateurs qui ne partagent pas les mêmes contextes et ne suivent pas les mêmes contraintes. Cela dit, un problème sérieux et crucial n'est autre que l'interopérabilité des systèmes d'informations géographiques qui ont besoin énormément et fréquemment de coopérer et partager leurs ressources, pour répondre à leurs besoins.

L'interopérabilité est l'une des meilleures solutions dans le sens où elle permet non seulement de fournir aux utilisateurs la vision et l'impression que le système "virtuel" sur lequel ils travaillent est unique et homogène, alors qu'en réalité l'information se trouve distribuée dans des environnements hétérogènes, mais elle leur donne également la possibilité d'accéder d'une manière simple, transparente et efficace aux informations demandées, quelles que soient leur localisation ou leurs modalités de stockage, mais il ne faut pas oublier que cette partageabilité doit être réalisée à travers des mécanismes assurant le bon fonctionnement des systèmes en question.

Cependant, avant de se réjouir des avantages de l'interopérabilité, il est essentiel d'étudier les techniques permettant de la réaliser et parmi lesquelles "**l'utilisation des ontologies**" se montre très intéressante.

3. Les ontologies : concepts et définitions

Tout d'abord, une ontologie au sens global possède plusieurs définitions. La première, mais aussi la plus fréquemment utilisée, est apparue pour le projet ARPA Knowledge sharing, elle définit l'ontologie comme "une spécification explicite d'une conceptualisation" [GRU 93]. Elle peut être aussi définie comme une formalisation à un certain niveau de complexité d'informations structurées [SMI 01], mais il est possible également de proposer une version simplifiée et plus facile à comprendre qui présente l'ontologie comme "une hiérarchie de termes reliés par des relations générique/spécifiques [FLE et Al 09]", ou encore un ensemble de concepts (classes) ainsi que les relations existant entre ces concepts.

L'ontologie qui a pour rôle de prendre en compte et gérer l'information géographique est appelée "géo-ontologie" ou ontologie d'un domaine relevant de la géographie. Elle est définie par les différentes spécificités auxquelles est liée la multidimensionnalité des objets géographiques.

Cependant il faut bien faire la différence entre les géo-ontologies relatives à une thématique (agriculture, hydrologie, l'urbain...) et qui doivent leur appellation au domaine qu'elles traitent, de celles purement spatiales (d'où l'appellation "ontologies spatiales") qui se focalisent sur la spécification de concepts spatiaux dans un but de représentation cartographique.

Il faut aussi prendre en compte, avant de construire une géo-ontologie, l'identification des utilisateurs concernés ainsi que les différents contextes qui rentrent en jeu. Cela dit, une géo-ontologie se présente sous forme de vues multi-utilisateurs pour des besoins multi-contextes.

Le contexte peut être défini comme étant "un ensemble de conditions et influences environnantes qui permettent de comprendre la situation en faisant d'elle une situation unique". En termes de méthodologie de construction des géo-ontologies, il est possible d'en distinguer deux types, le premier basé sur la structure de l'objet géographique et le deuxième sur son fonctionnement.

4. L'approche proposée (INSIGO)

L'objectif principal de notre modèle "**INSIGO**" est de résoudre le problème de l'interopérabilité des SIGs. Il s'agit donc de lever toute sorte d'ambiguïté qui peut être due à l'hétérogénéité qui existe entre les différents SIGs communiquant entre eux.

Le modèle que nous proposons dans la présente étude est non seulement basé sur les ontologies mais aussi sur les SMAs (Système Multi-Agents) dans le but d'unir les forces de ces deux grands concepts afin d'aboutir à une approche à la fois concise et consistante.

4.1. Présentation de l'approche

INSIGO est une approche basée sur un modèle cylindrique réalisé à travers la connexion de plusieurs cellules de forme hexagonale de façon circulaire. Cela signifie qu'en partant d'une cellule donnée et en parcourant ses voisines¹, il est possible d'y retourner. La figure 1 illustre la composition de chaque cellule :

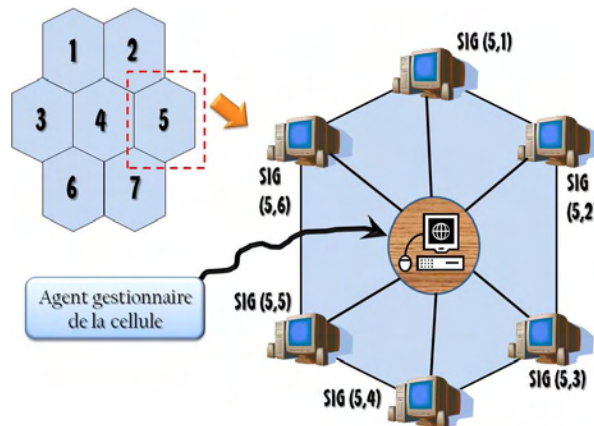


Figure 1. La composition des cellules

Sur chaque tête de l'hexagone se trouve un SIG muni d'une adresse composée du numéro de la cellule à laquelle il appartient ainsi que son numéro dans la cellule (ex : SIG(4,3) est le SIG numéro 3 de la cellule 4). Chaque SIG possède trois voisins directs et appartient pratiquement à trois cellules voisines et est doté, par conséquent, de trois adresses différentes avec lesquelles il est reconnu dans chaque cellule à laquelle il appartient(adresse locale).

Dans le but de réaliser l'interopérabilité, chaque SIG sera doté d'un module appelé "module d'interopérabilité" composé des éléments suivants :

- Une **ontologie de domaine** modifiée à chaque mise à jour du SIG, celle-ci

1. Le nombre de voisins est six exactement

contiendra l'ensemble des concepts qui relèvent du domaine que vise le SIG (agriculture, météorologie, urbanisme, etc.) en question ainsi que les rapports qui les relient. Cette ontologie subira des révisions chaque fois que le SIG subit des mises à jour influentes, c'est-à-dire des mises à jour qui ajoutent, suppriment ou modifient des concepts ou des relations entre ces concepts.

– Une **base de données** pour enregistrer les adresses des SIGs ayant déjà répondu à des requête antérieure afin d'y accéder directement. Cette base de données contient aussi les adresses locales du SIG pour chaque cellule à laquelle il appartient.

La connexion des SIGs de la cellule est assurée, comme le montre la figure (1), grâce à un agent appelé **gestionnaire de cellule** ou encore **GC**. Du moment que chaque cellule possède son GC, les gestionnaires sont à leur tours connectés entre eux. Chaque gestionnaire de cellule est connecté au GCs qui lui sont directement voisins soit six GC au nombre des cellules voisines comme le montre la figure (2).

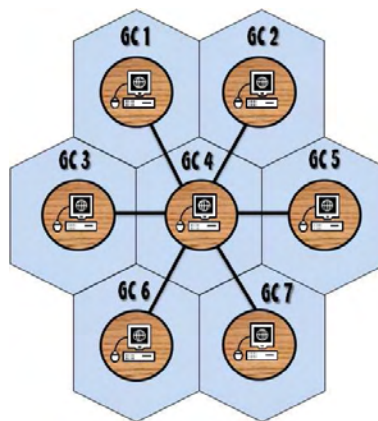


Figure 2. Connexion d'un GC avec ses voisins directs

Un gestionnaire de cellule possède également un "module d'interopérabilité" composé des éléments suivants :

- Un module de traduction des requêtes en langage ontologique
- Une ontologie d'application de la cellule qui assurera l'analyse syntaxique et l'orientation des requêtes vers les SIG de la cellules (plus précisément vers les ontologies de domaines)
- Une base de données qui contiendra :
 - La liste des cellules voisines et les SIGs partagés avec ces cellules ;
 - Les adresses des SIGs qui ont répondu à des anciennes requêtes ;

Les gestionnaires sont connectés à un agent appelé **gestionnaire global de cellule** ou encore **GGC**.

Le GGC est un agent désigné pour jouer le rôle d'intermédiaire entre les différents GCs en particulier ceux qui ne sont pas voisins directs. En dépit de son appellation qui donne l'impression que sa structure est plus complexe que celle des GCs, le GGC est en réalité une version simplifiée des gestionnaire des cellules et il ne fait que transmettre les requêtes qui lui arrivent vers le destinataire désiré. L'utilisation d'un tel agent permet de réduire les coûts de connexion entre les GCs surtout lors de l'ajout d'une nouvelle cellule au réseau. La figure (3) illustre la connexion entre les gestionnaires des cellules et le gestionnaire global.

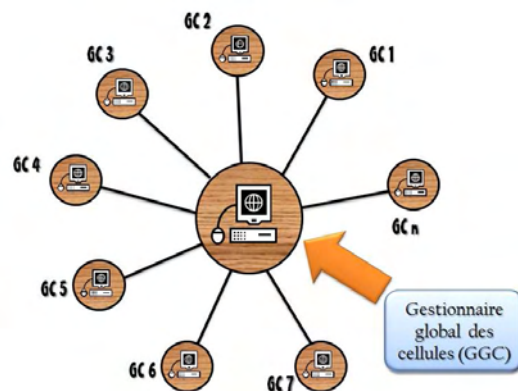


Figure 3. Connexion des GC avec le gestionnaire globale

La présence de agents gestionnaire de cellules crée deux niveau dans notre approche : le niveau SIGs et le niveau gestionnaires. les deux niveaux sont construits sur la même base, c-à-d une connexion avec les voisins directs et une connexion avec un agent assurant le rôle de gestion.

4.2. *Le traitement des requêtes*

Le traitement des requêtes se fait à base des deux niveaux de "INSIGO" : SIGs et gestionnaires.

4.2.1. *Le niveau des SIGs*

Le SIG interrogeant tente, en premier lieu, de répondre à la requête localement, et renvoie la réponse à l'utilisateur si elle existe et sinon il transmet la requête aux SIGs

qui lui sont directement voisins.

Chacun de ces derniers tente un traitement local de la requête à condition qu'il soit homogène avec le SIG interrogeant. La requête doit être écrite dans le même langage que celui du SIG voisin. Dans le cas contraire, un rejet de la requête est signifié au SIG envoyant la requête (les rejets sont mentionnés par les points rouges dans le diagramme de séquence niveau SIG).

Si la réponse existe, elle sera transmise dans le chemin inverse qu'a parcouru la requête munie de l'adresse du SIG ayant répondu à la requête pour l'enregistrer dans le module BDD du SIG interrogeant, sinon, les SIGs voisins transmettrons à leur tour la requête à leurs voisins et le cycle continue tant qu'il n'y a pas de rejet ou que la requête retourne au SIG interrogeant sans réponse(modèle cylindrique circulaire). Le diagramme de séquence correspondant au traitement des requêtes au niveau des SIGs est illustré par la figure (4).

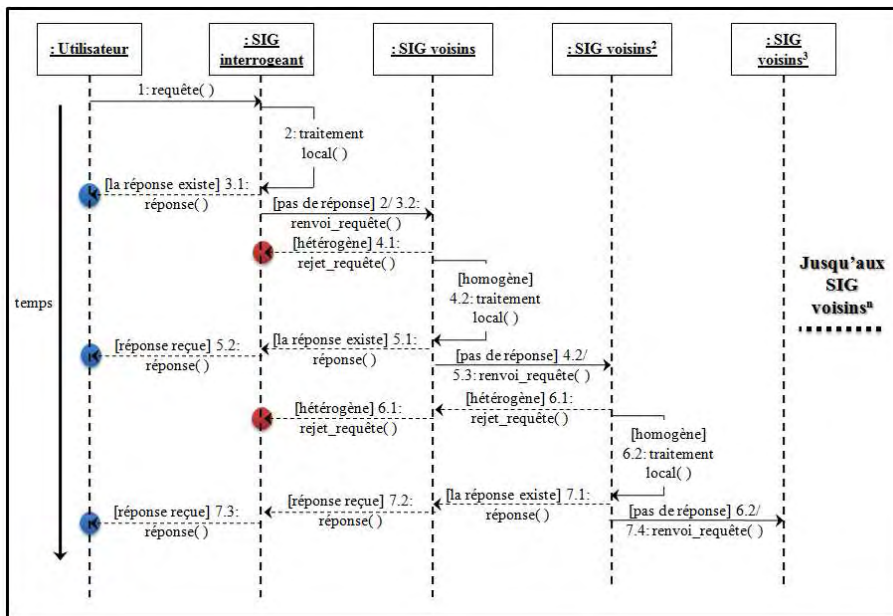


Figure 4. Diagramme de séquence, niveau SIG

La figure (5) montre l'organigramme du fonctionnement de l'approche au niveau des SIGs.

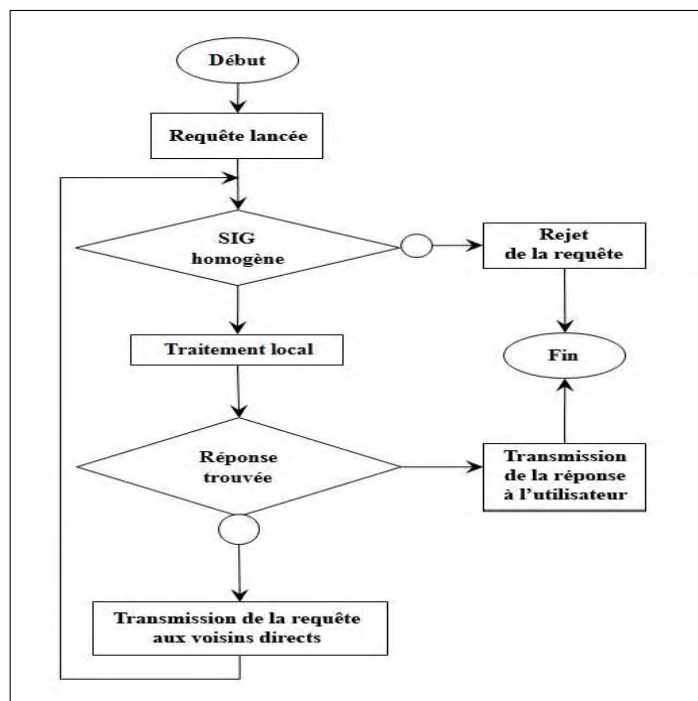


Figure 5. L'organigramme du fonctionnement du niveau des SIGs

4.2.2. Le niveau des gestionnaires

Le deuxième diagramme illustre le passage au deuxième niveau (niveau gestionnaire) à partir des points rouges sur le premier diagramme tel que ces points représentent les rejets de la requête. Une fois la requête rejetée ou revenue au SIG interrogant sans réponse, ce dernier la transmet au gestionnaire de la cellule qui utilisera son module de traduction afin de transformer la requête en langage ontologique dans le but d'enlever toute sorte d'ambiguïté due à l'hétérogénéité entre les SIGs.

Une fois la traduction terminée, la nouvelle requête est vérifiée syntaxiquement par l'ontologie d'application pour éviter toutes sortes d'erreurs, ensuite le GC accédera aux ontologies de domaine de chaque SIG de sa cellule sauf l'interrogeant dans une tentative de trouver un SIG qui peut répondre à la requête et transmettre la réponse au SIG interrogant. Si cette tentative échoue, le GC consulte ses voisins qui vont à leurs tours consulter les SIGs de leurs cellules et ainsi de suite jusqu'à ce que la réponse soit trouvée.

A ce moment, le GGC intervient pour assurer une communication directe entre le GC qui détient la réponse et le GC interrogant afin de lui transmettre cette réponse

qui sera munie de l'adresse du SIG lui ayant répondu. cette adresse est enregistrée à la fois dans le module BDD du SIG interrogeant ainsi que dans le module BDD du GC. La figure (6) montre le diagramme de séquence correspondant au traitement des requêtes au niveau des gestionnaires.

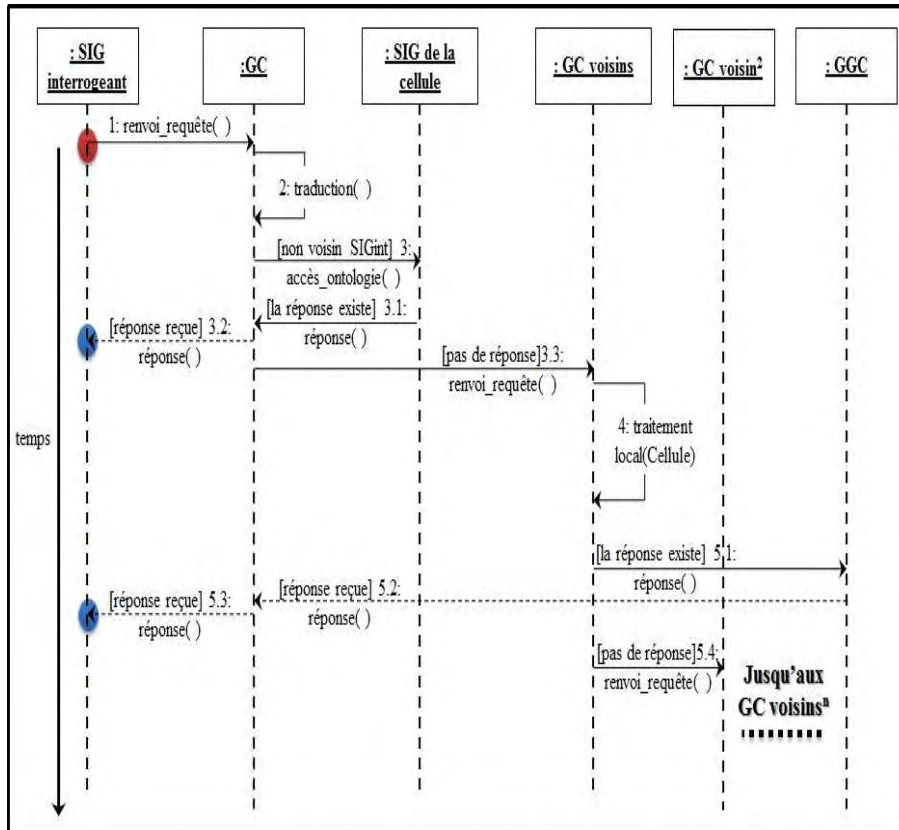


Figure 6. Diagramme de séquence niveau gestionnaire

La figure (7) montre l'organigramme du fonctionnement de l'approche au niveau des gestionnaires.

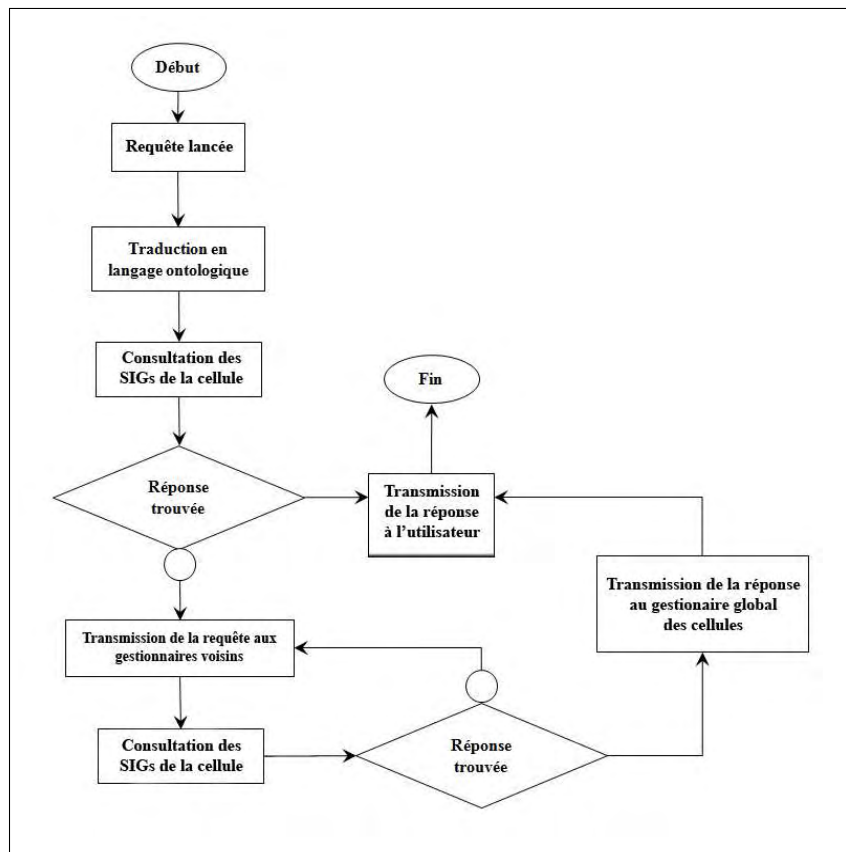


Figure 7. L'organigramme du fonctionnement du niveau des GCs

4.3. Avantages de l'approche

"INSIGO" adopte un traitement basé sur le même principe, traitement local suivi d'une consultation des voisins jusqu'à ce que la réponse soit trouvée ; cette uniformité du principe facilite la réalisation de cette approche. D'autre part, l'utilisation des agents pour la gestion des cellules au lieu de SIGs, permet au SIG de se consacrer à répondre aux requêtes et au GC d'en gérer le trafic. Il faut aussi noter le plus important à savoir l'utilisation des ontologies (la traduction de la requête en langage ontologique) permet d'assurer une homogénéité entre les SIGs sans pour autant être obligé d'uniformiser leurs langages internes.

5. Conclusion

La communication entre les systèmes d'informations géographiques est devenue une nécessité incontournable (catastrophes naturelles, météo, recherche scientifique ...). Cependant, partager les informations est une tâche variée et complexe en même temps. Cela est dû non seulement à la spécificité de chaque système, voire son organisation et le langage qu'il utilise, mais aussi à son type, le mode de transfert des données échangée et le coût des technologies utilisées pour répondre au besoin des utilisateurs.

L'interopérabilité pourra sans doute relever tous ces défis et en même temps rassembler toutes les préoccupations qui y sont attachées. En connaissance de tous cela, le besoin de techniques assurant cette interopérabilité se montre plus que nécessaire. Parmi ces techniques nous avons cité l'utilisation des applications et services Web, ainsi que les modèles basés sur des logiques terminologiques ou encore les modèles sémantiques.

En dépit des avantages de ces techniques, nous avons ciblé l'utilisation des ontologies comme technique qui se montre capable de résoudre le problème dans plusieurs domaines et disciplines de l'informatique pure à la médecine en passant par la biologie et la géographie.

Cette efficacité de l'ontologie est exprimée par sa définition simplifiée. L'association des ontologies et des agents informatiques pourra donner naissance à des structures flexibles et facile à implémenter, ce qui permet une large panoplie d'idées et d'architectures parmi lesquelles se situe notre approche "**INSIGO**" basée sur une connexion de cellules hexagonales ainsi que les agents gestionnaires de cellules.

Le traitement des requêtes se fait à deux niveaux en se basant sur un traitement local suivi d'une consultation des voisins directs (au niveau SIG tout comme au niveau GC), sachant que le modèle cylindrique construit de plusieurs cellules hexagonales permet un voisinage circulaire important. Enfin, bien que cette approche soit structurée en deux niveaux, ces derniers utilisent le même principe de base.

6. Bibliographie

- [KHO 07] *Conception d'une nouvelle ontologie pour l'interopérabilité des systèmes d'informations géographique*. Hacene BELHADEF et Dr Mohamed-Khireddine KHOLLADI, 4th International Conference : Sciences of Electronic, Technologies of Information and Telecommunications March 25-29, 2007 - TUNISIA
- [GRU 93] Gruber T.R., 1993, *A translation Approach to Portable Ontology Specifications*, Technical report KSL 92-71, Knowledge systems laboratory, Stanford University, California, 27p
- [SMI 01] Smith B., 2001, *Geographical categories : an ontological investigation*, in international journal of geographical information science, Vol.15 No. 7, pp. 591-612.
- [FLE et AI 09] Yann FLETY, Marie-Hélène DE SEDE, Vers une géo-ontologie pour les Systèmes Energétiques Territoriaux, XVI rencontres de Rochebrune, "Ontologies et dynamique des systèmes complexes", Janvier 2009, Megève.
- [GAN 02] Gandon F., 2002, *Distributed artificial intelligence and knowledge mangement : Ontologies and multi-agent systems for a coperate semantic web*, Thèse de Doctorat, University of Nice and INRIA, 486p.
- [SMI 04] Smith B., 2004, Beyond concepts : ontology as reality representation, in Proceedings of the international Conference on Formal Ontology and Information Systems FOIS 2004, Turin, Italy, 4-6 November 2004, 12p.
- [LEC et AI 98] Eric Leclercq, Djamel Benslimane, Kokou Yetongnon, *"ISIS : une architecture multi- agents pour l'interopérabilité des SIG"*, colloque SMAGETn Cemagref-Engref, Clermont-Ferrand, France 1998.
- [NOY et AI 00] *"Développement d'une ontologie 101 : Guide pour la création de votre première ontologie"* Natalya F. Noy et Deborah L. McGuinness Université de Stanford, Stanford, CA, 94305
- [MOR et AI 97] *Le Robert Collège* Daniel MORVAN, François GÉRARDIN, Annick DEHAIS, Brigitte VIENNE, Marie-José BROCHARD, Alette LUCOT, Alain REY, Dictionnaires IE ROBERT, 1997
- [PHI et AI 05] *De l'interopérabilité des SIG* Claude Philipona, Alexandre Fellay, bulletin HEC 71, Entreprise Camptocamp SA, FRANCE, 2005
- [DEG et AI 97] *Les enjeux de l'interopérabilité sémantique dans les systèmes d'information de santé* Patrice Degoulet, Marius Fieschi, Christophe Attali, Informatique et Gestion Médicalisée, Volume 9 Springer-Verlag France, Paris, 1997
- [PSY et AI 03] *Apport de l'ingénierie ontologique aux environnements de formation à distance*, Valéry PSYCHÉ, Olavo MENDES et Jacqueline BOURDEAU. Revue STICEF, 10, 2003
- [MEL 07] *Réalisation de l'interopérabilité sémantique des systèmes, basée sur les ontologies et les flux d'information* Thèse pour obtenir le grade de docteur de l'université de Savoie, Naçima MELLAL, Laboratoire d'Informatique, Systèmes, Traitement de l'Information et de la Connaissance, Université de Savoie, 2007
- [GUA et AL 95] *Ontologies and knowledge bases : Towards a terminological clarification*, Nicola GUARINO et Pierdaniele GIARETTA. In N MARS, réd., Towards Very Large Knowledge Bases : Knowledge Building and Knowledge Sharing, pages 25-32. IOS Press, 1995.

AOC, une ontologie formelle pour la modélisation de systèmes complexes en géographie

Patrice Langlois — Eric Daudé — Baptiste Blanpain — Emmanuel Sapin

UMR CNRS 6228 "IDEES", équipe MTG

Université de Rouen

1 rue Thomas Becket, bât. IRED, 76821 Mont-Saint-Aignan Cedex

Patrice.langlois@univ-rouen.fr

RÉSUMÉ. Cet article présente un méta-modèle nommé AOC (Agents-Organisations-Comportements), qui s'exprime par une ontologie formelle et sert de cadre conceptuel à une plateforme de simulation de systèmes complexes spatialisés, compatible avec les problématiques de dynamiques géographiques les plus larges. Le modélisateur géographe doit pouvoir concevoir son modèle et le formaliser dans un langage cohérent relativement à son cadre de pensée. AOC offre ainsi la possibilité de structurer les relations entre agents via le concept d'organisation, qui permet de créer des modèles multi-niveaux. De plus, une formalisation générale de la notion de comportement permet de modéliser la dynamique d'une classe très large de systèmes.

MOTS-CLÉS : Système complexe, système multi-agents, ontologie géographique, méta-modèle, dynamique spatiale, plateforme de modélisation.

ABSTRACT. This paper presents a meta-model named AOC (Agents-Organisation-Behaviour) that is described by a formal ontology. This meta-model is the theoretical concept of a simulation platform of spatial complex systems that are compatible with the largest dynamic geographic problems. The geographer should be able to conceive his model and to formalise it in a coherent language according to his way of thinking. Thanks to the concept of 'organisation', AOC offers the possibilities to structure the relationships between agents and to create multi-level models. Moreover, a general formalisation of the notion of behaviour is used to simulate the dynamic of very large types of systems.

KEYWORDS: Complex system, multi-agent system, geographic ontology, meta-model, spatial dynamic, simulation platform.

1. Introduction

Une problématique fondamentale en informatique moderne est l'étude des systèmes dits complexes, dans lesquels des agents indépendants peuvent interagir pour s'influencer entre eux et faire évoluer le système. Parmi les systèmes complexes, les systèmes multi-agents, dans lesquels des agents artificiels opèrent collectivement et de façon décentralisée pour accomplir une tâche ou simuler un phénomène, sont étudiés dans de nombreuses disciplines (informatique, géographie, sociologie, économie, biologie, etc.), par des informaticiens comme par des non informaticiens (Phan et al., 2007)

La nécessité d'un cadre théorique dans lequel le modélisateur peut penser son modèle et le formaliser est grandissante. Dans cet article, nous présentons l'ontologie d'un paradigme, c'est-à-dire un cadre conceptuel associé à un moyen de produire une certaine classe de modèles (Ramat, 2006). Ce paradigme, résumé par l'acronyme AOC (Agents-Organisations-Comportements), sert de base à une plateforme intégrée permettant la conception, le codage et la simulation de systèmes complexes en géographie (programme ANR et FEDER « MAGEO »).

Notre approche s'est inspirée initialement de l'ontologie AGR (Agent-Groupe-Rôle) de Ferber (Ferber, 1995), (Ferber et al., 2005), en y ajoutant le concept d'organisation multi-échelles, qui s'applique aussi bien à la structuration spatiale que sociale. Par ailleurs, le concept d'agent, dans le contexte AOC, s'applique aussi bien à des individus sociaux qu'à des entités structurantes de l'environnement. En effet, l'environnement est, pour le géographe, souvent autant le sujet d'étude que les individus qui s'y meuvent, même si la temporalité de sa dynamique est différente de celles des individus qui s'y déplacent.

AOC est une manière de représenter le monde à travers des modèles, par combinaison de quelques concepts simples à travers une certaine formalisation. Cet article présente le paradigme AOC sous forme d'une ontologie formelle orientée utilisateur, indépendante de toute application informatique.

On peut dire que l'ontologie présentée est formelle car ses entités sont abstraites (universelles) et permettent de générer des objets concrets (particuliers) dans un modèle. Le paradigme AOC fournit un cadre théorique au modélisateur qui est indépendant de l'informatique et dans lequel il peut penser son modèle, pour le décrire ensuite dans un langage utilisateur cohérent par rapport à ce cadre théorique. Ce cadre doit donc être facilement transposable à sa discipline, à ses thématiques, à ses méthodes et à son langage ; il est donc proche de sa manière de penser, à l'instar du langage à base de règles de SpaCelle (Dubos-Paillard et al., 2003). Ce paradigme offre un ensemble de concepts à la fois suffisamment abstraits pour que le modélisateur puisse y puiser une grande variété de réalisations, mais aussi suffisamment simples afin qu'il puisse les utiliser facilement. Enfin, le contexte théorique de la plateforme ne doit pas être alourdi et obscurci par des aspects

techniques purement informatiques n'ayant rien à voir avec la problématique du modélisateur. Dans cet article, nous ne décrivons donc pas une ontologie de la plateforme elle-même, mais l'ontologie de la classe des modèles qu'elle peut générer.

2. Définition des concepts de AOC

Après avoir exposé les concepts préalables d'espace et de temps, nous décrivons les concepts spécifiques à AOC que sont les agents, les organisations et les comportements.

2.1. *L'espace et le temps*

Toute modélisation géographique se situe à une certaine échelle d'espace et de temps. Cette échelle se caractérise par une extension (un domaine d'espace et de temps) et une granularité (discrétisation). L'espace géographique est constitué d'un référentiel géométrique paramétrable (nombre de dimensions, origine géographique, orientation des axes, unités, pas élémentaires, extension du domaine et métrique). De même le temps est représenté par un référentiel monodimensionnel qui définit une temporalité (par une origine, une unité de temps, un pas de temps élémentaire, une extension). A un référentiel d'espace on associe un référentiel de temps qui lui est compatible en termes de discrétisation et d'extension, aussi nous parlerons de référentiel d'espace-temps.

Les problématiques géographiques utilisent souvent plusieurs échelles de représentation des phénomènes. A chaque échelle est associé un référentiel d'espace-temps. Néanmoins il est nécessaire de définir une correspondance entre ces référentiels, en prenant le référentiel le plus fin (dit référentiel général) comme base de tous les autres.

Dans l'ontologie AOC, les objets du monde se localisent dans des référentiels d'espace-temps, associés à des organisations spatiales qui ont pour rôle de gérer les contraintes de déplacement et les interactions avec les autres objets. Ces organisations spatiales sont abordées dans la partie 2.3.

2.2. *Agent*

Dans le paradigme des systèmes multi-agents, un agent est une entité qui possède une certaine individualité, une indépendance, une certaine autonomie d'action, et une connaissance partielle de son environnement. Un agent possède une mémoire (dont la valeur à un instant donné représente son état) lui donnant la possibilité de conserver, à travers le temps, certaines de ses caractéristiques matérielles, énergétiques ou informationnelles. La dynamique d'un agent s'exprime par des changements internes de son état et des changements externes, par ses

actions qui modifient l'état des objets de son environnement. Elle est décrite dans la partie 2.4.

Un agent représente donc une entité du monde réel ayant un comportement. Cette entité peut être aussi bien une maison qu'un être humain, voire une entité abstraite.

A un instant donné, un agent peut être à la fois « positionné » socialement et localisé dans l'espace géographique.

Agent géographique

Un agent géographique est un agent localisé (pays, ville, individu) qui peut avoir une dynamique. Il appartient souvent à une couche d'information géographique, en général importée d'un SIG (Système d'Information Géographique). Il possède une emprise spatiale (ponctuelle, linéaire, ou surfacique), appelée aussi son support. Cette localisation peut être exprimée dans le repère de l'espace général ou être relative à un composant sur lequel il se localise (par exemple, un voyageur peut avoir des coordonnées terrestres absolues, ou avoir seulement une localisation relative dans le train dont il est passager). Un agent géographique possède donc des attributs spatiaux qui servent à décrire sa forme et sa localisation dans un référentiel donné.

2.3. Organisation

Si l'agent modélise ce qui est de l'ordre de l'individuel ou de l'élémentaire, dans AOC, une organisation modélise ce qui est de l'ordre du collectif. Mais ne nous trompons pas, une organisation n'est pas une collection d'objets ou un ensemble d'éléments ; une organisation est la structure qui permet à ces objets ou à ces éléments de se relier, de se séparer, de se combiner, de communiquer, d'interagir entre eux. En somme, l'organisation est ce qui fait la différence entre un ensemble d'objets et un système d'objets.

Définition : une organisation O définie sur un ensemble A d'agents, appelés ses composants, est une structure S munie d'un ensemble de règles d'évolution et de gestion des interactions avec les autres agents du système.

Cette définition, très générale, recouvre des façons très diverses de structurer des collectivités d'agents. Ces différentes approches peuvent néanmoins être classées en plusieurs catégories, comme nous allons le voir.

Si l'organisation structure les agents dans le référentiel spatial, on parle d'organisation spatiale, sinon on parle d'organisation sociale. L'adjectif social est pris ici dans un sens très large, qui va au-delà des exemples précédents.

Organisations spatiales

Les organisations spatiales disposent d'une position et d'un volume dans le référentiel spatial du modèle ; elles y organisent les agents les uns par rapport aux autres. Parmi les organisations spatiales, on peut distinguer :

- Les organisations formées de composants définissant un maillage de l'espace. Ces maillages peuvent être réguliers (rectangles, hexagones, triangles), ou irréguliers (découpages administratifs).
- Les organisations qui portent sur des composants matériels dissociés les uns des autres, donc ne représentant pas une partition de l'espace. Ces composants peuvent occuper des volumes nuls (semis de points), ou plus ou moins étendus dans une ou plusieurs dimensions (réseaux routiers faits de tronçons et de nœuds, ou les agglomérations dans un espace national).
- Les organisations qui portent sur des composants situés, mais non matériels, ce qui leur confère la propriété d'être interpénétrables. Les composants ne décrivent donc pas nécessairement une partition du volume, puisqu'ils peuvent couvrir des volumes d'intersection non nulle (zones d'influence religieuse, zones d'influence commerciale, etc.).

Organisations sociales

Les organisations dites sociales sont les organisations qui structurent leurs composants les uns par rapport aux autres sur d'autres critères que leur position spatiale. Les deux catégories remarquables sont :

- Les organisations qui structurent un ensemble d'agents abstraits (c'est-à-dire sans réalité matérielle) les uns par rapport aux autres. Un exemple peut être l'ensemble des catégories socioprofessionnelles (CSP), où chaque catégorie prend le statut d'agent ; un autre exemple est celui des différentes fonctions d'une entreprise, à savoir les fonctions d'administration, de production, de recherche, de direction, etc. Ces différents composants abstraits (qui, rappelons le, possèdent le statut d'agents), trouvent leur intérêt en permettant à d'autres agents de s'en déclarer membre (nous parlons de « passagers »), ce qui confère à ces derniers des compétences associées au composant qui les accueille. Ce point concerne également toutes les organisations qui agencent des composants les uns par rapport aux autres ; il sera développé dans la partie 3.1.
- Les organisations qui structurent socialement des agents (par exemple des individus sociaux), par des liens qui expriment des relations pouvant être de natures variées : liens mono ou bidirectionnels, liens étiquetés ou non (texte, pondération), etc.

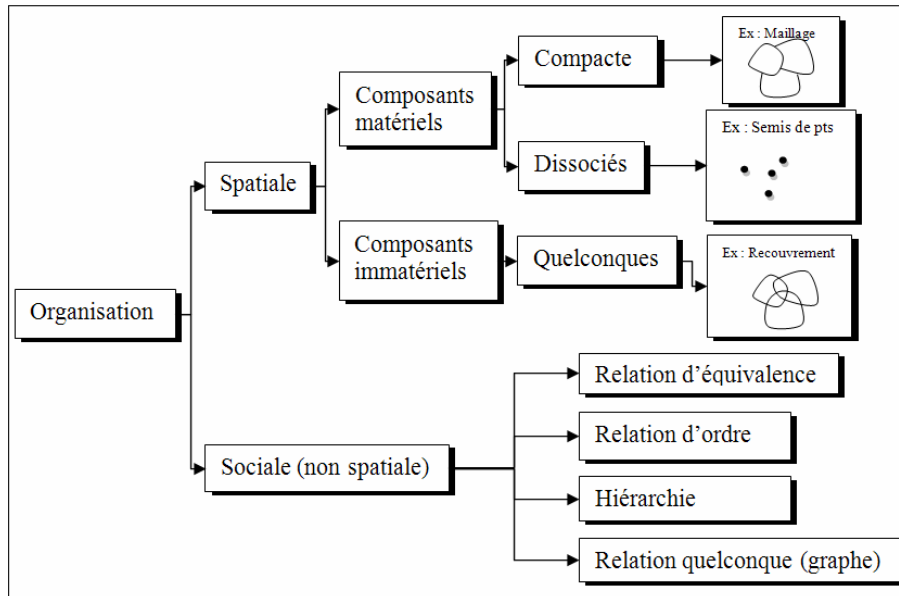


Figure 1. Hiérarchie des organisations

2.4. Compétence, comportement et dynamique

La dynamique du système résulte du comportement de ses agents qui agissent et interagissent entre eux et avec leur environnement.

La dynamique d'un agent est l'expression de ses compétences au cours du temps, (c'est-à-dire le processus d'exécution de ses compétences). Cette dynamique sera appelée comportement de l'agent.

Le comportement d'un agent peut se manifester de différentes manières (non exclusives entre elles) :

- Affecter son support spatial : si celui-ci subit une transformation isométrique qui, à chaque pas de temps, effectue une translation suivie d'une rotation, on dit que l'agent est mobile (son support reste invariant par rapport à un repère local attaché à l'agent). Si ses attributs spatiaux subissent un autre type de transformation géométrique, on dit que l'agent est déformable.
- Lorsque la transformation s'opère sur des attributs non spatiaux de l'agent (internes ou relationnels) on parle de dynamique évolutive (interne ou relationnelle) de l'agent.
- Lorsque des agents ou des objets (matière, flux, énergie, etc.) se transforment au cours du temps en d'autres agents ou d'autres objets,

comme des réactions chimiques, des dynamiques foncières (remembrement), ou des possibilités de reproduction ou de réplication, on parlera de dynamique transformationnelle.

Une compétence peut être innée (permanente), lorsqu'elle est constitutive de l'agent, mais peut également être acquise (temporaire), lorsqu'elle lui est offerte par une organisation (ou un de ses composants) dont il est passager à un instant donné ; l'agent peut ainsi adapter son comportement à son environnement social et spatial. Par exemple, dans une organisation qui décrit la « structure familiale d'un ménage », le composant « père » dispose de compétences spécifiques à ce rôle, ce qui permet à un agent passager de ce composant, d'acquérir, en plus de ses compétences innées, ces compétences temporaires. Mais ensuite, ce même agent humain partant travailler, endossera le rôle d'automobiliste et adoptera ainsi les compétences liées à la conduite automobile, afin de se déplacer sur l'organisation spatiale de type réseau routier.

Le comportement global d'un agent est en général la combinaison de processus plus simples, que l'on appelle comportements élémentaires, et qui sont l'expression (l'exécution en informatique) de compétences élémentaires, celles-ci étant des transformations de l'agent en fonction de son état présent et de celui de son environnement.

Formalisation de compétences simples

Notre modèle de compétence part du principe qu'à un instant donné, il peut, ou non, être pertinent d'exécuter une compétence donnée. Tout agent possède des compétences innées ou acquises, qu'il exécute ou pas en fonction du contexte. Selon le contexte social, spatial ou temporel, il sera pertinent ou non d'exécuter une *action* en particulier. Dans le modèle comportemental proposé, la pertinence liée à l'environnement présent sera exprimée par une clause *pertinence* associée à chaque compétence. Cette clause est une fonction qui renvoie une valeur réelle entre 0 et 1 à partir d'une évaluation de l'état de l'agent et de son environnement.

Les bases de cette approche ont été développées dans le modèle d'agents cellulaires décrit dans (Dubos-Paillard et al., 2003), (Langlois, 2010), basé sur le couple *compétence* = \langle *pertinence, action* \rangle . Nous y avons ajouté un troisième terme nommé *satisfaction*, pour obtenir une modélisation de compétence sous forme d'un triplet :

$$\textit{compétence} = \langle \textit{pertinence}, \textit{action}, \textit{satisfaction} \rangle.$$

La notion de pertinence d'une compétence permet de dépasser le simple conditionnement de l'exécution d'une action à une évaluation logique (de type SI pertinent ALORS action). En effet, la pertinence d'une compétence a pour but d'être évaluée et comparée à celles des autres compétences qu'un agent pourrait utiliser à un instant donné. Si plusieurs compétences sont en concurrence à un instant donné, seule la plus pertinente verra son action exécutée par l'agent.

L'action décrit les opérations exécutées par l'agent, autant sur ses attributs que sur son environnement.

La satisfaction est un nombre réel entre 0 et 1 représentant l'évaluation de la bonne exécution de l'action associée. Elle vaut zéro si celle-ci n'a pas été exécutée. Elle peut servir de pertinence pour une action suivante, lors de l'enchaînement séquentiel de deux comportements ; elle peut également permettre de quantifier l'état d'avancement vers la réalisation d'un objectif.

Formalisation de compétences complexes

La clause pertinence offre donc la possibilité de choisir, parmi une liste de compétences, celle qu'il convient d'exécuter à un instant donné. Le choix le plus naturel consiste à choisir la plus pertinente en cet instant (on note cette opération C_1 / C_2), mais on peut également choisir une compétence tirée au hasard ($C_1 ? C_2$), avec éventuellement une pondération par la pertinence.

D'autres opérations permettent de faire se succéder l'exécution de plusieurs compétences dans le temps : la plus classique est l'exécution séquentielle d'une série de compétences élémentaires. Dans le cas de deux compétences C_1 et C_2 , leur séquence s'écrit « $C_1 ; C_2$ » et signifie simplement qu'à l'exécution de C_1 succèdera celle de C_2 .

D'autres possibilités sont l'exécution en parallèle de plusieurs compétences (la marche met en jeu de nombreux mouvements élémentaires simultanés), ou encore l'exécution séquentielle avec un *GO/NOGO* après l'exécution de chaque compétence : c'est la séquence conditionnelle « $C_1 ; > C_2$ », qui signifie qu'après l'exécution de C_1 , la compétence C_2 ne sera examinée que si la satisfaction de C_1 n'est pas nulle ; C_2 ne sera donc jamais exécutée si l'action de C_1 ne l'a pas été.

Il est possible d'ordonner les compétences de plusieurs autres façons, par exemple en séquence dans l'ordre décroissant des pertinences (exécution de plusieurs tâches avec un ordre de priorité), ou encore en exécutant une compétence autant de fois qu'il est nécessaire, soit jusqu'à ce qu'une autre compétence devienne pertinente (un agent se déplace jusqu'à ce qu'il ait trouvé un objet à ramasser), soit jusqu'à ce qu'un événement se produise (un agent se déplace jusqu'à ce qu'on lui dise « stop »).

3. Relations entre les concepts de AOC

3.1. La relation de positionnement / localisation

Pour modéliser le fait qu'un ensemble A d'agents qui sont les composants d'une organisation, se localisent et se déplacent dans un environnement lui-même formé d'un ensemble d'agents B composants d'une autre organisation, on utilise une fonction de positionnement de A vers B , qui associe à tout agent de A un agent au

plus dans B . Les liens de cette relation sont des couples (passager, hôte) de la forme $(a, b) \in A \times B$, qui s'expriment par une phrase du type « le passager a est sur l'hôte b ». Si de plus le composant b est un « conteneur spatial » le passager a est alors localisé relativement à ce composant (qu'on appelle alors son localisateur). Par exemple, si b est un tronçon de route (élément linéaire), a peut être localisé par son abscisse curviligne relative à b . Si b est un carreau (élément surfacique) a peut être localisé par deux coordonnées relativement à b .

Par exemple, un agent « wagon de voyageurs » peut être modélisé en lui adjoignant une organisation spatiale composée de sièges et d'un couloir qui sont les composants de l'organisation du wagon, par abus de langage on dira simplement que ce sont les composants du wagon. Les voyageurs sont positionnés dans le wagon, comme passagers, chacun étant localisé sur un siège ou sur le couloir, qui sont les hôtes. Les passagers peuvent éventuellement se déplacer dans le wagon, selon certaines règles. A une autre échelle, un wagon peut être composant d'un train, si l'on considère qu'il est fixe dans la composition de ce train, même si le train est mobile (avec ses wagons et ses voyageurs) sur la ligne qu'il emprunte (il est alors passager sur un tronçon de ligne du réseau ferré). La notion de composant et de passager est donc relative à une organisation.

Les passagers sont, via une relation de localisation, les usagers de l'organisation. Ils interagissent avec elle et, éventuellement, la font évoluer. Un agent peut être l'hôte de plusieurs passagers, mais un agent ne peut être passager que d'un unique composant hôte d'une couche d'organisation. Un agent peut en revanche être passager de plusieurs organisations à la fois. Ainsi, un homme sera passager d'un service d'une entreprise, d'une CSP, et d'une maille spatiale. Un agent peut également être passager de plusieurs couches d'une même organisation, par exemple une maison est à la fois dans une commune, un département et une région de l'organisation administrative du territoire français.

3.2. Mobilité, processus d'interaction

L'organisation possède des méthodes et des données de gestion collective. Elle est capable d'autoriser ou refuser l'arrivée, le départ, la création ou la suppression d'un nouvel agent comme passager sur un hôte. Elle gère la mobilité des passagers sur ses composants-hôtes et d'une manière générale les interactions.

La dynamique de la relation de localisation définit le processus de mobilité des agents à travers les compétences offertes aux passagers par l'organisation-hôte. Ainsi, la mobilité apparaît comme un exemple simple (et fréquent) de processus d'interaction, dans lequel un agent a passé d'un lieu i à un lieu j au cours d'un pas de temps élémentaire. Il y a donc interaction entre les lieux i et j qui échangent simultanément leur contenu.

D'une manière générale un processus d'interaction est défini sur un ensemble d'agents A (appelés hôtes dans (Langlois et al., 2007) sous le contrôle d'un ensemble M de médiateurs. Par exemple la collision entre deux boules de billard, la

prédation d'un loup sur un mouton, la transmission d'une maladie d'une personne infectée à une personne saine, sont des interactions élémentaires qui entrent dans des processus d'interaction. Ainsi, un processus d'interaction est composé de trois entités, $P = \langle A, M, f \rangle$. A est l'ensemble des agents (A peut être hétérogène : les loups et les moutons, ou homogène : les boules de billard). M est l'ensemble des médiateurs de l'interaction à travers lesquels s'effectuent les interactions élémentaires (M est appelé parfois le vecteur). La fonction f représente ici de manière théorique le processus, c'est-à-dire le mécanisme d'évolution des états des agents, lorsque les médiateurs opèrent sur eux. Les médiateurs peuvent être représentés par les « particules » (vision discrète) qui sont échangées ou par le flux ou le champ (vision continue) qui s'applique lors d'une interaction élémentaire entre deux agents de A . Par exemple, dans l'interaction des boules de billard, le médiateur est la quantité de mouvement échangée sur le lieu de la collision. Dans la prédation, la proie disparaît au profit de l'énergie gagnée par le (ou les) prédateur(s). L'épidémie se propage entre les individus par transmission et multiplication du vecteur (bactérie ou virus). Comme pour la mobilité, c'est encore l'organisation des agents de A qui gère collectivement l'interaction.

3.3. Modèles multi-niveaux : Emboîtement agent – organisation – agents

L'organisation apparaît toujours comme un intermédiaire entre un agent global et ses agents internes. Un agent n'est jamais considéré comme contenant directement une collection d'agents, mais comme contenant une ou plusieurs organisations, qui gèrent des composants internes et des passagers.

Un agent est dit terminal (ou simple ou élémentaire), s'il ne contient pas d'organisation interne, et par conséquent pas de composant interne. Dans le cas contraire il est non-terminal (ou composé ou complexe). L'agent est alors un système composé d'organisations d'agents plus élémentaires en interaction. Il possède une double description, externe en tant qu'agent, et interne en tant que système formé de ses composants internes.

Cette approche permet une modélisation multi-échelles d'un modèle, par raffinements successifs à partir d'un agent qui englobe tous les autres. On peut ainsi envisager de modéliser des systèmes multicouches, tels que les maillages géographiques sur plusieurs échelles (par exemple, pays-région-département-canton).

L'agent qui englobe tous les autres peut être appelé agent-système. Son organisation interne décrit les entités et la dynamique du modèle. Ses entrées sont connectées à l'interface homme-machine, permettant à l'utilisateur de paramétrer le modèle et d'interagir en cours de simulation. Ses sorties permettent de visualiser et/ou de mémoriser les observations opérées sur le système au cours de la simulation.

4. Schéma ontologique

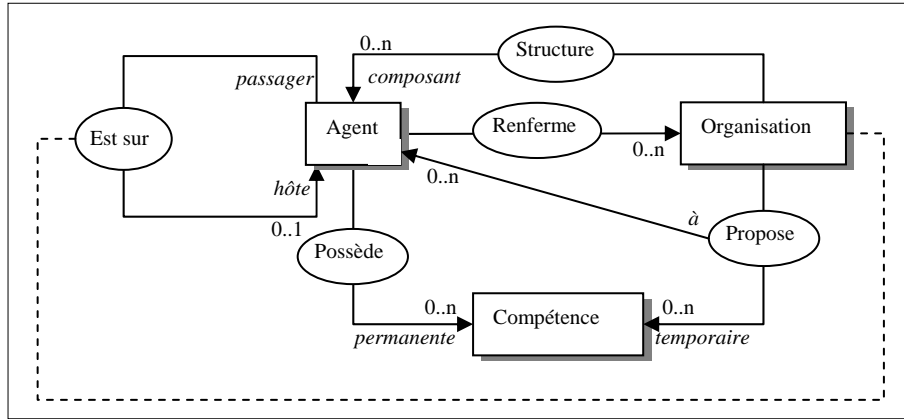


Figure 2. Schéma simplifié du paradigme AOC

Le schéma de la Figure 2 résume les principales relations entre les concepts d'agent, d'organisation et de comportement : une organisation structure un ensemble d'agents qui sont alors composants de l'organisation. Mais un agent complexe peut aussi renfermer en son sein des organisations qui structurent ses constituants internes. Une compétence peut être utilisée de manière permanente par un agent, ou bien être proposée par une organisation de manière temporaire aux passagers. Une relation de positionnement/localisation « est sur » est une fonction définie entre deux ensembles d'agents, les passagers et les hôtes. Les contraintes de localisation et de mobilité entre les passagers sont gérées par l'organisation des hôtes.

5. Application

Le modèle des termites (Wilensky, 1997) a été choisi pour être modélisé avec l'approche AOC. Ce modèle peut être décrit de la manière suivante : des termites se déplacent sur un espace discret formé d'un carroyage, une case ne pouvant contenir plus d'un termite à la fois. Des grains sont disposés dans les cases (un grain au plus par case). Les termites ramassent les grains qu'elles trouvent, les portent au cours de leur déplacement et les déposent à proximité d'autres grains. Un grain est donc, à un instant donné, soit dans une case, soit porté par un termite. A l'initialisation tous les grains sont dans des cases.

Montrons comment ce modèle peut être formalisé dans le paradigme AOC.

Agents

- Agent *Case* : possèdera un attribut booléen, *YaGrain*, indiquant s'il y a un grain dans la case ;
- Agent *Termite* : possèdera un attribut booléen, *PorteGrain*, indiquant si le termite porte un grain.

Ici les Termites, ainsi que les Cases, sont des agents terminaux.

Organisations

Terrain : Carroyage de *Cases* (par exemple un tore avec voisinage de Moore).

Relation de localisation des *Termites* vers les *Cases* : un termite est passager d'une case et une seule, et il ne peut y avoir plus d'un termite sur une case.

Compétences

Les seuls agents ayant une dynamique propre sont les termites. Cette dynamique se résume aux trois compétences suivantes :

- *SeDéplacer* : le termite se déplace vers une case voisine choisie au hasard ;
- *Prendre* : dans le cas où sa case contient un grain, le termite le ramasse ;
- *Déposer* : dans le cas où le termite porte un grain et se trouve sur une case vide, s'il y a au moins un grain dans son voisinage, alors il dépose le grain qu'il porte.

Traduisons ces compétences sous forme de triplets $\langle \textit{Pertinence}, \textit{Action}, \textit{Satisfaction} \rangle$:

$\textit{SeDéplacer} = \langle P_1, A_1, S_1 \rangle$

$P_1 = 1$ s'il y a au moins une case voisine sans termite, 0 sinon.

$A_1 =$ Se déplacer vers une case sans termite du voisinage choisie au hasard.

$S_1 = 1$ lorsque l'on s'est déplacé, 0 sinon.

$\textit{Prendre} = \langle P_2, A_2, S_2 \rangle$

$P_2 = 1$ si le termite ne porte aucun grain et que sa case en contient un, 0 sinon.

$A_2 =$ Le grain passe de la case au termite.

$S_2 = 1$ lorsque un grain est passé au termite, 0 sinon.

$Déposer = \langle P_3, A_3, S_3 \rangle$

P_3 = Le termite porte un grain, il n'y en a pas dans sa case, et il y en a au moins un dans une de ses cases voisines.

A_3 = Le grain passe du termite à la case.

S_3 = 1 lorsque le grain est déposé, 0 sinon.

Ces compétences doivent maintenant être ordonnancées dans le temps, à l'aide des opérations décrites dans la partie 2.4. Nous proposons deux façons de procéder, mais d'autres approches sont évidemment possibles.

Première approche

Lors d'un pas de temps, un termite peut soit prendre, soit déposer (s'il le peut). Cette première action est nécessairement suivie d'un déplacement (si celui-ci est possible). Le comportement d'un termite peut donc être formulé de la façon suivante :

$(Prendre \mid Déposer) ; SeDéplacer$

Il se peut qu'un termite ne puisse ni prendre ni déposer, dans ce cas le termite va juste se déplacer, si son environnement le permet. S'il ne peut pas se déplacer, il restera bloqué tant que les termites de son voisinage ne lui auront pas laissé de place.

On vérifie facilement que la succession de plusieurs pas de temps va permettre à un termite de réaliser l'ensemble d'un cycle ($SeDéplacer, Prendre, SeDéplacer, Déposer$).

Seconde approche

Une autre façon de formaliser ce processus consiste à réaliser l'ensemble du cycle ($SeDéplacer, Prendre, SeDéplacer, Déposer$) au cours d'un unique pas de temps : le termite (à vide) commence par se déplacer jusqu'à arriver sur une case où il pourra ramasser un grain, puis il se déplace jusqu'à arriver sur une case où il pourra déposer son grain :

JUSQUA $Prendre$ REPETER $SeDéplacer$;

JUSQUA $Déposer$ REPETER $SeDéplacer$;

La compétence *SeDéplacer* va ainsi, par exemple dans la première ligne, être exécutée plusieurs fois, jusqu'à ce que la compétence *Prendre* devienne pertinente, et soit exécutée.

Lorsqu'il vient de déposer, à la fin d'un cycle, le termite recommence le cycle suivant en se déplaçant au moins une fois avec de prendre. Il ne peut donc pas ramasser à nouveau le grain qu'il vient de déposer, à moins que son environnement l'ait empêché de se déplacer.

6. Conclusion

La vision systémique et complexe du monde conduit les chercheurs à investir l'univers des méthodologies individus-centrées, tels que les automates cellulaires et les systèmes multi-agents. Dans ces univers complexes, une grande variété de facteurs sociaux, économiques, naturels et géographiques interagissent et affectent la dynamique des objets étudiés. Que ce soit en biologie, en économie ou en géographie, cette prise en compte de la complexité des relations entre des entités d'un même niveau et de niveaux différents sert de levier au besoin de modèles de simulation informatique, afin de comprendre et prévoir les évolutions possibles de tels systèmes.

Le paradigme AOC permet au chercheur de penser et construire des modèles multi-niveaux et dynamiques. Cette possibilité est offerte grâce à la récursivité de la notion d'agent qui autorise l'emboîtement des échelles sociales et géographiques. Les dynamiques qui s'exercent au sein de ces modèles peuvent être aussi bien pensées au travers d'une modélisation de compétences propres aux organisations et à leurs composants qu'au niveau des agents eux-mêmes. A l'emboîtement des échelles géographiques et des niveaux d'organisation répond l'emboîtement des compétences. Ainsi des compétences complexes, évolutives, peuvent être construites à partir de compétences élémentaires.

Le paradigme AOC est formalisé dans un schéma XML qui permet d'écrire des modèles sous forme de documents XML, indépendamment de toute plateforme de simulation. Cette méthode permet ainsi la communication et le partage de modèles au sein d'une communauté de modélisateurs. La plateforme MAGéo, qui intègre ce schéma XML, permet une modélisation relativement simple basée sur de l'interaction graphique et des règles non algorithmiques, adaptée à des utilisateurs non informaticiens.

7. Références bibliographiques

Dubos-Paillard E., Guermond Y., Langlois P., 2003. Analyse de l'évolution urbaine par automate cellulaire : le modèle SpaCelle. *L'espace géographique*, vol 4, p. 357-378.

- Ferber J., 1995. *Les systèmes multi-agents. Vers une intelligence collective*. InterEditions, Paris.
- Ferber J., Michel F., Baez J., 2005. AGRE: Integrating Environments with organizations. In *E4MAS'04*, Australie, p. 127-134.
- Langlois P., Daudé E., 2007, Concepts et modélisation de la diffusion, *Cybergeo*, n°. 364, <http://cybergeo.revues.org/index2898.html>, 23p.
- Langlois P., 2010. *Simulation des systèmes complexes en géographie, fondements théoriques et applications*. Hermès-Lavoisier, Paris.
- Phan D., Amblard F., 2007. *Agent-based modelling and simulation in the social and human sciences*.
- Ramat E., 2006. Introduction à la modélisation et à la simulation à événements discrets. In Phan D.& Amblard F., 2007, *Agent-based modelling and simulation in the social and human sciences*, p. 87-103.
- Wilensky, U., 1997. NetLogo Termites model. <http://ccl.northwestern.edu/netlogo/models/Termites>. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL.

TaxoMap Framework appliqué à l’alignement d’ontologies géographiques dans le projet *GéOnto*

Fayçal Hamdi, Chantal Reynaud, Brigitte Safar

CNRS - Université de Paris-Sud 11 (LRI)
INRIA Saclay Ile-de-France (LEO)
Parc Orsay Université 4 rue Jacques Monod
91893 Orsay France

Faycal.Hamdi, Chantal.Reynaud, Brigitte.Safar@lri.fr

RÉSUMÉ. Les outils actuels d’alignement d’ontologies ne sont pas performants sur tous les domaines ni quelles que soient les ontologies. La qualité des mappings obtenus peut être améliorée si le processus d’alignement tient compte des spécificités des ontologies alignées. L’environnement TaxoMap Framework a été conçu pour aider un expert du domaine à spécifier des traitements sur les mappings produits par l’outil d’alignement TaxoMap, afin de les raffiner ou bien de fusionner, restructurer ou enrichir des ontologies. Nous proposons une démonstration de deux modules implémentés dans cet environnement : le module d’alignement et le module de raffinement.

ABSTRACT. The current ontology alignment tools are not efficient on all domains and whatever the ontologies to be aligned. The quality of the results can be improved if the alignment process take into account the specificities of the ontologies. The environment TaxoMap Framework was conceived to help a domain expert to specify treatments, based on mappings generated by the alignment tool TaxoMap, in order to refine or merge, restructure, or enrich ontologies. We propose a demonstration of two modules implemented in this environment: the alignment module and the refinement module.

MOTS-CLÉS : Alignement d’ontologies, raffinement d’ontologies

KEYWORDS: Ontology alignment, ontology refinement

1. Le module "Alignement"

TaxoMap est un logiciel d'alignement de taxonomies développé au sein de l'équipe IASI du LRI (Université Paris-Sud). L'outil a été conçu pour découvrir des alignements entre des ontologies $O = (C, H)$ exprimées en OWL et supposées décrire le même domaine d'application. C est un ensemble de concepts où chaque concept est caractérisé par un ensemble de labels et H est une hiérarchie de subsomption qui contient un ensemble de relations *isA* établis entre les nœuds correspondants aux concepts. Le processus d'alignement est un processus orienté qui cherche à relier par un *mapping* chaque concept d'une ontologie source à un unique concept de l'ontologie cible.

La découverte des mappings repose sur des techniques variées, terminologiques ou structurelles, qui s'appuient sur l'utilisation de l'analyseur morpho-syntaxique Tree-Tagger [SCH 94]. Celui-ci permet un paramétrage en fonction de la langue, une lemmatisation et une catégorisation des mots qui composent un label. Les sorties de l'analyseur sont utilisées par une mesure de similarité appliquée aux labels des concepts vus comme des ensembles de tri-grammes [LIN 98]. Pour chaque concept c_s de l'ontologie source O_S , la mesure de similarité identifie les concepts de l'ontologie cible O_T , qui ont les meilleures similarités et sont candidats au mappings avec c_s . Les techniques d'alignement appliquées en séquence permettent ensuite de déterminer le concept le plus pertinent parmi les candidats et quelle type de relation (relations d'équivalence, de subsomption ou de proximité) doit être établie entre c_s et ce concept.

2. Le module "Raffinement"

Les tests effectués sur les taxonomies mises à disposition par notre partenaire le COGIT de l'IGN dans le projet *GéOnto*, ont montré que le module d'alignement fournissait de très bons résultats (précision de 92,3%) mais que ces derniers pouvaient encore être améliorés.

Une étude des traitements d'améliorations souhaitées par les experts a montré que celles-ci étaient souvent spécifiques aux ontologies alignées. Pour ne pas faire de *TaxoMap* un outil uniquement dédié à l'alignement de taxonomies topographiques, nous avons développé un environnement permettant aux experts du domaine de spécifier et de mettre en oeuvre des traitements supplémentaires sur des alignements préalablement obtenus. Dans cette démonstration, l'environnement est utilisé pour raffiner l'alignement construit par *TaxoMap*. Il peut aussi être utilisé pour tout autre traitement prenant appui sur les résultats d'un alignement entre ontologies, comme la fusion ou l'enrichissement d'ontologies.

L'environnement permet une spécification déclarative des traitements basés sur les résultats d'un alignement particulier et concernant des ontologies particulières, en utilisant un ensemble de primitives génériques et prédéfinies. Les traitements sont spécifiés sous la forme de pattern composés d'une partie condition et d'une partie action. La partie condition s'exprime au travers d'un ensemble de primitives (prédicats)

permettant d'expliciter l'ensemble des caractéristiques qui doivent être vérifiées pour l'application du pattern. Les caractéristiques peuvent être des relations structurelles entre des concepts ou des relations terminologiques sur leur labels ou d'autres relations.

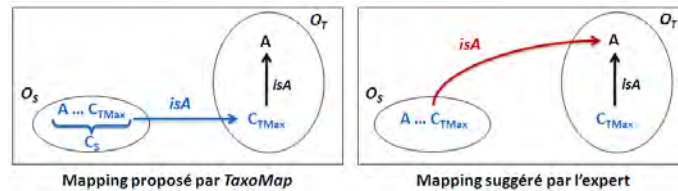


Figure 1. Illustration d'un Pattern

La partie contexte du Pattern :

$$\exists x \exists y (isAStrictInclusion(x, y) \wedge \exists z (isSubClassOf(y, z, O_T) \wedge strictInclusionLabel(z, x)))$$

La partie solution du Pattern :

$$Delete_Mapping(x, y, _) \wedge Add_Mapping(x, z, isA)$$

Les traitements qui peuvent être spécifiés dépendent des caractéristiques des ontologies concernées et de la tâche à exécuter (raffinement d'alignement ou enrichissement d'ontologies par exemple). Ces traitements sont associés à des modules de spécification indépendants, un pour chaque tâche, chacun ayant leur propre vocabulaire. Le module est extensible et a priori applicable à n'importe quel traitement basé sur des résultats d'alignement.

3. Bibliographie

- [LIN 98] LIN D., « An Information-Theoretic Definition of Similarity », *In Proceedings of the 15th International Conference on Machine Learning*, 1998, p. 296–304.
- [SCH 94] SCHMID H., « Probabilistic Part-of-Speech Tagging Using Decision Trees », *Proceedings of the International Conference on New Methods in Language Processing*, 1994, p. 44–49.

Vers un système en ligne d'aide à la découverte du contenu de bases de données géographiques

Mechouche Ammar, Abadie Nathalie, Prouteau Emeric, Mustière Sébastien

Laboratoire COGIT, Institut Géographique National

73 Avenue de Paris

94160 Saint-Mandé, FRANCE

ammar.mechouche@ign.fr, nathalie-f.abadie@ign.fr, emeric.prouteau@ign.fr, sebastien.mustiere@ign.fr

RESUME. Nous présentons un prototype d'application Web proposant à un utilisateur de découvrir le contenu de bases de données géographiques. Cette application repose sur des ontologies pour décrire et annoter les bases de données décrites par le système. Cette application inclut une solution de cartographie en ligne pour afficher les données.

ABSTRACT. We present a Web application prototype enabling any user to discover geographic databases' content. It uses ontologies to describe and annotate the different geo-databases considered by the system. A Web mapping application is added to display geo-data samples.

MOTS-CLÉS: Bases de données géographiques, spécifications, ontologie, IHM, cartographie en ligne.

KEYWORDS: Geographic databases, specifications, ontology, user interface, Web mapping.

1. Introduction

Il peut être difficile d'appréhender la complexité des diverses sources de données géographiques désormais disponibles (Craglia et al., 2008). Nous présentons ici un prototype d'application Web proposant à un utilisateur de découvrir le contenu des bases de données de l'IGN en lui permettant de savoir comment et où sont représentées les données qui répondent à son besoin.

2. Architecture du système

Notre système est composé de trois modules principaux. Les choix techniques effectués pour l'implémentation du système sont présentés en figure 1.

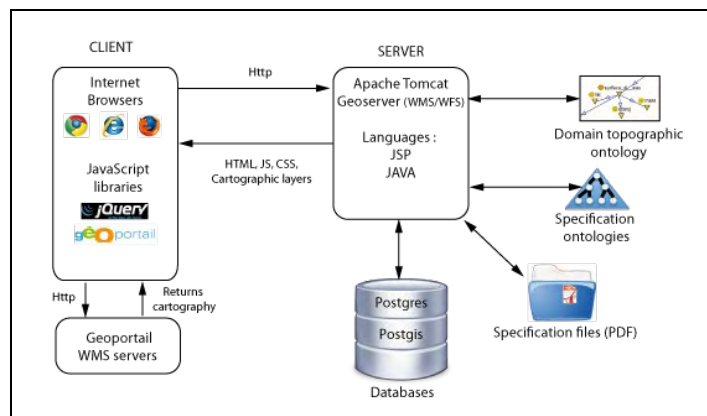


Figure 1: Implémentation du système.

Le module de recherche aide l'utilisateur à exprimer sa requête en lui proposant, à l'aide d'une application d'auto-complétion, d'interroger le système à l'aide d'un vocabulaire unifié fourni par une ontologie globale du domaine de la topographie. Nous utilisons une ontologie bilingue (Français et Anglais), de plus de 700 concepts, créée à partir des textes de spécifications de bases de données géographiques à l'aide d'outils de traitement automatique du langage naturel (Abadie et al., 2008).

Le module d'extraction d'informations recherche les données disponibles correspondant aux termes de la requête de l'utilisateur, en utilisant des ontologies locales de spécifications qui formalisent les spécifications de contenu de chacune des bases de données considérées (Abadie et al., 2010). Chaque classe du schéma d'une base de données y est traduite en OWL et annotée à l'aide des concepts issus de l'ontologie du domaine de la topographie. Des connaissances supplémentaires tirées de ses spécifications, comme des critères de sélection du type « *La classe 'Rivière' comprend tous les cours d'eau permanents de plus de 10 mètres de large* », sont formalisées et ajoutées dans l'axiome utilisé pour l'annoter. Toutes ces connaissances concernant des classes annotées par le concept de l'ontologie globale recherché par l'utilisateur sont ainsi récupérées et renvoyées à l'utilisateur.

Le module cartographique affiche sous forme de cartes les données identifiées par le module d'extraction d'informations afin de permettre à l'utilisateur de visualiser simplement les diverses données disponibles correspondant au thème qui l'intéresse.

3. Conclusion

L'interface Web, composée de trois parties, est présentée en figure 3 : Une zone de texte (en haut) pour saisir les requêtes des utilisateurs, des onglets (à gauche),

correspondant chacun à une base de données et servant à afficher les informations envoyées par le système sur les données issues de cette base et qui correspondent à la requête de l'utilisateur. La troisième, à droite de la page, comporte l'affichage cartographique des données.

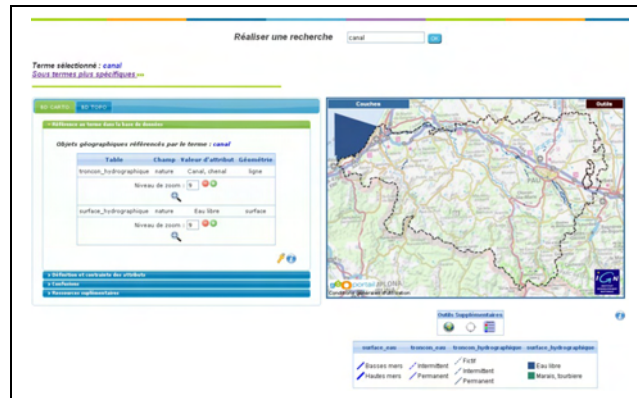


Figure 1: Interface Web du prototype implémenté.

4. Remerciements

Cette recherche a été réalisée dans le cadre du projet GeOnto, financé par l'Agence Nationale de la Recherche (ANR-O7-MDCO-005).

5. Bibliographie

Abadie N. and Mustière S., « Constitution d'une taxonomie géographique à partir des spécifications de bases de données ». *Colloque International de Géomatique et d'Analyse Spatiale, SAGEO'08*, Montpellier, France, 2008.

Abadie N., Mechouche A., Mustière M., « OWL-based formalization of geographic databases specifications ». *17th International Conference on Knowledge Engineering and Knowledge Management, EKAW'10*, Lisbonne, Portugal, 11-15 octobre 2010.

Craglia M, Goodchild M, Annoni A, Camara G, Gould M, Kuhn W, Mark D, Masser I, Maguire D, Liang S, Parsons E. « Next-Generation Digital Earth - A Position Paper from the Vespucci Initiative for the Advancement of Geographic Information Science ». *International Journal of Spatial Data Infrastructures Research*. 3; 2008. p. 146-167. JRC47746 1.5 Article contribution to other periodicals.