

Rapport semestriel d'activité - coordonnateur
Programme MDCO - Edition 2007
Projet GEONTO – 2ème semestre 2010

Identification

Acronyme du projet	GEONTO
Numéro d'identification de l'acte attributif	ANR-07-MDCO-05
Coordonnateur (société/organisme)	LRI – Université Paris-Sud
Période couverte (date à date)	01/07/2010 – 31/12/2010
Période couverte (t0+n mois à t0+m mois)	T0+30 à T0+36
Rédacteur (nom, téléphone, email)	Chantal Reynaud, 01 72 92 59 87 Chantal.reynaud@lri.fr
Date	2 février 2011

URL de la page web du projet et date de dernière mise à jour

<http://geonto.lri.fr>

La dernière mise à jour a été effectuée le 2 février 2011.

Activités de coordination des activités du projet

(lister les réunions, visites, ...)

Réunion plénière :

18/11/2010 : réunion de bilan semestre 6 – LRI, COGIT, IRIT, LIUPPA - Orsay

Réunions de travail par lot :

Nous n'avons pas eu de réunions par lot. Le travail de chacun a consisté à poursuivre les développements précédemment spécifiés. Les échanges entre partenaires se sont faits essentiellement par courrier électronique ou téléphone.

Synthèse

Numéro du Partenaire	Conformité des résultats obtenus aux prévisions	Conformité de la consommation des ressources par rapport aux prévisions	Difficultés particulières
1	Légèrement inférieure aux prévisions mises à jour dans la demande de prolongation du projet.	Supérieure. Un stagiaire ingénieur 1 ^{ère} année a été financé sur un autre contrat.	Alignement non directement applicable sur l'ontologie construite automatiquement nécessitant de revoir l'approche pour aboutir à l'ontologie topographique visée. Retard dans la livraison du logiciel de réconciliation d'instances pour la réconciliation d'ontologies du fait de

			l'embauche d'un ingénieur du 1/04/10 au 15/07/10 puis d'un ingénieur 1ère année ¹ (stage) du 1/10/10 au 01/02/11.
2	conforme	Légèrement inférieure aux prévisions (fin anticipée de 1,5 mois du contrat de post-doc engagé sur budget ANR)	aucune
3	Conforme au nouveau calendrier	Conforme aux nouvelles prévisions	Pas de difficulté particulière
4	conforme	Conforme (compte tenu du décalage précédemment annoncé en terme de recrutement du doctorant)	aucune
Synthèse	Légèrement inférieure aux prévisions mises à jour dans la demande de prolongation du projet.	Légèrement supérieure. Un stagiaire ingénieur 1 ^{ère} année a été financé sur un autre contrat.	Alignement non directement applicable sur l'ontologie construite automatiquement nécessitant de revoir l'approche pour aboutir à l'ontologie topographique visée. Retard dans la livraison du logiciel de réconciliation d'instances pour la réconciliation d'ontologies du fait de l'embauche d'un ingénieur du 1/04/10 au 15/07/10 puis d'un ingénieur 1ère année ² (stage) du 1/10/10 au 01/02/11.

Faits marquants

Indiquer les résultats et/ou réalisations marquants. Préciser s'ils peuvent ou non faire l'objet de communications externes par l'ANR et la Délégation ANR-CI.

Les travaux du lot 1 ont porté sur les aspects suivants :

- Sous-lot 1.1 : mise au point d'outils d'extraction de concepts et de relations

La chaîne de traitement requise pour analyser le langage naturel présent dans les documents de spécification en utilisant la plateforme *LinguaStream* a été finalisée. Elle exploite un lexique du domaine construit à partir du corpus, des patrons spécifiant l'expression linguistique des relations sémantiques et prenant aussi en compte la structure du document, et le texte présent dans les définitions ou d'autres parties du corpus. Elle permet d'identifier des termes pouvant donner lieu à des concepts, ainsi que des relations entre ces termes, pouvant donner lieu à des propriétés de concepts ou des relations entre concepts. Elle fournit une représentation en OWL de propositions de concepts et de relations. Les différentes expressions de concepts et de relations sont maintenant couvertes de manière exhaustive. Les concepts et relations trouvés ont été évalués de manière quantitative et qualitative.

Le noyau d'ontologies construit automatiquement doit être étudié par le COGIT qui intégrera manuellement dans *TopoCarto_Cogit* les propriétés provenant des définitions des spécifications. L'enrichissement à partir des concepts est à automatiser : Etude de l'existence des termes dans le thesaurus Rameau et enrichissement de *TopoCarto_Cogit* à l'aide de *TaxoMap FrameWork*.

¹ Ce stage a été financé sur un autre contrat.

² Ce stage a été financé sur un autre contrat.

- Sous-lot 1.2 et 1.3 : Enrichissement et restructuration d'une ontologie existante

La version définitive de la chaîne de traitement « AugmOnto » développée par le LIUPPA qui alimente le processus d'enrichissement est maintenant opérationnelle. Les sorties sont fournies en SKOS (pour les besoins de l'équipe de l'IRIT) ou RDF ou OWL (pour les besoins de l'équipe du LRI). Une publication a été acceptée à la conférence internationale GeoS 2011.

L'IRIT a implémenté une partie des principes définis au S5 pour guider l'enrichissement de l'ontologie à partir des termes trouvés dans le corpus grand public par le LIUPPA. Les termes sont situés par rapport à des entrées du thesaurus Rameau. Les propositions d'enrichissement sont disponibles dans un fichier contenant des fragments d'ontologie composés d'un concept, de son concept générique et des concepts plus spécifiques, ainsi que des termes équivalents. Dans chaque fragment, on cherche à reconnaître au moins un des concepts de l'ontologie BD-Topo pour y intégrer ce fragment. Un algorithme de parcours et d'analyse de fichiers contenant l'ensemble des propositions d'enrichissement a été implémenté. La poursuite des travaux sur l'enrichissement a donné lieu à la proposition d'un stage de M2 Recherche à l'IRIT.

En parallèle, le LRI a également travaillé sur les sorties d' « AugmOnto » développé par le LIUPPA et a défini des patterns de validation de compatibilité de domaine autorisant ou non l'enrichissement. Les résultats doivent donner lieu au livrable n°9 envoyé en mars 2011. Une méthodologie d'enrichissement d'ontologies à partir de très volumineuses ontologies généralistes a également été étudiée. Des expérimentations sont en cours avec l'ontologie Yago. Les premiers travaux sur l'enrichissement réalisés par le LRI ont été publiés à EGC 2011.

Ces travaux sur l'enrichissement doivent se poursuivre de la façon suivante :

- Enrichissement de TopoCarto_Cogit à partir des concepts venant des définitions des spécifications générées par l'IRIT en appliquant TaxoMap Framework sur des graphes construits à partir de ces concepts situés dans Rameau.
- Enrichissement de TopoCarto_Cogit à partir des propriétés venant des définitions des spécifications générées par l'IRIT par intégration manuelle par le COGIT.

Le travail sur la restructuration d'ontologie est prévu en utilisant également TaxoMap Framework, un environnement dont l'implémentation a été effectuée. La méthodologie de restructuration reposera sur l'utilisation de patrons.

Les travaux du lot 2 réalisés sont les suivants :

- Sous-lot 2.1 : Alignement d'ontologies

Suite aux développements sur TaxoMap et TaxoMap Framework qui étaient terminés, le livrable 8 a pu voir le jour. Il contient une description complète de la version courante de TaxoMap incluant ainsi toutes les adaptations réalisées pour répondre aux besoins du projet Geonto, l'environnement TaxoMap Framework pour le raffinement de mappings et les tests réalisés dans le cadre du raffinement des mappings générés par le processus d'alignement entre Topo-Cogit et Carto-Cogit. Ces travaux ont été valorisés via des publications et des démonstrations (EKAW 2010, OAEI 2010, Atelier OntoGeo).

Nous prévoyons d'appliquer TaxoMap dans sa version courante pour aligner l'ontologie construite automatiquement par l'IRIT, une fois revue par le COGIT, de façon à obtenir une ontologie reflétant réellement le point de vue des textes, avec TopoCarto_Cogit. Cet alignement sera utile pour le travail de thèse de N. Abadie afin de faire le lien entre la base de données et l'ontologie.

- Sous-lot 2.2 : Réconciliation d'instances pour l'alignement d'ontologies

L'implémentation de l'approche combinant alignement d'ontologie et réconciliation d'instances décrite dans le livrable n°10 et livré en juillet 2010 est réalisée. Le logiciel devra être testé et validé avant d'être livré en juin 2011.

- Sous-lot 2.3 : Analyse des différences entre ontologies

Les travaux sur la comparaison d'ontologies ont été finalisés. Des tests ont été réalisés. Le travail a été valorisé via des publications (Atelier Ontology Matching et OntoGeo). Il a donné lieu à l'élaboration du livrable n°12 remis en février 2011.

Les travaux du lot 3 réalisés sont les suivants :

- Sous-lot 3.1 : Indexation automatique du contenu des documents

Finalisation de la conception et de la spécification du module logiciel « Indexation automatique du contenu des documents » par le LIUPPA, avec validation des toponymes à partir de ressources locales ou distantes et construction d'une sortie XML.

Conception d'une structure XML pour la production des index après récupération automatique locale ou distante des géométries des toponymes valides.

- Sous-lot 3.2 : Intégration, accès aux schémas de bases de données et évaluation

Le module logiciel « Indexation automatique de contenu de documents » a été conçu, spécifié et partiellement implémenté.

Publications liées au projet :

Conférences et ateliers internationaux (mono-partenaires)

N. Abadie, A. Mechouche, S. Mustière, OWL based formalisation of geographic databases specifications, EKAW 2010, 17th International Conference on Knowledge Engineering and Knowledge Management, Poster, 11th October-15th October 2010, Lisbon, Portugal.

F. Hamdi, C. Reynaud, B. Safar, Pattern-based Mapping Refinement, EKAW 2010, 17th International Conference on Knowledge Engineering and Knowledge Management, 11th October-15th October 2010, Lisbon, Portugal.

F. Hamdi, B. Safar, N. Niraula, C. Reynaud, TaxoMap alignment and refinement modules: results for OAEI 2010, Ontology Alignment Evaluation Initiative (OAEI) 2010 Campaign - ISWC Ontology Matching Workshop, Shanghai International Convention Center, Shanghai, China, Nov. 7, 2010.

Mechouche A. , N. Abadie, S. Mustière, 2010. Alignment based measure of the distance between potentially common parts of lightweight ontologies. *Fifth International Workshop on Ontology Matching (OM 2010)*, Shanghai, Nov. 7.

F. Saïs, N. Niraula, N. Pernelle, M.-C. Rousset, LN2R a knowledge based reference reconciliation system: OAEI 2010 results, Ontology Alignment Evaluation Initiative (OAEI) 2010 Campaign - ISWC Ontology Matching Workshop, Shanghai International Convention center, Shanghai, China, Nov. 7, 2010.

Conférences et ateliers d'audience nationale (mono-partenaires)

M.-N. Bessagnet, M. Gaio, E. Kergosien, C. Sallaberry, Extraction automatique d'un lexique à connotation géographique à des fins ontologiques dans un corpus de récits de voyages, TALN 2010, 19-23 juillet, Montréal.

F. Hamdi, C. Reynaud, B. Safar, 2010. TaxoMap Framework appliqué à l'alignement géographiques dans le projet GéOnto, Atelier OntoGeo associé à SAGEO'2010, Toulouse, 18 novembre 2010, pp. 51-53.

Mechouche A., Abadie N., Prouteau E., Mustière S. 2010. Vers un système en ligne d'aide à la découverte du contenu des bases de données géographiques. Démonstration à l'atelier OntoGéo, Toulouse, 18 nov. 2010.

Revue internationale (multi-partenaires)

L'article du consortium à la conférence Sageo de novembre 2009 a été sélectionné pour être publié dans la Revue Internationale de Géomatique.

Articles acceptés présentés en 2011 : GeoS 2011, EGC 2011 (3 articles), TSI (article accepté à JFO 2009 retenu pour un numéro spécial TSI)

Soumissions : TALN 2011, Revue RIG (multi-partenaires), Conférence « Int. Cartographic Conference ».

Difficultés rencontrées

L'avancement des travaux a été ralenti par le fait que l'alignement n'est pas directement applicable sur l'ontologie construite automatiquement. Ceci a nécessité de revoir l'approche pour aboutir à l'ontologie topographique visée. Nous avons convenu :

- De construire une ontologie, reflétant les textes, restreint au monde réel, à partir de l'ontologie construite automatiquement par l'IRIT, revue par le COGIT.
- D'aboutir à une ontologie topographique par enrichissement de TopoCarto_Cogit à l'aide (1) des propriétés venant des définitions des spécifications (intégrées par le COGIT à la main), (2) des concepts venant des définitions des spécifications intégrés automatiquement, (3) des « mini-ontologies » générées par « AugmOnto » à partir de termes extraits de textes grand public situés dans le thesaurus Rameau.
- De disposer de mises en correspondance entre le noyau d'ontologie généré par l'IRIT et revue par le COGIT et TopoCarto_Cogit.

Par ailleurs, nous devons souligner la difficulté que nous avons eu à définir une chaîne de traitement paramétrable et applicable à d'autres textes que ceux conformes à la DTD. Certaines parties du programme d'analyse des fichiers XML sont complètement spécifiques au document BD-Topo, ce qui remet en question la possibilité de produire un programme unique pour toutes les spécifications des bases de données du Cogit.

Un autre point difficile est l'intégration des traitements (réalisés avec différents logiciels et langages : Java puis XSLT et Prolog dans LinguaStream) et des résultats. L'application de TaxoMap sur les fragments d'ontologie transmis par le LIUPPA en sortie d'AugmOnto n'a, par ailleurs, pas été immédiat et a nécessité des ajustements.

Suivi des livrables du projet (d'après le planning accepté lors de la demande de prolongation)

(exemple, le tableau initial est celui contenu en annexe 1)

	Libellé	Nat.	Partenaires	Date	08 S1	08 S2	09 S1	09 S2	10 S1	10 S2	11 S1
T0	Coordination – Communication										
T0a	Mise en place d'une page web pour le projet		Tous	Début 2008	A						
T0b	Mise à jour page Web		Tous	Régulièrement	A	A	A				
T0c	Réunion de lancement		Tous	18/01/08	A						
T0d	Réunion de bilan semestre 1		Tous	13/06/08	A						
T0e	Réunion de bilan semestre 2		Tous	23/01/09		A					
T0f	Réunion de bilan semestre 3		Tous	30/06/09			A				
T0g	Réunion de bilan semestre 4		Tous	04/12/09				A			
T0h	Réunion de bilan semestre 5		Tous	21/06/10					A		
T0i	Réunion de bilan semestre 6		Tous	18/11/10						A	
T1	Lot 1 Construction et enrichissement d'ontologies										
T1a	Mise au point d'outils d'extraction de concepts et de relations : rapport intermédiaire	R	IRIT, LIUPPA, COGIT	Fin S2		X	A				
T1b	Mise au point d'outils d'extraction de concepts et de relations	Logiciel	IRIT	Fin S3			A				
T1c	Enrichissement d'une ontologie existante à partir de textes à l'aide des outils d'extraction et à partir des ressources lexicales	Logiciel	LIUPPA	Fin S3			A*	A*	A		
T1d	Mise au point d'outils d'extraction de concepts et de relation	R Module logiciel	IRIT	Fin S4				X		A	
T1e	Enrichissement d'une ontologie existante à partir de textes à l'aide des outils d'extraction	R Logiciel	LRI	Fin S6						R2	
T2	Lot 2 Appariement d'ontologies hétérogènes										
T2a	Alignement d'ontologies : rapport intermédiaire	R	LRI, COGIT	Fin S2		A					
T2b	Réconciliation d'instances pour l'alignement d'ontologies	R Logiciel	LRI	Fin S5					A	R1 (Log)	
T2c	Alignement d'ontologies	R Logiciel	LRI	Fin S6						A	
T2d	Analyse des différences entre ontologies pour faire ressortir les différences de points de vue sous-jacentes	R	COGIT	Fin S6						A	
T3	Lot 3 Exploitation des ontologies créées										
T3a	Intégration et accès aux schémas des bases de données	R	Cogit	Fin S3			A				
T3b	Indexation automatique de contenu de documents	R	LIUPPA	Fin S3			A				

Nat. : CR = Compte-rendu, R = rapport, ...

X : prévision initiale

A : atteint – A* : version livrée non finale

R1, R2, ... : reprévision

Commentaires

Préciser en particulier la raison de chaque reprévision de livrables (Ri)

Concernant le livrable n°7 (tâche T1d), sa livraison est intervenue en octobre 2010.

En T0+30, concernant la tâche T1c « Enrichissement », nous avons livré la chaîne de traitement intégrant « AugmOnto » réalisant l'extraction de termes candidats (livrable n°4). La suite du travail d'enrichissement portant sur l'utilisation de TaxoMap FrameWork fait l'objet de la tâche T1e

et du livrable n°9. Nous livrerons le rapport décrivant les résultats à ce jour et la version de TaxoMap Framework contenant les patterns d'enrichissement utilisés en mars 2011.

Concernant le livrable n°10 « Réconciliation d'instances pour l'alignement d'ontologies » (tâche n°2b), la livraison du logiciel est prévue pour Juin 2011 une fois celui-ci testé et validé. Le rapport a été livré en juin 2010.

Liste des CDD recrutés par des établissements publics dans le cadre du projet

Lister ici tous les CDD recrutés depuis le début du projet.

Numéro du Partenaire	Nom	Prénom	Qualifications	Date de recrutement	Durée du contrat (en mois)
1	HAMDI	Fayçal	Stagiaire recherche M2	10/03/2008	6 mois
1	HAMDI	Fayçal	Doctorant	05/11/2008	36 mois
4	NGUYEN	Van Tien	Doctorant	17/11/2008	36 mois (renouvelable par année)
2	MECHOUCHE	AMMAR	Post-doctorant	18/05/2009	16,5 mois (contrat initial de 18 mois achevé 1,5 mois avant la fin, le candidat ayant été nommé sur un poste d'ATER)
3	LAIGNELET	Marion	Post-doctorant	01/10/2009	12 mois à 4/5 de temps
3	CAPELLE	Jérôme	Stage L3	01/07/2009	1 mois
1	NIRAULA	Nobal	Ingénieur	01/04/10	3 mois et 2 semaines
2	PROUTEAU	Emeric	Stagiaire (master)	8/3/2010	5,4 mois

Equipements achetés par les partenaires dans le cadre du projet

Lister ici tous les équipements achetés depuis le début du projet

Numéro du Partenaire	Désignation	Date d'achat	Prix d'achat (en Euros)	Part financée par l'aide ANR (en Euros)
1	Mac Pro (sans écran)	Décembre 2008	2 036,94	2 036,94
4	Disque Dur 250	Mars 2008	119,79	119,79
4	2 Mémoires DDR 333Mhz 1go	Mai 2008	157,87	157,87
4	2 Mémoires SODIMM DDR 333Mhz 1go	Juin 2008	124,38	124,38
4	Portable pour doctorant	Décembre 2008	1205,38	1205,38
4	Ecran de bureau pour doctorant	Décembre 2008	249,00	249,00
3	Ordinateur individuel	Mars 2009	1300,00	1300,00
1	PC portable	Décembre 2009	1227,09	1227,09
2	Deux PC	Septembre 2009	2 x 822,95	2 x 822,95
3	4 PC, écrans, licences logiciels	Octobre 2009	5204,00	5204,00
4	Disque dur	Mars 2009	119,79	119,79
4	2 mémoires DDR 333Mhz 1 go	Mai 2009	157,87	157,87
4	2 mémoires SODIMM DDR 333 Mgz 1 go	Juin 2009	124,38	124,38
4	Portable pour doctorant + écran de bureau	Décembre 2009	1454,38	1454,38
4	Portable pour chercheur + station + écran de bureau	Mai 2010	2578,79	2578,79

