

30/01/2009

Rapport semestriel d'activité -coordonnateur
Programme MDCO - Edition 2007
Projet GEONTO – 2^{ème} semestre 2008

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)	1/07/2008 – 31/12/2008
Période couverte (t0+n mois à t0+m mois)	T0+6 à T0+12
Rédacteur (nom, téléphone, email)	Chantal Reynaud, 01 72 92 59 87, chantal.reynaud@lri.fr
Date	28 janvier 2009

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 30/01/2009.

Activités de coordination des activités du projet

Réunion plénière :

23/01/2009 : Réunion de bilan semestre 2 – Orsay – Tous les partenaires étaient présents (LRI, COGIT, IRIT, LIUPPA). Diane Penel (Délégation ANR - Calcul intensif) a présenté le fonctionnement des projets ANR.

Réunions de travail par lot :

Beaucoup de réunions avaient été organisées au démarrage du projet de façon à bien définir le travail à réaliser et à bien se coordonner entre partenaires. Le semestre 2 a permis d'avancer dans la réalisation des travaux. Il a nécessité beaucoup moins de réunions.

22/07/2008 : réunion de travail – Lot 2 – LRI, COGIT, Orsay

21/11/2008 : réunion de travail – Lot 1 – IRIT, LIUPPA, Toulouse

20/01/2009 : réunion de travail – Lot 2 – LRI, COGIT – Saint Mandé

Synthèse

Numéro du Partenaire	Conformité des résultats obtenus aux prévisions	Conformité de la consommation des ressources par rapport aux prévisions	Difficultés particulières
1	Conformes	Inférieure	Doctorant recruté le 5/10/08 (date prévue dans le projet : 1/1/08)
2	Conformes	Conforme	Aucune
3	Conformes	Inférieure	Etudiant de niveau post-doctoral non recruté
4	Conformes	Inférieure	Recrutement de doctorant tardif
Synthèse	Conformes	Inférieure pour 3 partenaires sur 4 – Conforme pour 1 partenaire	Difficultés de recrutement de personnel pour 3 partenaires sur 4.

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 tout d'abord porté sur la mise au point d'outils d'extraction de concepts et de relations (sous-lot 1.1) au sein de chaînes de traitement appliquées d'une part aux spécifications de la base géographique Topo-Cogit, d'autre part aux textes du corpus « récits de voyage ».

Ainsi, nous avons continué le développement en GATE de l'outil d'extraction à partir de patrons de façon à prendre en compte la structure hiérarchique des balises XML du document de spécification associée à BD-Topo, ainsi que le texte de certaines parties de ce document. Ces développements concernent l'extraction de concepts et de relations. Des expérimentations ont permis d'obtenir une ontologie OWL construite à partir des spécifications de BD-Topo.

Par ailleurs, des patrons de repérage de concepts et de relations géographiques spécifiques aux itinéraires contenus dans les textes de récits de voyage ont été définis ainsi que des techniques de repérage basées sur des règles grammatico-sémantique. Une implémentation dans une chaîne de traitement a été réalisée au sein de l'environnement LINGUASTREAM.

Dans le cadre des travaux sur l'enrichissement d'ontologie à partir de textes (sous-lot 1.2), nous avons réalisé un inventaire de termes associés aux entités géographiques nommées dans le corpus « récits de voyage ». Nous avons, par ailleurs, construit manuellement une ontologie minimaliste décrivant des itinéraires, ontologie qui devra ensuite être peuplée et améliorée en exploitant les informations contenues dans le corpus « récits de voyage ». L'étude de cette ontologie en liaison avec les textes doit permettre, par ailleurs, d'améliorer les patrons syntaxico-sémantiques utilisés dans la chaîne de traitement LINGUASTREAM (cf. ci-dessus sous-lot 1.1). L'objectif est également de faire coopérer cette ontologie avec celles de l'IGN.

Les travaux effectués dans le cadre du lot 2 ont consisté à poursuivre les tests réalisés sur les ontologies du Cogit. Les résultats ont été validés par le Cogit à l'aide de l'outil de visualisation d'ontologies et d'alignement, AlignViz, développé au cours du 1^{er} semestre au LRI. Une analyse de

la validité des propositions faites par l'outil d'alignement TaxoMap a conduit à identifier des améliorations et des extensions de l'outil, qui ont été présentées au Cogit lors de la réunion du 20/01/09 et à l'ensemble des partenaires à la réunion de bilan du 2^{ème} semestre (23/01/2009).

En parallèle, un travail sur le partitionnement d'ontologies a été mené dans le cadre d'un stage de master 2 Recherche. Les développements à venir de TaxoMap devraient s'appuyer, en partie, sur les résultats de ce travail pour identifier des contextes sémantiques proches dans deux ontologies distinctes.

Les travaux effectués dans le cadre du lot 3 ont porté essentiellement sur le sous-lot 3.2 avec la poursuite des travaux de thèse sur l'intégration de bases de données géographiques. Des premiers tests d'intégration ont été effectués pour tester la fiabilité de l'approche au sein de la plateforme GéOxygène du Cogit. Par ailleurs, le travail consistant à réaliser une interface de consultation des spécifications des bases a été finalisé.

Les travaux du sous-lot 3.1 sont retardés. Ce retard s'explique d'une part par les difficultés de recrutement du LIUPPA, d'autre part par le fait que nous ne disposons pas encore d'ontologie pouvant être exploitée dans cette tâche. Ce travail peut débuter maintenant que nous avons une vision plus formalisée de l'information que nous souhaitons extraire des textes.

Publications liées au projet :

Revue

Safar B., Reynaud C., Alignement d'ontologies basé sur des ressources complémentaires : illustration sur le système TaxoMap, TSI, à paraître en 2009.

Conférences et ateliers internationaux

Hamdi F., Zargayouna H., Safar B., Reynaud C., 2008, TaxoMap in the OAEI 2008 alignment contest, Proc. of OAEI-2008 in cooperation with the ISWC Ontology Matching Workshop, 26 oct., Karlsruhe, Germany.

Conférences nationales

Abadie N., Mustière S., 2008, Création d'une taxonomie géographique à partir des spécifications de bases de données. In SAGEO'08, Montpellier.

Hamdi F., Safar B., Zargayouna H., Reynaud C., Partitionnement d'ontologies pour le passage à l'échelle des techniques d'alignement, EGC'2009, 27-30 janvier 2009, Strasbourg.

Rapports de stage

Hamdi F., Partitionnement d'ontologies pour l'alignement, Rapport de stage de Master Recherche en Informatique, Université Paris-Sud, Orsay, 2008.

Horel T., Interface de consultation des spécifications de bases de données géographiques. Stage de 2^{ème} année d'Ecole d'Ingénieurs de l'ESTIA.

Minard A.-L., Recherche et analyse de ressources terminologiques liées à la topographie. Rapport de stage de Master 1 « Traitement automatique des langues », Université Lille 3.

Difficultés rencontrées

Les difficultés rencontrées concernent les recrutements :

- un doctorant a été recruté au LRI le 05/10/08 alors que le projet prévoyait l'embauche à partir du 01/01/08
- un post-doctorant sera recruté début 2009 à l'IRIT alors que son embauche avait été prévue pour 03/2008
- un doctorant a été recruté en 11/2008 au LIUPPA alors que son embauche avait été prévue en 09/2008

Pour l'instant, les résultats sont conformes aux prévisions mais ils ne sont toutefois pas aussi satisfaisants que nous l'aurions espéré, du fait de ces difficultés de recrutement. Ces difficultés expliquent également le fait que les travaux du sous-lot 3.2 sous la responsabilité du LIUPPA soient retardés.

Suivi des livrables du projet

	Libellé	Nat.	Partenaires	Date	08 S1	08 S2	09 S1	09 S2	10 S1	10 S2
T0	Coordination – Communication									
T0a	Mise en place d'une page web pour le projet		Tous	Début 2008	X					
T0b	Mise à jour page Web		Tous	Régulièrement	X	X				
T0c	Réunion de lancement		Tous	18/01/08	X					
T0d	Réunion de bilan semestre 1		Tous	13/06/08	X					
T0e	Réunion de bilan semestre 2		Tous	23/01/09		X				
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 2008		X	R1			
T2	Lot 2 Appariement d'ontologies hétérogènes									
T2a	Alignement d'ontologies : rapport intermédiaire	R	LRI, COGIT	Fin 2008		A				

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

X : prévision initiale

A : atteint

R1, R2, ... : reprévision

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 M2 recherche	10/03/2008	6 mois
1	HAMDI	Fayçal	Doctorant	05/11/2008	24 mois (avec l'objectif de prolonger de 12 mois)
4	Nguyen	Van Tien	Doctorant	17/11/2008	1 an renouvelable

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

Liste des livrables joints au présent rapport (uniquement pour les rapports de fin d'année)

Les livrables du projet sont fournis par le coordonnateur.

Numéro du livrable	Désignation	Forme/Support
2	Alignement d'ontologies	Rapport sous forme électronique

L'envoi du livrable n°1 « Mise au point d'outils d'extraction de concepts et de relations » est retardé. Il sera envoyé fin février 2009.