

Rapport semestriel d'activité – coordonnateur Programme MDCO – Edition 2007

Projet GEONTO - 1^{er} semestre 2008

Identification

Acronyme du projet	GEONTO
Numéro d'identification de l'acte attributif	ANR-07-MDCO-005
Coordonnateur (société/organisme)	LRI – Université Paris-Sud
Partenaire (société/organisme)	LRI – Université Paris-Sud
Période couverte (date à date)	01/01/2008 – 30/06/2008
Période couverte (t0 + n mois à t0 + m mois)	T0 à T0 + 6
Rédacteur (nom, téléphone, email)	Chantal Reynaud, 01 72 92 59 87
Date	15 juillet 2008

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 15 juillet 2008.

Activités de coordination des activités du projet

(lister les réunions, visites, ...)

Réunions plénières :

18/01/2008 : Réunion de lancement du projet – Orsay - LRI, COGIT, IRIT, LIUPPA

13/06/2008 : Réunion bilan semestre 1 – Pau - LRI, COGIT, IRIT, LIUPPA

Réunions de travail par lot :

15/02/2008 : Réunion de travail – Lot 1 - IRIT, LIUPPA – Toulouse

12/03/2008 : Réunion de travail – Lot 1 – IRIT, LIUPPA - Toulouse

28/03/2008 : Réunion de travail – Lot 1 - COGIT, IRIT, LIUPPA– Pau

29/04/2008 : Réunion de travail – Lot 1 – COGIT, IRIT, LIUPPA - Pau

13/05/2008 : Réunion de travail – Lot 1 - COGIT, IRIT, LIUPPA – Pau

21/05/2008 : Réunion de travail – Lot 2 - LRI, COGIT – Orsay

30/05/2008 : Réunion de travail – Lot 1 - IRIT, LIUPPA – Toulouse

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

Synthèse

Numéro du partenaire	Conformité des résultats obtenus aux prévisions	Conformité de la consommation des ressources par rapport	Difficultés particulières
----------------------	---	--	---------------------------

		aux prévisions	
1	Supérieurs	Inférieure	Difficultés à recruter un doctorant à compter du 1/1/08
2	Conformes	Conforme	-
3	Conformes	Inférieure	Pas de post-doc recruté.
4	Conformes	Inférieure	Stagiaires non rémunérés au LIUPPA
Synthèse	Conformes pour 3 partenaires sur 4 – Supérieurs pour 1 partenaire	Inférieure pour 3 partenaires sur 4. Conforme pour 1 partenaire	Difficultés de recrutement de personnel pour 2 des partenaires

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 études menées pendant les 6 premiers mois ont permis d'établir des résultats concernant l'extraction de concepts et de relations (sous-lot 1.1) mais aussi d'affiner et de reformuler les objectifs du projet.

1) Les ontologies de l'IGN déjà existantes sont des ontologies de référence ; elles ne sont pas utilisées comme point de départ du sous-lot 1.1. Le sous-lot 1.1 extrait des concepts, relations, propriétés des spécifications des bases de l'IGN et des récits de voyage du LIUPPA. Les ontologies de l'IGN déjà existantes sont utilisées en phase de validation.

2) L'approche par patron s'avère pertinente, moyennant la définition d'un processus supervisé car dans les documents de spécification, certaines formulations ou mise en forme peuvent avoir des significations différentes. Nous avons défini un certain nombre de patrons adaptés au corpus des spécifications et aux récits de voyage, pour des relations d'hypéronymie (classe à sous-classe).

3) Deux chaînes de traitement **génériques** seront mises au point. Elles intégreront des outils dont **l'usage est répandu** et mettront en œuvre une méthodologie **reproductible**. L'une permettra le traitement des spécifications associées aux schémas de bases de données, l'autre traitera les récits de voyage. Le prototypage de telles chaînes de traitements basées sur les environnements GATE et LINGUASTREAM a déjà débuté. Ces travaux ont été menés conjointement par le LIUPPA et l'IRIT.

4) Les types de connaissances pouvant être extraites des spécifications sont essentiellement des concepts et des relations hiérarchiques, de classes à sous-classes ou de composition « partie-tout » entre données spatiales. Toutefois, dans le cadre de GEONTO, l'accent devra être mis sur la définition de relations et de propriétés, les relations pouvant être des relations hiérarchiques ou non. Pour rendre les différents processus d'extraction plus productifs, nous envisageons alors élargir le corpus en explorant des textes pédagogiques ou encyclopédiques en géographie (cours, extraits de Wikipédia). Ces différentes ressources seront utilisées dans le processus d'enrichissement de l'ontologie TopoCarto-COGIT de l'IGN, qui est l'ontologie issue de la fusion manuelle de Topo-COGIT et Carto-COGIT.

Les travaux menés ces 6 premiers mois ont montré la nécessité de bien différencier :

- les ontologies pouvant être générées de manière semi-automatique à partir des documents de spécifications et à l'aide des chaînes de traitements qui seront implémentées,

Projet ANR-07-MDCO-005 GEONTO

- l'ontologie de référence, qui se veut une représentation beaucoup plus riche des connaissances du domaine de la géographie pertinentes pour la cartographie, la description d'itinéraires, etc. Celle-ci sera obtenue à partir de TopoCarto-COGIT enrichie à l'aide de ressources ontologiques, topologiques, textuelles et de Wikipedia.

Les travaux effectués dans le cadre du lot 2 se sont déroulés conformément aux prévisions. Les expérimentations ont été réalisées sur TOPO-COGIT et CARTO-COGIT dans leur version actuelle (deux ontologies fournies par l'IGN, qui ont été respectivement élaborées à partir des bases TOPO et CARTO). Des améliorations de l'outil d'alignement TaxoMap ont toutefois déjà été réalisées. Un outil logiciel d'aide à la validation des résultats obtenus, destiné aux experts, a également été mis au point.

Les travaux effectués dans le cadre du lot 3 ont consisté à initier la thèse « Intégration des bases de données à partir de la formalisation de leur spécifications ». Un travail sur la conception d'une interface de consultation des spécifications a été réalisé ainsi qu'une recherche et une analyse de ressources terminologiques liées à la topographie (ces ressources seront en particulier utiles au lot 1 et au sous-lot 3.2 (état de l'art)).

Publications liées au projet :

P. Loustau, T. Nodenot, M. Gaio « Spatial decision support in the pedagogical area : Processing travel stories to discover itineraries hidden beneath the surface » In Proceedings of the 11th AGILE International Conference on Geographic Information Science (AGILE 2008) The European Information Society: Taking Geoinformation Science One step Further, pp. 359-373, LNG&C, Springer

B. Safar, C. Reynaud, « Alignement d'ontologies basé sur des ressources complémentaires : illustration sur le système TaxoMap », article soumis à la revue TSI, en cours de révision. (*article non communicable avant de connaître la décision finale*)

J.A. Alla, Outil de visualisation d'ontologies et de validation d'alignement, Rapport de Projet de fin d'études d'Ingénieur - Ecole Mohammadia, Rabat.

Difficultés rencontrées

Les difficultés rencontrées concernent le recrutement de doctorant (LRI) ou d'étudiant de niveau post-doctoral (IRIT) aux dates indiquées dans le projet. Ces difficultés seront prochainement résolues de la façon suivante :

- doctorant recruté à partir de début octobre 2008 par le LRI
- recrutement d'un stagiaire ingénieur en octobre 2008 à l'IRIT et d'un post-doctorant à partir de janvier 2009.

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	A	Tous	Début 08	X					

Projet ANR-07-MDCO-005 GEONTO

T0b	Réunion de lancement	A	Tous	18/01/08	X					
T0c	Réunion du comité de pilotage	A	Tous	13/06/08	X					
T0d	Mise à jour page Web	A	Tous	Tout au long du semestre	X					

Nat. : CR = compte rendu, R = rapport

X : prévision initiale

A : atteint

R1, R2, ... : reprévision.

Commentaires

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

Les livrables du tableau initial de l'annexe 1 n'ont pas été modifiés.

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 mars 2008	6 mois
2	MINARD	Anne-Lyse	Master 1 TAL	5 mai 2008	2 mois
2	HOREL	Thomas	Ecole d'Ingénieur 2A	13 mai 2008	3 mois

Equipements achetés par des 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)
4	Disque Dur 250	18/03/2008	119,79€	119,79€
4	2 Mémoires DDR 333 Mhz 1go	23/06/2008	157,87€	157,87€
4	2 Mémoires SODIMM DDR 333Mhz 1go	15/05/2008	124,38€	124,38€

Projet ANR-07-MDCO-005 GEONTO

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

Les livrables du projet sont fournis par le coordonnateur.

Aucun livrable n'est à joindre au rapport en T0+6.