

Rapport semestriel d'activité - coordonnateur
Programme MDCO - Edition 2007
Projet GEONTO – 2^{ème} semestre 2009

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) | 01/07/2009 – 31/12/2009 |
| Période couverte (t0+n mois à t0+m mois) | T0+18 à T0+24 |
| Rédacteur (nom, téléphone, email) | Chantal Reynaud, 01 72 92 59 87 Chantal.reynaud@lri.fr |
| Date | 1 ^{er} février 2010 |

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 20/01/2010.

Activités de coordination des activités du projet

(lister les réunions, visites, ...)

Réunion plénière :

04/09/2009 : Evaluation mi-parcours : LRI, COGIT, IRIT, LIUPPA – Paris

04/12/2009 : réunion de bilan semestre 4 – Saint Mandé

Réunions de travail par lot :

02/11/2009 : réunion de travail – Lot 1 – LRI, IRIT – Orsay

06/11/2009 : réunion de travail – lot 1 – IRIT, LIUPPA - Pau

18/11/2009 : réunion de travail – Sous-lot 2.2 – LRI, COGIT – Orsay

18/11/2009 : réunion de travail – sous-lot 2.1 – LRI, COGIT – Orsay

Synthèse

| Numéro du Partenaire | Conformité des résultats obtenus aux prévisions | Conformité de la consommation des ressources par rapport aux prévisions | Difficultés particulières |
|----------------------|---|---|--|
| 1 | inférieurs aux prévisions | conforme | Retard pris dans les travaux du fait de l'embauche d'un doctorant uniquement à partir d'octobre 2008 |
| 2 | conforme | conforme | aucune |
| 3 | Inférieures aux prévisions | Consommation inférieure aux prévisions (moindre dépense de la | Calendrier différé de 6 mois |

| | | | |
|----------|--|--|---|
| | | ressource salaire car le recrutement d'une étudiante post-doctorale a été différé de 8 mois) | |
| 4 | conforme | Conforme (compte tenu du décalage annoncé en terme de recrutement du doctorant) | |
| Synthèse | Résultats inférieurs aux prévisions pour 2 partenaires sur 4 | Conforme pour 2 partenaires sur 4, inférieure pour 2 partenaires | Des retards dans l'embauche du personnel pour 3 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 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 de textes qui permet de produire une ontologie à partir de spécifications a été réécrite de façon à être davantage paramétrable et plus flexible (ajustement plus facile des règles à la structure du document traité, application partielle des règles possible). Les labels associés aux concepts OWL sont mieux exploités. La chaîne de traitement mise au point a été validée sur deux documents de spécification XML mis à disposition par le COGIT et associés aux bases de données BD-Topo et BD-Carto. L'ontologie construite à partir des spécifications associées à BD-Carto, BD-CARTO-IRIT, a été mise à disposition des autres partenaires du projet. Le logiciel de construction d'ontologie à partir de spécifications a été mis en ligne.

Les travaux sur l'extraction de concepts et de relations vont se poursuivre à l'IRIT avec le travail de Marion Laignelet, post-doctorante à l'IRIT. Des indices linguistiques ont déjà été identifiés afin de mieux structurer les concepts appris. Par ailleurs, un processus de vérification et correction de l'ontologie apprise basée sur l'utilisation d'outils d'alignement est en cours de spécification.

Ces travaux ont fait l'objet de deux publications (SAAKM 2009, TIA 2009)

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

Un travail a été réalisé conjointement entre l'IRIT et le LIUPPA de façon à définir un processus d'enrichissement de l'ontologie BD-Carto-IRIT à l'aide des termes trouvés dans des documents grand public et de relations trouvées dans le thésaurus Rameau. Par ailleurs, le statut de l'ontologie BD-Carto-IRIT et de son lien avec l'ontologie de référence déjà existante au COGIT a été précisé et les étapes d'alignement ont été définies afin d'enrichir ou de restructurer l'ontologie de référence avec BD-Carto-IRIT.

En parallèle, la plateforme Linguastream a été étendue pour permettre l'appariement de termes/requêtes et l'analyse des dépendances grammaticales.

Les travaux sur l'enrichissement ont fait l'objet d'une publication à JFO'09.

Les travaux du lot 2 réalisés sont les suivants :

- Sous-lot 2.1 : Alignement d'ontologies

L'adaptation de l'outil d'alignement TaxoMap a été poursuivi en introduisant des techniques complémentaires et en spécifiant et développant TaxoMap Framework dont l'objectif est de faciliter la spécification de traitements d'affinement de mappings et, à terme, de fusion, restructuration ou enrichissement d'ontologies. Ces travaux ont donné lieu à 3 publications.

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

Après une comparaison des différences entre les résultats obtenus par l'outil de réconciliation de l'IGN et l'outil LN2R du LRI, des adaptations de LN2R ont été proposées. Par ailleurs, des scénarii qui permettant de combiner les deux outils en tirant parti de leurs avantages respectifs ont été définis.

- Sous-lot 2.3 : Analyse des différences entre ontologies

Ce travail avait débuté en T0+18. Il s'est poursuivi par une étude portant sur la comparaison globale d'ontologies.

Les travaux du lot 3 réalisés sont les suivants :

- Sous-lot 3.1 : Le module logiciel « Indexation automatique du contenu des documents » a été conçu et implémenté.
- Sous-lot 3.2 : Continuation de la thèse sur l'intégration de bases de données géographiques. La structure générale d'un modèle de représentations en OWL2 des spécifications des bases de données géographiques a été mise au point. Ce modèle est le point d'entrée des outils de comparaison de bases de données.

L'ensemble des travaux réalisés dans le cadre du projet GEONTO a été publié à SAGEO 2009.

Publications liées au projet :

Chapitre de livre, Conférences et ateliers internationaux (mono-partenaires)

- F. Hamdi, B. Safar, C. Reynaud, H. Zargayouna, **Alignment-based Partitioning of Large-scale Ontologies**, Chapter of the book *Advances in Knowledge Discovery and Management, Studies in Computational Intelligence*, H. Briand, F. Guillet, G. Ritschard and D. Zighed (Ed.), Springer, accepté, à paraître.
- F. Hamdi, B. Safar, N. Niraula, C. Reynaud, **TaxoMap in the OAEI 2009 alignment contest**, *Ontology Alignment Evaluation Initiative (OAEI) 2009 Campaign - ISWC Ontology Matching Workshop*, Westfields Conference Center near Washington, DC, 2009.
- Kamel M., Aussenac-Gilles N., **How can document structure improve ontology learning?** (regular paper). *Semantic Authoring, Annotation and Knowledge Markup Workshop - collocated with K-CAP 2009 (SAAKM 2009), Redondo Beach, California (USA), 01/09/2009*, S. Handschuh, M. Sintek (Eds.), [CEUR Workshop Proceedings](#), p. 1-8, 2009.

Conférences nationales (mono-partenaires)

- F. Hamdi, C. Reynaud, B. Safar, **L'approche TaxoMap Framework et son application au raffinement de mappings**, *Congrès Francophone Reconnaissance des Formes et Intelligence Artificielle, RFIA 2010, Caen, 19-22 janvier 2010*.
- M. Kamel, N. Aussenac-Gilles. **Utiliser la Structure du Document dans le Processus de Construction d'Ontologies**. *Conférence Internationale sur la Terminologie et l'Intelligence Artificielle (TIA 2009), Toulouse (France), 18-20/11/09*, M.-C. L'Homme, S. Szulman (Eds.), [IRIT](#), (en ligne, 13pages), 2009.
<http://www.irit.fr/TIA09/thekey/articles/kamel-aussenac.pdf>

Conférences nationales (multi-partenaires)

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** (regular paper). *Journées Francophones sur les Ontologies (JFO 2009), Poitiers (France), 03-04/12/2009*, Bellatreche L., Kassel G., Thiran P. (Eds.), [ACM SIGAPPFR](#), p. 11-20, 2009.

S. Mustière, N. Abadie, N. Aussenac-Gilles, M.-N. Bessagnet, M. Kamel, E. Kergosien, C. Reynaud, B. Safar, **GéOnto : Enrichissement d'une taxonomie de concepts topographiques**, *Spatial Analysis and GEomatics (SAGEO), Paris, 25-26 Nov. 2009*.

Difficultés rencontrées

Les difficultés rencontrées concernent le retard dans le recrutement de personnel :

- doctorant au LRI recruté le 5/10/2008 au lieu du 1/1/2008
- post-doctorant au Cogit recruté fin mai 2008 au lieu de début 2008
- post-doctorant à l'IRIT recrutée uniquement en octobre 2009

Les dates de début des travaux ont été toutefois respectées mais on note un retard dans la progression du travail réalisé.

Ces retards nécessiteront une demande de prolongation du projet d'un peu moins d'un an de façon à assurer le financement de la dernière année de thèse du doctorant du partenaire 1 (LRI) et à atteindre les objectifs annoncés.

Suivi des livrables du projet

(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 |
|-----------|---|----------------------|---------------------|---------------|----------|----------|----------|----------|----------|----------|
| 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 | | |
| 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 | R1 | |
| T1d | Mise au point d'outils d'extraction de concepts et de relation | R Module logiciel | IRIT | Fin S4 | | | | A | R2 | |
| T2 | Lot 2 Appariement d'ontologies hétérogènes | | | | | | | | | |
| T2a | Alignement d'ontologies : rapport intermédiaire | R | LRI, COGIT | Fin S2 | | 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

R1, R2, ... : reprévision

Commentaires

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

Le livrable n° 4 (T1c) portant sur l'enrichissement d'une ontologie existante à partir de textes à l'aide des outils d'extraction et à partir de ressources lexicales a donné lieu à la livraison de la version V2, comme indiquée dans le rapport d'activité du semestre précédent. La version complète du logiciel intégrant l'outil d'enrichissement sera transmise en T0+30.

Concernant le livrable n°7 (T1d), une première version du logiciel est livrée à T0+24. Le rapport ainsi qu'une version plus élaborée seront livrées à T0+30.

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 | 24 mois (avec l'objectif de prolonger de 12 mois) |
| 4 | NGUYEN | Van Tien | Doctorant | 17/11/2008 | 1 an renouvelable |
| 2 | MECHOUCHE | AMMAR | Post-doctorant | 18/05/2009 | 18 mois |
| 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 |

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 |
|---|---|---------------|---------|---------|

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 |
|--------------------|--|---|
| 3 | Mise au point d'outils d'extraction de concepts et de relations | Démonstration + Logiciel accessible sur le site web du projet (rubrique livrables) |
| 4 | Enrichissement d'une ontologie existante à partir de textes à l'aide des outils d'extraction et à partir de ressources lexicales | Logiciel accessible sur le site web du projet (rubrique livrables) |
| 5 | Intégration et accès aux schémas de bases de données | Rapport sous forme électronique joint |
| 6 | Indexation automatique de contenu de documents | Rapport sous forme électronique joint |
| 4 | Enrichissement d'une ontologie existante à partir de textes à l'aide des outils d'extraction et à partir de ressources lexicales | Logiciel version V2 |
| 7 | Mise au point d'outils d'extraction de concepts et de relations | Logiciel |