

Compte rendu

Réunion semestrielle n°6 – Projet ANR MDCO – GEONTO

Toulouse le 18 novembre 2010 – 8h – 13h

Présents : N. Abadie (COGIT), N. Aussenac (IRIT), M.-N. Bessagnet (LIUPPA), M. Gaio (LIUPPA), Y. Mrabet (LRI), S. Mustière (COGIT), N. Pernelle (LRI), C. Reynaud (LRI), B. Safar (LRI). C. Sallaberry (LIUPPA)

Ordre du jour :

8h00 : Accueil

8h30 – 11h00 : Point sur les derniers livrables

Livrable 7 – IRIT : Mise au point d’outils d’extraction de concepts et de relations

Livrable 10 – LRI : Réconciliation d’instances pour l’alignement d’ontologies

Livrable 8 – LRI : Alignement d’ontologies

Livrable 9 – LRI : Enrichissement d’une ontologie existante à partir de textes à l’aide des outils d’extraction et à partir de ressources lexicales

Livrable 12 – COGIT : Analyse des différences entre ontologies pour faire ressortir les différences de points de vue sous-jacents

11h00 – 12h00 : Point sur le lot 3

Indexation automatique de contenu de documents - LIUPPA

Intégration et accès aux schémas des bases de données et validation - COGIT

12h00 - 13h00 : Discussion suite du projet

13h 00 : Déjeuner

1. Point sur le livrable 7 (IRIT) : Nathalie Aussenac

La livraison du livrable n°7 a été effectuée très récemment.

La partie provenant de l’analyse des définitions contenues dans les spécifications contient des fragments d’ontologies : des concepts (884 avec éventuellement des doublons), des propriétés (en petit nombre), mais pas forcément de liens entre ces éléments.

Par ailleurs, l’IRIT a construit un « noyau » d’ontologie automatiquement à partir des spécifications des bases de données de l’IGN. Ce « noyau » n’est pas aujourd’hui directement exploitable, notamment pour l’alignement par le LRI (cf. échanges IRIT – LRI de juin/juillet 2010)

Si on replace ces travaux dans le cadre général du projet, les tâches qui restent à effectuer sont les suivantes :

- 1) Enrichir TopoCarto_Cogit avec les fragments ontologiques extraits des définitions des spécifications des BD par l'IRIT.
- 2) Enrichir TopoCarto_Cogit avec l'ontologie « noyau » fournie par l'IRIT **restreint au monde réel**.
- 3) Intégrer tous les résultats produits par l'IRIT dans une même ontologie qui reflètera le point de vue des textes. Cette ontologie sera utilisée (1) pour l'enrichissement de TopoCarto_Cogit, (2) pour faire le lien entre la BD et l'ontologie dans le travail de thèse de N. Abadie.

Décisions prises :

- L'IRIT évalue rapidement en interne les fragments d'ontologie extraits de l'analyse des définitions des spécifications
- L'enrichissement de TopoCarto_Cogit à partir des concepts venant des définitions des spécifications est à automatiser. Le LIUPPA regarde si les termes existent dans Rameau et renvoie le graphe associé au LRI pour enrichissement à l'aide de TaxoMap Framework. Le COGIT intègre à la main dans TopoCarto_Cogit les propriétés venant des définitions des spécifications
- L'IRIT étudie les problèmes mis en évidence par le LRI dans l'ontologie « noyau », et effectue les corrections mineures possibles.
- Le COGIT étudie l'ontologie « noyau » et les fragments d'ontologie extraits des définitions des spécifications. Il construit manuellement (car nécessite des connaissances du domaine) une ontologie reflétant le point de vue des textes à partir de la production de l'IRIT.
- Le LRI enrichit TopoCarto_Cogit avec l'ontologie reflétant les textes qui sera produite par le COGIT **restreint au monde réel** (cette partie est à extraire par le LRI en enlevant les relations « représente »).

L'IRIT a proposé un sujet de stage de master sur l'enrichissement qui portera principalement sur l'enrichissement des relations.

2. Point sur les livrables 8 et 9 (LRI) : Brigitte Safar

Livrable 8 : Alignement

TaxoMap a été adapté (techniques nouvelles intégrées) et étendu par la conception de TaxoMap Framework permettant de raffiner les mappings générés en prenant en compte les spécificités des ontologies alignées.

Les expérimentations ont porté sur l'alignement de Topo_Cogit avec Carto_Cogit. Le Cogit a validé les mappings obtenus. Le rapport final est déjà écrit et sera livré en décembre conformément aux prévisions.

Livrable 4 et 9 : Enrichissement

Le LIUPPA a repéré des entités nommées. Des sous-graphes liés aux termes ont été extraits de Rameau → mini-ontologies fournies en entrée du processus d'enrichissement.

Le LRI a cherché à aligner ces sous-graphes avec Topo_Cogit.

→ pattern de validation de compatibilité de domaine autorisant ou non l'enrichissement

Le domaine est supposé proche si au moins 2 mappings d'équivalence sont trouvés entre la « mini-ontologie » associée au terme traité et Topo_Cogit.

Exemples :

- 1 exemple avec grotte qui marche très bien
- 1 exemple avec constructions souterraines pour illustrer un cas de contexte non valide. Selon le cogit, il est pertinent d'agir ainsi automatiquement. Les points de vue sont différents. L'enrichissement peut être possible mais pas automatiquement → Identification d'un cas qualifié de « semi-pertinent ».
- D'autres exemples permettent d'identifier des cas tout à fait hors contexte. Ex : bidonville
1 seul mapping is-A et tout un ensemble de concepts non mappés → à rejeter

Autres exemples illustrant les problèmes qui se posent :

- Ex. de Caverne : 2 mappings d'équivalence → contexte bon → enrichissement des fils
Pb : 5 ajouts faux → il est nécessaire de valider après l'enrichissement.
- Ex constructions rurales. 3 mappings d'équivalence → bon contexte mais tout ce qui sera importé ne sera pas bon.

Il reste à faire pour décembre :

- **recupérer du LIUPPA un échantillon plus gros de « mini-ontologies » Rameau**
- **implémenter le pattern de validation de compatibilité de domaine et faire l'exécution**
- **faire valider par le Cogit : le plus vite sera le mieux.**

Le LRI a également proposé une méthodologie d'enrichissement d'ontologies quand de très grosses ontologies sont considérées et qu'il s'agit d'ontologies généralistes. Expérimentations faites sur Yago.

→ recours au partitionnement mais le processus doit être adapté car l'identification des ancres est basé sur les mappings d'équivalence. Ceci n'est pas suffisant quand on est en présence d'ontologies généralistes.

Rq : l'idée que le LIUPPA récupère Yago et fasse la même chose que sur Rameau a été émise. Ceci n'est toutefois pas possible car Yago est en anglais.

En décembre : on prévoit de terminer l'implémentation de l'approche adaptée aux ontologies généralistes. La validation sera faite plus tard.

Rq : C'est en anglais : du bonus pour le cogit mais pas d'utilité à court terme.

3. Point sur le livrable n°10 (LRI) : Nathalie Pernelle

L'approche de réconciliation a été conçue et décrite dans le rapport livré en juillet 2010. L'implémentation est en retard du fait de difficultés à recruter.

Planning prévisionnel pour l'implémentation:

- 1^{er} sem. 2010 : un ingénieur a ré-implémenté LN2R pour une utilisation combinée. Ceci a permis de participer à OAEI 2010 (3^{ème}).
- 15 octobre – 15 janvier : embauche d'un stagiaire L3 pour implémenter le module d'alignement

Le logiciel sera livré en juin 2011, de façon à avoir le temps de valider le logiciel.

4. Point sur le livrable n°12 (COGIT) : Ammar Mechouche

Ce livrable porte sur l'analyse des différences entre ontologies. Le travail a été réalisé par A. Mechouche, en tant que post-doctorant au COGIT. Ammar a quitté le COGIT au 1^{er} septembre 2010. Le travail a été terminé. Le rapport final sera livré en décembre accompagné d'un prototype logiciel implémentant l'approche.

La méthode proposée compare seulement les parties analogues des ontologies, donne des indices et évalue les niveaux de détail. A rapprocher de ce qui est fait à Stanford. Le problème principal a été de trouver des mesures *interprétables*. La simplicité de la mesure a donc été un critère important.

5. Point sur le lot 3.2 (COGIT) : Sébastien Mustière et Nathalie Abadie

Le travail du stagiaire E. Prouteau portant sur un système d'aide en ligne pour découvrir le contenu de bases de données géographiques est terminé. Il fera l'objet d'une démonstration à l'atelier OntoGeo.

Concernant la thèse de Nathalie Abadie :

Le modèle de spécification formel a été présenté à EKAW. C'est le modèle actuel implémenté par Mouna Kamel à l'IRIT. Ce modèle a été testé sur de petits exemples de BD réelles. Il doit être un peu retouché pour être plus conforme en ce qui se fait en géomatique. La communauté a beaucoup recours aux ontologies de haut niveau → DOLCE préconisé. Le modèle sera fini en fin d'année civile.

Des exemples sur l'hydrographie ont été implémentés. **Nathalie aurait besoin de l'ontologie enrichie assez vite** pour ses expérimentations.

Ce qui reste à faire = l'alignement au niveau axiomes et l'intégration de l'outil de réconciliation d'instances du Cogit.

La fin de la thèse devra correspondre avec la fin du projet.

Rq de Sébastien M. : La partie interaction avec l'expert a été un peu négligée dans le projet. Pas assez identifiée dans le projet au départ. Devra être mis en avant dans les conclusions tirées du travail réalisé.

6. Point sur le lot 3.1 (LIUPPA) : Christian Sallaberry et Mauro Gaio

La chaîne de traitement pour faire de l'indexation de contenu sur la base de l'ontologie initiale (TopoCartoCogit) + Rameau a été réalisée. Pour la suite, il faudrait avoir une idée de la version finale de l'ontologie pour évaluer plus finement les textes, notamment pour faire des évaluations de la qualité de l'index.

Le modèle existe mais comment le remplir assez vite pour avancer aussi bien le lot 3.1 que le 3.2 → **le Cogit et le LIUPPA doivent se voir assez vite pour s'organiser et se répartir les tâches.**

→ engagement par ailleurs du LRI pour fournir d'ici fin décembre l'ontologie TopoCartoCogit enrichie avec les éléments de Rameau

Logiciel à livrer en T0+46 : pas de problème

Alignement entre ontologie enrichie et ontologie de l'IRIT suffit, pas besoin d'enrichir. Enrichissement de TopoCarto_Cogit le plus possible à partir de Rameau et Yago.

Ce qui est issu de l'IRIT sera étudié par le Cogit pour en faire un tout intégré. Il faudra prévoir non de l'utiliser pour enrichir l'ontologie enrichie mais de faire un alignement avec.

A faire :

- Regarder si la réalité de la BD et ce qu'a produit Mouna K. de l'IRIT est comparable, quitte à faire des modifications à la main.
- Prévoir une partie venant du LIUPPA dans le livrable sur l'enrichissement. Cette partie concernera le module d'extraction. **Prévoir un plan assez rapidement et le communiquer au LIUPPA.**

7. Questions diverses :

1. L'argent de l'ANR est-il utilisable après la fin du projet ? Si oui, pendant combien de temps ? CR se renseigne.
2. Sébastien a été approché (et l'IRIT aussi) par des industriels pour un projet de l'armée d'intégration d'informations géographiques. Pourra être mis en avant dans le projet Geonto comme une industrialisation. Contact suite au site de Geonto.
3. Approche de collègues de l'IGN anglais. Ont travaillé sur les ontologies et aimerait travailler avec l'IGN.
4. Sagem Défense (Cergy): plein de documentations techniques. Besoin d'une ontologie pour indexer : ontologie de navigation.
5. Suite de Geonto ? Les différents partenaires n'ont pas pris de décision sur le montage d'un projet faisant suite à Geonto. Le COGIT a eu plusieurs projets acceptés l'an passé. Il ne peut donner de suite cette année à Geonto. Le LRI est déjà impliqué dans plusieurs réponses à appels à projets cette année.

Fin de la réunion à 12h45.

Chantal Reynaud