

Compte rendu de la réunion de lancement du projet GeOnto

~

Vendredi 18 janvier 2008

Lieu de la réunion : INRIA Saclay Ile de France - Parc Orsay-Université, 4 rue Jacques Monod, Orsay – Salle G008 ~ 10h – 16h30

Présents :

LRI/IASI : Nathalie Pernelle, Brigitte Safar, Chantal Reynaud, Haïfa Zargayouna

COGIT : Nathalie Abadie, Sébastien Mustière

IRIT : Mouna Kamel, Ollivier Haemmerlé

LIUPPA : Mauro Gaio, Pierre Loustau, Christian Sallaberry

Excusés :

Nathalie Aussenac (IRIT), Françoise Angrand (ANR)

1. Introduction à la journée (Chantal Reynaud)

Présentation de l'ordre du jour. Rappel des objectifs généraux du projet, du contexte, du domaine d'application retenu et présentation synthétique des différents lots.

2. Présentations des différents partenaires

Chaque partenaire a brièvement présenté son équipe de recherche (cf. photocopies de transparents jointes) et les ressources, techniques ou approches apportées dans le projet.

COGIT/IGN : Sébastien Mustière

- Illustration des difficultés de l'appariement des schémas des bases de données TOPO et CARTO du fait de l'insuffisance de sémantique et du besoin d'exploiter les spécifications associées aux schémas de ces bases qui, elles, sont riches en information.
- Présentation des spécifications associées aux schémas des bases : des documents volumineux, très riches, structurés (structure parfois différente d'un schéma à un autre) mais comportant quand même du texte libre.
- Présentation des approches de construction d'ontologies à partir de spécifications et d'alignement déjà expérimentées au COGIT et des problèmes associés : les ontologies obtenues ne sont pas exhaustives et contiennent des erreurs (relations entre concepts parfois mal typées).
- Bilan sur les ressources disponibles : 2 taxonomies correspondant à deux bases de données géographiques (TOPO et CARTO), 1 taxonomie issue de leur fusion (réorganisation manuelle partielle)
- Présentation de travaux réalisés en appariement géographique de données.
- Résultats attendus rappelés : ontologie(s) plus riche(s) et plus juste(s), méthodologies de construction, d'alignement et de comparaison d'ontologies, exploitation des ontologies pour comprendre le point de vue décrit, intégrer des schémas de bases de données, accéder aux données géographiques via un portail.

IRIT : Mouna Kamel, Ollivier Haemmerlé

- Présentation de l'approche d'extraction de relations par patrons lexico-syntaxiques et son intégration au sein d'un processus de construction d'ontologie.
- Présentation de deux outils utilisés à l'IRIT : Cameleon développé à l'IRIT (gratuit et disponible – bibliothèque de 70 patrons validés – patrons définis par des non-

informaticiens) et la plate-forme d'ingénierie linguistique Gate (libre d'accès – très utilisée – application successive de ressources linguistiques – utilisable par des informaticiens : patrons définis en Jape - définition possible de classes sémantiques au sein des patrons). Le choix entre les deux outils peut aussi dépendre de l'objectif visé : enrichir, peupler, annoter, indexer une ontologie.

- Présentation du problème de l'extraction de relations entre des termes appartenant à des phrases différentes.
- Rappel de la contribution au projet : extraction de concepts à partir d'outils éprouvés, extraction de relations à partir de patrons lexico-syntaxiques (adaptation de patrons, définition de nouveaux patrons exploitant le type de document, sa structure), rechercher les paramètres d'une relation en exploitant les relations argumentatives, identifier les relations exprimées sur plusieurs phrases.

LIUPPA : Christian Sallaberry, Mauro Gaio, Pierre Loustau

- Présentation de l'approche de recherche d'information géographique par contenu du projet DESI : des traitements faits à l'échelle du syntagme nominal construit autour d'entités spatiales, des recherches de motifs d'une granularité supérieure s'appuyant sur l'extraction d'entités spatiales et temporelles, sur l'exploitation de connaissances complémentaires et sur la mise en œuvre de mécanismes de raisonnement spécifiques.
- Présentation du modèle pour représenter les entités spatiales faisant une distinction entre ces entités selon qu'elles peuvent être géo-localisées ou non (entités spatiales absolues ou relatives) et du modèle pour représenter les entités temporelles calendaires (absolues et relatives également). Ces modèles peuvent faire l'objet de traitements informatiques de façon, par exemple, à reconstruire un itinéraire à partir de son évocation dans un texte.
- Présentation du fonds documentaire patrimonial mis à disposition par la médiathèque de Pau. Les documents (récits de voyage) composant ce fonds décrivent le territoire local.

LRI : Brigitte Safar, Nathalie Pernelle

- Présentation de l'outil d'alignement de taxonomies TaxoMap : contexte d'application, types de mappings produits, nature des ressources mises en correspondance, caractéristiques du processus mis en œuvre, description des techniques composant le système et de la façon dont elles sont appliquées.
- Présentation du problème de la réconciliation des données et de l'approche LN2R, totalement automatique, combinant une approche logique inférant des décisions de réconciliation (ou de non réconciliation) sûres et une approche numérique fondée sur un calcul de similarité.
- Rappel des objectifs dans le projet : tester et étendre l'outil TaxoMap pour gérer l'hétérogénéité structurelle du point de vue de la précision, de la qualité et de la fiabilité de la représentation, analyser les résultats d'appariement dans un but de fusion, de réorganisation d'ontologie ou de compréhension des points de vue modélisés, combiner des techniques de recherche de mises en correspondance entre concepts et entre instances.

3. Point organisation

- Mise en place d'un site Web au plus tard à T0+6 sur lequel seront déposés les ressources mises à disposition du projet par les partenaires ainsi que les données relatives aux réunions organisées (ordre du jour, compte rendu, transparents des présentations). Responsable : Chantal Reynaud.

- Le COGIT communique des versions .doc des spécifications de bases de données géographiques dès que possible.
- Le LIUPPA communique très prochainement des documents textuels extraits d'un corpus documentaire numérisé constitué de récits de voyages.
- L'IGN dispose d'un lexique franco-anglais des termes géographiques (origine différente des ontologies). Il pourra être communiqué aux partenaires du projet.
- L'IRIT et le LIUPPA se mettent en relation pour travailler sur le problème de l'extraction des concepts et des relations. Une réunion est prévue le 15/02 à Toulouse pour approfondir et comparer les outils d'extraction des deux partenaires.
- Le COGIT et le LRI prévoient de se réunir début mars (la date précise sera fixée ultérieurement) suite aux premières expérimentations qui seront effectuées au LRI en alignement d'ontologies sur les ontologies TOPO et CARTO transmises par le COGIT. Le COGIT transmettra également la taxonomie (ONTOGEO) issue de la fusion de CARTO et TOPO.
- Le comité de pilotage du projet se réunira à la fin du 1^{er} semestre 2008 et fera le point sur l'avancement du projet dans sa globalité. La réunion aura lieu début juin à Pau.

Fin de la réunion de lancement du projet Geonto à 16h30.

Chantal Reynaud