



Sujet :

Construction collaborative d'ontologies

Niveau : Bac + 5 - 3^{ème} année d'école d'ingénieur – Master

Durée du stage : 4 à 6 mois

Contexte

Les ontologies sont des formalisations des connaissances d'un domaine. Elles sont le fondement du Web sémantique qui consiste à associer des connaissances manipulables par des machines aux documents disponibles sur le Web. La construction d'ontologies reste à ce jour un travail principalement manuel réalisé par des experts.

Par ailleurs, les systèmes de traitement automatique des langues ont atteint un niveau tel qu'il est désormais possible d'effectuer une analyse syntaxique assez complète des documents en langage naturel, ce qui permet l'apprentissage à partir de corpus de connaissances du niveau sémantique. Ainsi, le laboratoire LIC2M a conduit plusieurs thèses permettant d'une part le repérage de relations entre entités d'un domaine et d'autre part la découverte des différents sens des mots du langage.

Enfin, le LIC2M est partenaire du projet Scribo dont un lot porte sur la construction collaborative d'ontologies. Les connaissances acquises automatiquement ont besoin d'être validées de façon à pouvoir les exploiter avec confiance dans des applications. Les experts et les utilisateurs doivent être associés à cette validation. Il est essentiel que les conséquences des actions de validation ou de rejet des éléments de connaissances puissent être directement visibles, au sein d'interfaces graphiques permettant de constater ces conséquences sur la structure globale des connaissances et sur les applications qui les exploitent. Le processus d'acquisition doit donc être considéré comme un processus collaboratif entre l'homme et la machine. Dans le cadre de Scribo, les outils de construction d'ontologies produisent des "tickets" comprenant les connaissances proposées ainsi que diverses informations accessoires permettant à l'utilisateur de les valider ou non. Ces tickets sont ensuite exploités au sein d'ateliers graphiques. L'API permettant de soumettre des documents à l'outil et à celui-ci de répondre avec un "ticket" utilise le framework UIMA

Objectifs du stage

Le but de ce stage est de reprendre les prototypes de logiciels développés par le LIC2M, de les consolider et de les intégrer pour produire les "tickets" attendus par le projet Scribo.

Le stagiaire pourra être amené à développer des interfaces graphiques en Qt pour mettre en valeur ses développements dans l'attente de la disponibilité des outils Scribo.

Domaines de spécialité requis :

Traitement Automatique des Langues, Analyse sémantique, Web sémantique, Ontologies

Moyens informatiques mis en œuvre :

C++, Qt

Autres moyens mis en œuvre (expériences, méthodes d'analyses, autres...)



COMMISSARIAT A L'ENERGIE ATOMIQUE
DRT/DTSI/SRCI
Relation avec les écoles : MARC MERGY
Tél : 01 46 54 81 58
E-Mail : marc.mergy@cea.fr

Centre : Fontenay-aux-Roses

Pôle ou Direction : DRT - LIST

Dépt/Service/Labo : DTSI/SRCI/LIC2M :

Laboratoire d'Ingénierie de la Connaissance Multimédia Multilingue

Web : <http://www-list.cea.fr/>

Encadrement

Nom de l'ingénieur responsable : Gaël De Chalendar Tél. : +33 1 46 54 80 18

E-mail : gael.de-chalendar@cea.fr Fax : +33 1 46 54 75 80

Secrétariat : DTSI/SRCI Tél. : +33 1 46 54 91 17

Nom du chef de laboratoire : Olivier Mesnard Tél. : +33 1 46 54 73 38