

Intitulé du projet de recherche :

Construction collaborative de ressources sémantiques à partir de textes.

Problématique

Les ressources sémantiques sont des ensembles plus ou moins structurés de connaissances utilisées pour la recherche d'information, la communication entre agents et plus largement la conception de bases de connaissances dédiées à des domaines d'activités comme la médecine, la construction automobile ou le droit. Sous ce vocable, nous désignons aussi bien des ontologies que des thésaurus. La construction de ressources sémantiques constitue aujourd'hui un enjeu majeur pour de nombreuses applications, mais le terme de « ressource sémantique » recouvre des réalités très hétérogènes qui diffèrent par le domaine couvert, le degré d'abstraction, le niveau de granularité ou le langage utilisé, qui peut être plus ou moins formel. Nous nous intéressons particulièrement aux ensembles structurés de termes non ambigus dans un domaine donné. La construction de telles ressources n'est généralement pas l'œuvre d'une seule personne mais de plusieurs. De plus, des problèmes de point de vue et de cohérence peuvent surgir, nécessitant des arbitrages.

Contexte

Au sein du LIPN, l'équipe « Représentation des Connaissances et Langage Naturel » (RCLN) réunit des compétences dans différents domaines du traitement automatique des langues (terminologie, analyse syntaxique, analyse de corpus) ainsi que de l'ingénierie des connaissances (construction, fusion, alignement d'ontologies).

Ce travail de postdoctorat se fera dans le cadre du projet LegiLocal dont l'équipe RCLN est partenaire. L'objectif de ce projet est de simplifier l'accès des citoyens aux données juridiques, publiques, locales en proposant aux collectivités une solution qui s'intègre aisément à leur site web et qui s'appuie sur les technologies du Web Sémantique. L'équipe RCLN est fortement impliquée dans la modélisation et constitution de ressources terminologiques et ontologiques juridiques. Elle développe une méthode de construction de ressources sémantiques à partir de textes, Terminae. Cette méthode repose sur l'étude des résultats d'outils de traitement automatique des langues pour aider à conceptualiser les éléments de connaissances présents dans le corpus spécifique du domaine et de l'application visée. L'une des caractéristiques de la méthode est de construire un réseau termino-conceptuel, sorte de passerelle entre le niveau linguistique et le niveau conceptuel. L'outil Terminae permet la mise en œuvre de cette méthode.

Objectifs de la recherche

Différentes modalités de collaboration sont possibles selon que la tâche d'acquisition est distribuée de manière complémentaire entre des pairs ou que différents points de vue collaborent à la même tâche.

Nous avons comme objectif d'étudier et de mettre en place une approche collaborative afin de permettre l'utilisation conjointe de plusieurs acteurs lors de la constitution d'une ressource.

Dans le cadre de LegiLocal, le processus d'acquisition est partagé par les partenaires du projet qui ont des profils différents (juristes, terminologues et informaticiens). La ressource sémantique créée servira à annoter des documents qui seront ensuite retournés aux différents utilisateurs (secrétaire de mairie, citoyen).

Missions

* Faire une étude bibliographique sur les méthodologies et les outils de construction collaborative de ressources sémantiques

* Définir une méthodologie de construction

* Proposer des méthodes prenant en compte de points de vue dans Terminae

* Mettre en œuvre cette méthodologie sur les jeux de données du projet

* Interagir avec les partenaires pour valider la méthodologie

* Rédiger des rapports et des publications

Profil recherché :

Informaticien(ne) spécialiste en ingénierie des connaissances.

Durée : 9 mois prolongeable à 18 mois

Début souhaité : dès que possible (au plus tard janvier 2013)

Modalité de dépôt de candidature :

Merci d'envoyer un CV détaillant la formation et l'expérience acquise, accompagné d'une liste de publications et éventuellement de lettres de recommandation à :

Haïfa Zargayouna (haifa.zargayouna@lipn.univ-paris13.fr)

Sylvie Szulman (sylvie.szulman@lipn.univ-paris13.fr)