

Sujet de thèse en Informatique

Titre : Identification d'entités dans le web sémantique

Encadrants : Michel Leclère (leclere@lirmm.fr) et Marie-Laure Mugnier (mugnier@lirmm.fr),
équipe RCR du LIRMM

Résumé :

L'une des problématiques majeures du web sémantique consiste à interconnecter les informations provenant de différentes sources en exploitant les connaissances qui leur sont associées.

Citons notamment l'initiative « linked data », qui vise à promouvoir la mise à disposition et l'interconnexion des données sur le web afin de favoriser l'émergence d'un web de données à côté du web de documents. Les formats et ontologies développées pour le web sémantique permettent cette ouverture des données sur le web.

Cependant pour assurer une réelle interconnexion des diverses sources de données il est nécessaire de ne pas dupliquer les références d'entités. Le problème qui se pose alors est : existe-t-il déjà une référence pour mon entité ? Si oui quelle est-elle ? Si non, comment puis-je décrire ma nouvelle référence créée de façon à ce qu'elle se distingue des autres ? Un problème similaire est le suivant : n'y-a-t-il pas des références différentes qui désignent en fait la même entité, et qu'il faudrait « regrouper » sous une seule référence ?

Pour résoudre ces problèmes, on peut s'appuyer sur la description sémantique des entités. Pour distinguer des entités qui ont le même « nom », on peut chercher ce qui différencie chacune de toutes les autres entités de même nom.

L'ABES (Agence bibliographique de l'enseignement supérieur), basée à Montpellier, dispose de données permettant de bâtir de tels référentiels dans le domaine documentaire : référentiel d'œuvres, d'auteurs, de collectivités... cf. site de l'ABES : <http://www.abes.fr>.

L'objectif de ce travail est de définir un cadre formel pour ces problèmes de référencement des données puis de proposer des solutions à ces problèmes. Les données et applications de l'ABES fourniront un cadre propice à l'expérimentation et l'évaluation des solutions proposées.