

Sujet de stage Master Recherche

Optimisation de l'alignement de modèles par la prise en compte de mesure de similarité sémantique

Problématique

De nombreuses applications ont besoin de mesurer le degré de similarité entre termes ou concepts. Parmi ces applications, nous pouvons citer la recherche d'information, l'indexation de documents, le résumé automatique de textes ou encore l'alignement ou l'extraction d'ontologies. Le choix d'une bonne mesure est important puisqu'il aura une grande influence sur la pertinence des résultats obtenus. D'autre part, il est désormais admis que le langage XML (eXtensible Markup Language) est devenu le standard pour échanger de l'information au travers d'Internet. XML Schéma (XSD) est un formalisme permettant de décrire la structure d'un document XML de façon beaucoup plus précise qu'une simple DTD (Document Type Definition). L'utilisation d'XSD s'avère adaptée à la définition de modèles sous forme de graphes. La transformation d'un modèle XSD à l'autre implique d'abord de déterminer les appariements possibles de façon semi-automatique (voire presque entièrement automatique) grâce à une « bonne » mesure de similarité contextuelle qui prenne en compte les contextes des nœuds, les sémantiques des arcs et les contraintes sur ces nœuds et ces arcs.

Nous proposons dans ce stage d'optimiser l'appariement de schémas XSD par l'amélioration d'un algorithme de mesure de similarité contextuelle existant (voir www.iut.univ-paris8.fr/~lamolle/recherche/publication). Puis, de proposer une extension pour l'alignement d'ontologies exprimées en OWL.

Mots-clé : mesure de similarité, notion de contexte, appariement de modèles.

Objectifs du stage

1. Faire une étude des différents algorithmes existants concernant la mesure de similarité sémantique dans le cadre de l'appariement d'arbres ou de graphes.
2. Amélioration de l'algorithme intégrant la notion de contexte de nœuds pour optimiser l'appariement de schémas en tenant compte de la sémantique des nœuds et des arcs et de leurs contraintes.
3. Concevoir un outil simple de visualisation de l'appariement de schémas. Cet outil doit permettre à des experts du domaine de valider ou modifier le résultat de l'appariement facilement.

Profil souhaité : Le candidat devra posséder des connaissances en Java, XML Schéma et XSLT.

Rémunération : 30% du SMIC

Contacts :

Myriam Lamolle (m.lamolle@iut.univ-paris8.fr) ou Chan Leduc (chan.leduc@iut.univ-paris8.fr)

IUT de Montreuil, 140 rue de la Nouvelle France, 93100 Montreuil