

Combiner approche logique et numérique pour la réconciliation de données et l'alignement d'ontologies

Marie-Christine Rousset

LIG (Laboratoire d'Informatique de Grenoble)
681 rue de la passerelle B.P. 72, 38402 St MARTIN D'HERES CEDEX (FRANCE)
Marie-Christine.Rousset@imag.fr
<http://www-lsr.imag.fr/Les.Personnes/Marie-Christine.Rousset/>

La réconciliation de données et l'alignement d'ontologies sont deux problèmes au coeur de l'intégration d'informations, auxquels de nombreux travaux se sont attaqués soit par une approche symbolique, soit par une approche numérique. Dans cet exposé, nous montrons que combiner une approche logique et une approche numérique permet d'une part d'obtenir des mises en correspondance moins empiriques et qui ont plus de sens, et d'autre part un meilleur passage à l'échelle en exploitant au mieux l'inférence logique pour élaguer certains calculs numériques. Nous nous appuyons sur la logique, les systèmes d'équations ou les probabilités comme des formalismes mathématiques complémentaires permettant de traduire et exploiter la sémantique d'un schéma médiateur ou d'une ontologie. Pour illustrer notre propos, nous présentons d'abord une méthode de réconciliation de références qui s'applique à des sources de données RDF conformes à un même schéma. Cette méthode commence par inférer des décisions certaines de réconciliation et de non réconciliation qui découlent logiquement des contraintes sémantiques déclarées au niveau du schéma. Puis elle calcule la similarité entre les paires de références non traitées par la méthode logique par la résolution d'un système d'équations modélisant les dépendances entre réconciliations découlant de la description sémantique associée à chaque référence. Nous présentons ensuite une méthode d'alignement d'ontologies qui combine inférence logique et classification bayésienne pour estimer la probabilité de mappings entre classes à partir des méta-données des instances déclarées ou inférées dans ces classes. Cette méthode s'appuie sur un algorithme qui, à partir d'un ensemble de mappings candidats qui peut être l'ensemble de tous les mappings possibles, fournit l'ensemble des mappings dont la probabilité dépasse un certain seuil. Cet algorithme passe à l'échelle grâce à l'exploitation d'une propriété de monotonie de la fonction de probabilité par rapport à un ordre sur les mappings fondé sur leur sémantique logique.