

Une mesure de distance dans l'espace des alignements entre parties potentiellement homologues de deux ontologies légères

Ammar Mechouche, Nathalie Abadie, Sébastien Mustière

Institut Géographique National, Laboratoire Cogit, 73 Av. de Paris, 94160 St-Mandé, France

Résumé. Nous proposons dans cet article une méthode qui calcule la distance entre ontologies dans un but d'aide à la décision sur la pertinence ou non de leur fusion. Cette méthode calcule la distance entre parties homologues de deux ontologies par rapport à leurs niveaux de détail et leurs structures taxonomiques, et ce en exploitant les correspondances produites par un alignement préalablement effectué entre ces ontologies, et en adaptant la méthode de la distance d'édition entre arbres ordonnés. Nous limitons notre étude ici aux ontologies légères, c'est-à-dire des taxonomies représentées en langages OWL, le langage d'ontologies pour le Web. Notre méthode a été implémentée et testée sur des ontologies réelles, et les résultats obtenus semblent prometteurs.

1 Introduction

Calculer la distance entre ontologies du même domaine a plusieurs intérêts. En effet, ceci permet d'améliorer la recherche d'ontologies sur le Web, afin de 1) retrouver des ontologies qui sont susceptibles de remplacer d'autres (David et Euzenat, 2008), 2) retrouver des ontologies qui peuvent en enrichir d'autres, 3) retrouver des ontologies sur lesquelles on peut propager une requête, 4) retrouver une communauté de personnes qui utilisent les mêmes ontologies ou des ontologies proches, dans le but d'établir des collaborations, etc. Evaluer la distance entre ontologies peut être aussi utile dans l'étude de l'évolution d'ontologies, pour savoir par exemple dans quelle mesure la structure d'une ontologie, mise à jour par plusieurs personnes, a évolué. Enfin, disposer d'une distance entre ontologies peut être utile afin d'aider l'utilisateur à savoir si deux ontologies peuvent être fusionnées ou juste mises en correspondance. Peu de travaux de la littérature ont abordé cette problématique (Maedche et Staab, 2002) (David et Euzenat, 2008) (Wang et al., 2008) (Ngan et al., 2009). De plus, les mesures proposées restent très globales et n'évaluent pas la distance selon des critères bien précis, ce qui les rend difficilement interprétables.

Dans notre cas, il s'agit de déterminer, parmi les ontologies décrivant les données disponibles, celles couvrant l'ensemble ou tout au moins une partie du domaine d'intérêt, ainsi que la différence du niveau de détail et de la structure des connaissances décrites par celles-ci. De telles informations constituent en effet une indication importante sur la proximité ou la complémentarité à la fois thématique et structurelle des différentes sources de données décrites par ces ontologies, permettant de juger par avance de la pertinence de leur éventuelle intégration en vue d'analyses conjointes.

En effet, dans le domaine géographique, qui est notre domaine d'intérêt, où les sources de données sont annotées en utilisant des ontologies hétérogènes, trois principaux critères doivent pouvoir être évalués par une mesure de distance: le premier critère concerne les thématiques décrites par les deux ontologies comparées. On cherche à savoir si l'ontologie source décrit exactement le même domaine que l'ontologie cible ou si elle fournit des connaissances