

Extraction de sous-parties ciblées d'une ontologie généraliste pour enrichir une ontologie particulière

Fayçal Hamdi, Brigitte Safar, Chantal Reynaud

LRI, Bât. 650, Université Paris-Sud 11
91405 Orsay Cedex France
{faycal.hamdi, safar, chantal.reynaud}@lri.fr,
<http://www.lri.fr/~hamdi>

Résumé. Différentes ressources ontologiques généralistes de très grande taille ont été développées de façon collective et sont aujourd'hui disponibles sur le web. Ainsi l'ontologie YAGO est une énorme base de connaissances décrivant plus de 2 millions d'entités. Afin de tirer parti de ce gigantesque travail collectif, nous montrons comment en extraire des sous-parties thématiquement focalisées pour enrichir une autre ontologie, dite cible, de taille plus limitée mais de domaine centré sur une application particulière ¹.

1 Introduction

L'ontologie YAGO (Yet Another Great Ontology) est une énorme base de connaissances, développée à l'Institut d'informatique Max Planck de Sarrebruck, décrivant plus de 2 millions d'entités, des personnes, des organisations, des villes et contenant plus de 20 millions de faits sur ces entités. Ses données proviennent de Wikipedia et sont structurées à l'aide de WordNet. De par sa taille et ses multi-thématiques, YAGO est difficile à utiliser directement pour des applications particulières. Afin de tirer parti plus aisément de ce gigantesque travail collectif, l'objectif de cet article est de montrer comment on peut en extraire des sous-parties thématiquement focalisées pour enrichir une autre ontologie dite cible, de taille plus limitée mais de domaine centré pour une application particulière.

Pour procéder à l'extraction des sous-parties pertinentes de YAGO, nous adaptons l'outil TaxoPart (Hamdi et al., 2009b), développé pour partitionner des ontologies de grande taille avant de les aligner. L'enrichissement de l'ontologie cible est ensuite effectué en procédant par alignement des sous-parties extraites de YAGO avec l'ontologie à enrichir, en utilisant l'outil d'alignement TaxoMap (Hamdi et al., 2009a). Les concepts extraits de YAGO apparaissant dans l'alignement produit sont considérés comme des concepts candidats pour l'enrichissement, liables à ceux de la cible par les relations établies dans les mappings.

Nous commençons par rappeler en section 2, le fonctionnement de TaxoPart, avant de présenter, en section 3, les adaptations faites pour l'utiliser au mieux dans ce contexte. Les résultats d'une expérimentation menée sur une ontologie topographique sont présentés en section 4, les travaux proches et nos conclusions en section 5.

1. Ce travail est financé par l'Agence Nationale de la Recherche (ANR) au travers du projet ANR-07-MDCO-005 pour la création, la comparaison et l'exploitation d'ontologies géographiques hétérogènes (<http://geonto.lri.fr/>).