

# Partitionnement d'ontologies pour le passage à l'échelle des techniques d'alignement

Fayçal Hamdi \*, Brigitte Safar\*  
Haïfa Zargayouna\*,\*\*, Chantal Reynaud\*

\*LRI, Université Paris-Sud, Bât. G, INRIA Futurs  
2-4 rue Jacques Monod, F-91893 Orsay, France  
{Faycal.Hamdi, safar, reynaud}@lri.fr,  
<http://www.lri.fr>

\*\*LIPN, Université Paris 13 - CNRS UMR 7030,  
99 av. J.B. Clément, 93440 Villetaneuse, France.  
[Haifa.Zargayouna@lipn.univ-paris13.fr](mailto:Haifa.Zargayouna@lipn.univ-paris13.fr)

**Résumé.** L'alignement d'ontologies est une tâche importante dans les systèmes d'intégration puisqu'elle autorise la prise en compte conjointe de ressources décrites par des ontologies différentes, en identifiant des appariements entre concepts. Avec l'apparition de très grandes ontologies dans des domaines comme la médecine ou l'agronomie, les techniques d'alignement, qui mettent souvent en œuvre des calculs complexes, se trouvent face à un défi : passer à l'échelle. Pour relever ce défi, nous proposons dans cet article deux méthodes de partitionnement, conçues pour prendre en compte, le plus tôt possible, l'objectif d'alignement. Ces méthodes permettent de décomposer les deux ontologies à aligner en deux ensembles de blocs de taille limitée et tels que les éléments susceptibles d'être appariés se retrouvent concentrés dans un ensemble minimal de blocs qui seront effectivement comparés. Les résultats des tests effectués avec nos deux méthodes sur différents couples d'ontologies montrent leur efficacité.

## 1 Introduction

Le développement rapide des technologies internet a engendré un intérêt croissant dans la recherche sur le partage et l'intégration de sources dispersées dans un environnement distribué. Le Web sémantique (Berners-Lee et al., 2001) offre la possibilité à des agents logiciels d'exploiter des représentations du contenu des sources. Les ontologies ont été reconnues comme une composante essentielle pour le partage des connaissances et la réalisation de cette vision. En définissant les concepts associés à des domaines particuliers, elles permettent à la fois de décrire le contenu des sources à intégrer et d'explicitier le vocabulaire utilisable dans des requêtes par des utilisateurs. Toutefois, il est peu probable qu'une ontologie globale couvrant l'ensemble des systèmes distribués puisse être développée. Dans la pratique, les ontologies de différents systèmes sont développées indépendamment les unes des autres par des communautés différentes. Ainsi, si les connaissances et les données doivent être partagées, il est essentiel d'établir des correspondances sémantiques entre les ontologies qui les décrivent. La tâche