

# Techniques structurelles d'alignement pour portails Web

Chantal Reynaud, Brigitte Safar

Université Paris-Sud XI, CNRS (L.R.I.) & INRIA (Futurs)

91405 Orsay cedex

{chantal.reynaud, brigitte.safar}@lri.fr

<http://www.lri.fr/~cr>

**Résumé.** Le travail décrit dans cet article a pour objectif d'unifier l'accès aux documents d'un domaine d'application. Il permet en particulier d'augmenter le nombre de documents accessibles à partir de portails Web sans en modifier l'interface d'interrogation. L'accès aux documents est supposé s'appuyer sur des taxonomies que nous proposons d'aligner. Cet article porte spécifiquement sur les techniques structurelles mises en œuvre, qui sont originales et particulières dans la mesure où elles sont adaptées au traitement de taxonomies dont les structures sont hétérogènes et dissymétriques. Nous présentons et analysons ensuite les résultats de trois expérimentations effectuées sur diverses taxonomies, des taxonomies réelles qui ont motivé notre approche ainsi que des taxonomies tests mises à disposition des chercheurs de la communauté.

## 1 Introduction

La recherche de documents pertinents sur le Web est une tâche encore souvent laborieuse. Le Web sémantique devrait faciliter ce travail en réalisant un appariement sémantique entre la requête de l'utilisateur et les documents indexés. Les techniques d'alignement de méta-données ou d'ontologies sont pour cela de la plus grande importance.

Notre travail porte sur de telles techniques, utilisables dans le contexte du Web. Notre objectif est de permettre un accès unifié via le Web aux documents d'un même domaine d'application. Plus précisément, nous proposons d'aligner la taxonomie d'un portail Web avec celle de documents externes de façon à augmenter le nombre de documents accessibles à partir de ce portail sans en modifier l'interface d'interrogation. La recherche de tels documents s'appuie en général sur des ontologies très simples, souvent réduites à des hiérarchies de classification, c'est-à-dire des taxonomies. Les approches basées sur des représentations plus fines, telles des représentations OWL, exploitent toute la richesse du langage. Elles ne sont pas adaptées au traitement de simples taxonomies.

Les taxonomies à appairer, tout en représentant la structure d'un même domaine, peuvent comporter des termes différents ou des termes identiques structurés différemment. Chaque concepteur utilise son propre vocabulaire. Il ne s'agit pas de représentations uniformes mais de vues spécifiques à chaque concepteur, certaines étant des descriptions avec un niveau de détail important, des concepts très spécialisés y sont alors représentés, d'autres étant des descriptions plus macroscopiques et se limitant à la représentation de concepts généraux. Selon le cas, la profondeur et la taille des taxonomies varient.