

Règles de propagation pour la création d'ontologies d'annotation de ressources

Lydia Abrouk, Pierre Pompidor
Danièle Hérin, Michel Sala

LIRMM
161, rue ada
34392, Montpellier
{abrouk,pompidor,dh,sala}@lirmm.fr,
<http://www.lirmm.fr/~{abrouk,pompidor,dh,sala}>

Résumé. L'annotation se distingue de l'indexation automatique par l'utilisation d'une ou plusieurs ontologies qui définissent un domaine global de référence permettant de cadrer et de normaliser les annotations effectuées, par ailleurs une ressource annotée doit l'être non pas par une liste de mots-clés, mais bien par une ou plusieurs ontologies.

Malheureusement, il est peu réaliste de penser que les centaines de millions de ressources mises à disposition sur le Web puissent être annotées par leurs auteurs. Pour résoudre ce problème, notre démarche consiste à indexer les documents en se basant sur l'ontologie globale et ensuite propager les annotations en utilisant des documents déjà annotés pour annoter d'autres documents référencés par ceux-ci.

La propagation des annotations suit des règles que nous proposons dans cet article. L'illustration est effectuée sur un corpus de livres dont le thème relève de l'informatique.

1 Introduction

Le Web est devenu dans tous les domaines la plus grande source d'informations, engendrant lors de la recherche d'informations de grandes difficultés pour retrouver les ressources les plus pertinentes, celles-ci provenant d'ailleurs de sources hétérogènes et parfois sécurisées (et dans ce cas accessibles seulement par un fournisseur donné). Une solution est d'annoter les ressources pour décrire non seulement leur contexte de création (noms des auteurs, date de parution, etc.), ce qui ne nous intéresse pas ici, mais également la sémantique de leur contenu. Dans cet article nous nous intéressons uniquement à ce deuxième point.

L'annotation est caractérisée par l'utilisation d'une ou plusieurs ontologies qui définissent un domaine global de référence permettant de cadrer et de normaliser les annotations effectuées. Nous pouvons par exemple citer DMOZ¹, une ontologie universelle utilisée par certains moteurs de recherche comme Google². Naturellement, l'annotation est d'autant meilleure qu'elle privilégie une intervention humaine, une annotation

¹www.dmoz.com

²www.google.com