

Réécriture de requêtes multimédias : approche basée sur l'usage d'une ontologie

Samira Hammiche, Salima Benbernou, Mohand-Saïd Hacid
Université Claude Bernard Lyon 1 – Bâtiment Nautibus
43, boulevard du 11 novembre 1918
69622 Villeurbanne - France
{shammich, sbenbern, mshacid}@liris.univ-lyon1.fr

Résumé. Nous proposons dans cet article une stratégie de réécriture de requêtes sur des données multimédias décrites moyennant le standard MPEG-7. Ce standard se base sur XML schéma qui permet de décrire la structure des données. Cependant, aucune sémantique n'est assignée à cette structure. Nous proposons d'étendre ce standard d'une ontologie permettant d'exprimer les connaissances du domaine. Ainsi, l'ontologie sera utilisée durant l'indexation des données multimédias et la réécriture de requêtes. Le but de la réécriture de requêtes est de transformer une requête initiale en une ou plusieurs requêtes équivalentes ou sémantiquement proches compte tenu des connaissances représentées dans l'ontologie.

1 Introduction

L'avènement du Web a contribué à la prolifération massive d'informations multimédias. Cette augmentation rend la recherche de ce type de données de plus en plus difficile. Afin de pallier ce problème, le standard MPEG-7 (Martinez 2003) a été proposé. La principale ambition de MPEG-7 est de rendre les informations multimédias aussi facile à trouver sur le Web que l'est le texte aujourd'hui.

Le standard MPEG-7 fournit différents outils pour décrire la plupart des informations du contenu multimédia. Il est basé sur XML schéma, ce qui permet de structurer et de contrôler les descriptions des contenus. Cependant, aucune sémantique n'est assignée aux descriptions. La formalisation de ces descriptions est une piste pour faciliter la recherche dans les documents MPEG-7. Nous proposons d'augmenter le standard MPEG-7 d'une ontologie du domaine, qui peut être utilisée comme source de descripteurs permettant d'indexer les données multimédias. Afin d'éviter les réponses vides, nous proposons de réécrire les requêtes, avant leur évaluation, en s'appuyant sur les mécanismes de raisonnement offert par le formalisme de spécification de l'ontologie. La requête réécrite est, par la suite, évaluée. A cet effet, nous nous basons sur l'appariement d'arbres comme stratégie de recherche vu la représentation arborescente des descriptions MPEG-7 et des requêtes.

Cet article est organisé comme suit. Dans la Section 2, nous présentons le standard MPEG-7, la représentation et l'interrogation des descriptions MPEG-7. Dans la Section 3, nous décrivons de façon générale le cadre de recherche. La Section 4 présente les ontologies et leur représentation en logique de description. La Section 5 décrit le processus de réécriture de requêtes sur des données multimédias à travers un exemple.